Large Sample Inference in Finite Population Problems

ANURAG DEY

Indian Statistical Institute

April 2024

INDIAN STATISTICAL INSTITUTE

DOCTORAL THESIS

Large Sample Inference in Finite Population Problems

Author:
ANURAG DEY

Supervisor:
PROBAL CHAUDHURI

A thesis submitted to the Indian Statistical Institute
in partial fulfilment of the requirements for
the degree of
Doctor of Philosophy in Statistics

Theoretical Statistics & Mathematics Unit

Indian Statistical Institute, Kolkata

April 2024

To my parents and my teachers

Acknowledgements

I would like to express my love and respect towards my family members for their enormous support. Their constant encouragement and motivation gave me the confidence to pursue my research work.

I would like to express my sincere gratitude to my supervisor, Prof. Probal Chaudhuri. His critical remarks and helpful suggestions have an immense impact on my research work. He has invested a huge amount of time on me, and I am indebted to him for whatever I have learnt from him.

I would like to express my deep gratitude to my teachers in Presidency University. They introduced me to the subject *Statistics*. I was benefited from the courses that I took during my M.Sc. in Indian Institute of Technology (IIT), Kanpur, and Ph.D. coursework in Indian Statistical Institute (ISI), Kolkata. I thank all my teachers in IIT and ISI.

I would like to thank Mausumi Bose, Aloke Kar, Sandip Mitra, Prabir Chaudhury, Arijit Chaudhuri, Anirvan Chakraborty, Tathagata Bandyopadhyay, Anil Kr. Ghosh, Arijit Chakrabarty, Samarjit Das, Rituparna Sen and Late Asit Baran Chakraborty for various discussions about my research work. Many of them served on review committees evaluating my research at different stages of my tenure as a Ph.D. student in ISI. I am also thankful to two anonymous reviewers, who carefully reviewed an earlier version of my thesis and provided me with several constructive criticisms and helpful suggestions. Those have greatly improved the presentation and the contents of the thesis. I would also like to thank my viva examiners Sugata Sen Roy and Arijit Chakrabarti for their valuable comments and remarks.

I thank all my seniors and juniors in ISI for various academic as well as non-academic discussions that I had with them. I am fortunate enough to have seniors like Joydeep da, Soham da, Sukrit da and Bishakh da, and juniors like Bilol, Bodhisatta, Subhodeep, Bhowal, Deborshi and Sourav. I also thank Arijit and Priyanka, who were my batch mates in ISI. I want to express my gratitude to all of my friends for their encouragement and support, especially Aakash, Sayak, and Shreyan.

I acknowledge ISI for providing me with research fellowship and wonderful research facilities. I am thankful to the administrative staff of Indian Statistical Institute, especially those in the Dean's Office and the Stat-Math Office in Kolkata, for their help and assistance at various times during my tenure as a Ph.D. student.

I would also like to thank Irish Social Science Data Archive (ISSDA) and its administrator Ruby O'Riordan for making Electricity Customer Behaviour Trial data available for my research work.

Anurag Dey

Anurag Dey

April 2024

Contents

List of symbols

$\mathcal{P}$=$\{1, 2, \ldots, N\}$	Finite population
N	Number units in a finite population
s	Sample
n	Number units in a sample
$\mathcal{S}$	Sample space
$\mathcal{A}$	Power set of $\mathcal{S}$
$P(s)$	Sampling design
$d(i, s)$	Sampling design weight
$(\Omega, \mathcal{F}, \mathbf{P})$	Probability space associated with superpopulation model
$P(s, \omega)$	Sampling design under superpopulation models
$\mathbf{P}^*$	A probability measure defined on the product space ($\mathcal{S} \times \Omega$, $\mathcal{A} \times \mathcal{F}$) as $\mathbf{P}^*(B \times E)$=$\int_E \sum_{s \in B} P(s, \omega) \mathrm{d}\mathbf{P}(\omega)$ for $B \in \mathcal{A}$ and $E \in \mathcal{F}$
y	Study variable
x	Size variable
z	Covariate
$\pi_1, \ldots, \pi_N$	Inclusion probabilities
$G_1, \ldots, G_n$	x-totals of the groups of the population units formed in the first step of the RHC sampling design
$\mathcal{H}$	Infinite dimensional separable Hilbert space
$\langle \cdot, \cdot \rangle$	Inner product in $\mathcal{H}$
$\|\| \cdot \|\|_{\mathcal{H}}$	Norm associated with $\mathcal{H}$
$\|\| \cdot \|\|$	Euclidean norm
$\|\| \cdot \|\|_{HS}$	Hilbert-Schmidt operator norm
$\otimes$	Tensor product
$\boxtimes$	Kronecker product

List of Figures

List of Tables

Chapter 1

Introduction

In sample survey, estimation of different finite population parameters like, mean, median, variance, coefficient of variation, correlation and regression coefficients, interquartile range, measure of skewness, etc. was considered extensively in the past. However, comparison of different estimators of the same parameters has been limited. Also, asymptotic theory for several estimators has not been adequately developed in the available literature. One of the main objectives of this thesis is to compare various estimators of finite population parameters under different sampling designs (with no non-response) and superpopulation models, and to identify asymptotically efficient estimators among them. Another objective of this thesis is to understand the role of auxiliary information in the implementation of different sampling designs and in the construction of different estimators.

Suppose that $\mathcal{P}$=$\{1, 2, \ldots, N\}$ is a finite population of size N, s is a sample of size n $(< N)$ from $\mathcal{P}$, and $\mathcal{S}$ is the collection of all possible samples having size n. Then, a sampling design $P(s)$ is a probability distribution on $\mathcal{S}$ such that $0 \leq P(s) \leq 1$ for all $s \in \mathcal{S}$ and $\sum_{s\in\mathcal{S}} P(s)$=1. In this thesis, we consider sampling designs having fixed sample size. Now, suppose that $X_1, \ldots, X_N$ denote the population values on a positive real-valued size variable x. In sample survey, these population values are assumed to be known and utilized to implement sampling designs as well as to construct estimators. In this thesis, we consider the following sampling designs.

Simple random sampling without replacement (SRSWOR): In SRSWOR, n units are selected from the population $\mathcal{P}$ such that any subset of n units has the same probability =$\left({}^N C_n\right)^{-1}$ of being selected.

Rejective sampling design ([40]): Suppose that $\alpha_1, \ldots, \alpha_N$ are such that $\alpha_i > 0$ for any i=$1, \ldots, N$ and $\sum_{i=1}^{N} \alpha_i$=1. Then, in the rejective sampling design, n units are first drawn with replacement, where the i^{th} population unit is selected with probability =α_i, for i=$1, \ldots, N$. If any population unit is selected in the sample more than once, the sample is rejected and the entire procedure is repeated until n distinct units are selected in the sample. SRSWOR is a special case of rejective sampling design.

High entropy sampling design ([4]): A sampling design $P(s)$ is called high entropy sampling design if $D(P||R)=\sum_{s\in\mathcal{S}} P(s)\log\big(P(s)/R(s)\big) \rightarrow 0$ as $n, N \rightarrow \infty$ for some rejective sampling design $R(s)$. Some examples of high entropy sampling designs are SRSWOR, Lahiri–Midzuno–Sen (LMS) sampling design and Rao–Sampford (RS) sampling design.

LMS sampling design ([55], [57] and [75]): In LMS sampling design, the first unit is selected from $\mathcal{P}$, where the i^{th} population unit has the probability $=X_i/\sum_{j=1}^{N} X_j$ of being selected for i=$1,\ldots,N$. Following the first draw, $n-1$ units are selected from the remaining $N-1$ units in $\mathcal{P}$ using SRSWOR. One can show that in this sampling design, the selection probability of a sample is proportional to the total of the values of the size variable x for the sampled units.

RS sampling design ([4]): In RS sampling design, a population unit is first selected in such a way that the i^{th} population unit has the probability $=X_i/\sum_{j=1}^{N} X_j$ of being selected for i=$1,\ldots,N$. After replacing this unit back into the population, $n-1$ units are drawn with replacement, where the i^{th} population unit is selected with probability $=\lambda_i(1-\lambda_i)^{-1}/\sum_{i=1}^{N}\lambda_i(1-\lambda_i)^{-1}$ for λ_i=$nX_i/\sum_{i=1}^{N} X_i$. If any population unit is selected in the sample more than once, the sample is rejected and the entire procedure is repeated until n distinct units are selected in the sample.

πPS sampling design ([4] and [9]): A sampling design is called πPS (i.e., inclusion probability π proportional to size) sampling design if its inclusion probabilities $\{\pi_i\}_{i=1}^{N}$ satisfy the condition π_i=$nX_i/\sum_{j=1}^{N} X_j$ for i=$1,\ldots,N$. RS sampling design is an example of πPS sampling designs.

High entropy πPS (HEπPS) sampling design: A sampling design is called a HEπPS sampling design if it is a high entropy sampling design as well as a πPS sampling design. It was shown by [4] that RS sampling design is a HEπPS sampling design.

Rao–Hartley–Cochran (RHC) sampling design ([66]): In RHC sampling design, $\mathcal{P}$ is first divided randomly into n disjoint groups, say $\mathcal{P}_1,\ldots,\mathcal{P}_n$ of sizes $N_1,\ldots,N_n$, respectively, by taking a sample of N_1 units from N units using SRSWOR, then a sample of N_2 units from the remaining $N-N_1$ units using SRSWOR, then a sample of N_3 units from the remaining $N-N_1-N_2$ units using SRSWOR and so on. Following this random split, one unit is selected from each group independently. For each r=$1,\ldots,n$, the q^{th} unit from $\mathcal{P}_r$ is selected with probability $=X'_{qr}/\sum_{l=1}^{N_r} X'_{lr}$, where X'_{qr} is the x value of the q^{th} unit in $\mathcal{P}_r$.

Stratified multistage cluster sampling design ([35] and [77]): Suppose that the finite population $\mathcal{P}$ is divided into H strata or subpopulations, where stratum h consists of M_h clusters for h=$1,\ldots,H$. Further, the j^{th} cluster in stratum h consists of N_{hj} units for j=$1,\ldots,M_h$. For any given h=$1,\ldots,H$, j=$1,\ldots,M_h$ and l=$1,\ldots,N_{hj}$, we assume that the l^{th} unit from cluster j in stratum h is the i^{th} unit in the population $\mathcal{P}$, where i=$\sum_{h'=1}^{h}\sum_{j'=1}^{M_{h'}} N_{h'j'} - \sum_{j'=j}^{M_h} N_{hj'} + l$. In stratified multistage cluster sampling design with SRSWOR, first a sample s_h of m_h $(< M_h)$ clusters is selected from stratum h under SRSWOR for each h. Then, a sample s_{hj} of r_h $(< N_{hj})$ units is selected from j^{th} cluster in stratum h if it is selected in the sample of clusters s_h in the first stage for h=$1,\ldots,H$. Thus the resulting sample is s=$\cup_{1\leq h\leq H, j\in s_h} s_{hj}$. The samplings in the two stages are done independently across the strata and the clusters. Under the above sampling design, the inclusion probability of the i^{th} population unit is π_i=$m_h r_h/M_h N_{hj}$ if it belongs to the j^{th} cluster of stratum h. Note that stratified multistage cluster sampling design with SRSWOR becomes stratified sampling design with SRSWOR, when N_{hj}=1 for any h=$1,\ldots,H$ and j=$1,\ldots,M_h$. Also, note that stratified multistage cluster sampling design with SRSWOR becomes multistage cluster sampling design with SRSWOR, when H=1.

Suppose that (Y_i, Z_i) is the value of (y, z) for the i^{th} population unit, where y is a finite/infinite dimensional study variable, z is a finite dimensional covariate, and i=$1,\ldots,N$. In sample survey, the population total of z is assumed to be known. Moreover, z is used to construct different estimators (e.g., generalized regression (GREG) estimator). The variables (z, x) are also known as auxiliary variables. Sometimes, we consider superpopulation models, where $\{(Y_i, Z_i, X_i) : 1 \leqslant i \leqslant N\}$ are assumed to be independently and identically distributed (i.i.d.) random elements on $(\Omega, \mathcal{F}, \mathbf{P})$.

In Chapter 2 of this thesis, several well known estimators of finite population mean and its functions are investigated under some standard sampling designs. Such functions of mean include the variance, the correlation coefficient and the regression coefficient in the population as special cases. We compare the performance of these estimators under different sampling designs based on their asymptotic distributions. Equivalence classes of estimators under different sampling designs are constructed so that estimators in the same class have equivalent performance in terms of asymptotic mean squared errors (MSEs). Estimators in different asymptotic-MSE equivalence classes are then compared under some superpopulations satisfying linear models. It is shown that the pseudo empirical likelihood (PEML) estimator of the population mean under SRSWOR has the lowest asymptotic MSE among all the estimators under different sampling designs considered in this chapter. It is also shown that for the variance, the correlation coefficient and the regression coefficient of the population, the plug-in estimators based on the PEML estimator have the lowest asymptotic MSEs among all the estimators considered in this chapter under SRSWOR. On the other hand, for any HEπPS sampling design, which uses the auxiliary information, the plug-in estimators of those parameters based on the Hájek estimator have the lowest asymptotic MSEs among all the estimators considered in this chapter. This chapter is based on [29].

Asymptotic equivalence of some specific estimators of the population mean under some sampling designs was shown earlier in [22] and [74]. [22] established asymptotic equivalence of the PEML and the GREG estimators by showing that under some conditions on sampling designs, the difference between these two estimators is asymptotically negligible in probability. On the other hand, [74] showed that the ratio estimator has the same asymptotic distribution under SRSWOR and LMS sampling designs. The result that the difference between two estimators is asymptotically negligible in probability is a stronger result than the result that the asymptotic distributions of these estimators are the same. However, none of these authors constructed asymptotic-MSE equivalence classes, which consist of several estimators of a function of the population means under several sampling designs. Comparisons of some estimators of the population mean under some sampling designs were also carried out in [1], [2], [24]) and [64] based on asymptotic MSEs. However, the above comparisons included neither the PEML estimator nor HEπPS sampling designs.

In Chapter 3 of this thesis, the Horvitz–Thompson (HT), the RHC and the GREG estimators of the finite population mean are considered, when the observations are from an infinite dimensional space. We compare these estimators based on their asymptotic distributions under some commonly used sampling designs and some superpopulations satisfying linear regression models. We show that the GREG estimator is asymptotically at least as efficient as any of the other two estimators under different sampling designs considered in this chapter. Further, we show that the use of some well-known sampling designs utilizing auxiliary information may have an adverse effect on the performance of the GREG estimator, when the degree of heteroscedasticity present in linear regression models is not very large. On the other hand, the use of those sampling designs improves the performance of this estimator, when the degree of heteroscedasticity present in linear regression models is large. We develop methods for determining the degree of heteroscedasticity, which in turn determines the choice of appropriate sampling design to be used with the GREG estimator. We also investigate the consistency of the covariance operators of the above estimators. We carry out some numerical studies using real and synthetic data and our theoretical results are supported by the results obtained from those numerical studies. This chapter is based on [30].

[12], [13], [14], [16], [15], etc. investigated different asymptotic properties of the HT and the model assisted estimators of the finite population mean, when population observations are from $\mathcal{C}[0,T]$, the space of continuous functions defined on $[0,T]$. The model assisted estimator can be related to the GREG estimator considered earlier in [22] for finite dimensional data. All these authors carried out their investigation under sampling designs, which satisfy some regularity conditions. These sampling designs include SRSWOR, stratified sampling design with SRSWOR, rejective sampling designs, etc. However, none of the above authors compared the HT and the model assisted estimators.

In Chapter 4 of this thesis, the weak convergence of the quantile processes, which are constructed based on different estimators of the finite population quantiles, is shown under various well-known sampling designs based on a superpopulation model. The results related to the weak convergence of these quantile processes are applied to find asymptotic distributions of the smooth L-estimators and the estimators of smooth functions of finite population quantiles. Based on these asymptotic distributions, confidence intervals can be constructed for several finite population parameters like the median, the α-trimmed means, the interquartile range and the quantile based measure of skewness. Comparisons of various estimators are carried out based on their asymptotic distributions. We show that the use of the auxiliary information in the construction of the estimators sometimes has an adverse effect on the performances of the smooth L-estimators and the estimators of smooth functions of finite population quantiles under several sampling designs. Further, the performance of each of the above-mentioned estimators sometimes becomes worse under sampling designs, which use the auxiliary information, than their performances under SRSWOR. Moreover, it is shown that the sample median is more efficient than the sample mean under SRSWOR, whenever the finite population observations are generated from some symmetric and heavy-tailed superpopulation distributions with the same superpopulation mean and median. In the cases of symmetric superpopulation distributions with the same superpopulation mean and median, it is also shown that the GREG estimator of the finite population mean is more efficient than the sample median under SRSWOR, whenever there is substantial correlation present between the study and the auxiliary variables. This chapter is based on [31].

Strong and weak versions of Bahadur type representations of the sample quantile process were shown under simple random sampling in [78]. A quantile process based on the sample quantile, which is obtained by inverting the Hájek estimator of finite population distribution function, was constructed under high entropy sampling designs in [26]. However, there is no result available in the literature related to the weak convergence of quantile processes based on quantile estimators like the ratio, the difference, and the regression estimators, which are constructed using auxiliary information. There is also no available result related to the weak convergence of a quantile process under RHC and stratified multistage cluster sampling designs.

In sample survey, construction of several estimators (e.g., GREG and ratio estimators of the finite population mean) and derivation of their properties involve some form of regression analysis. Regression analysis also plays an important role for statistical analysis of estimators, when sampling designs (e.g., πPS, LMS and RHC) use auxiliary information. In Chapter 5 of this thesis, estimators obtained from least square (LS), asymmetric least square (ALS), truncated least square (TLS), least absolute deviation (LAD) and quantile regression (QR) are considered, when the sample observations are drawn from a finite population using some sampling design. The asymptotic distributions of these estimators are derived under different sampling designs based on a superpopulation model. Comparisons of several estimators are also carried out based on their asymptotic distributions. From these comparisons, it is shown that the use of the auxiliary information in the design stage sometimes has an adverse effect on the performances of different estimators of parameters in finite populations. It is also shown that the estimators of the finite population mean constructed based on quantile and TLS regression become more efficient than the GREG estimator under various sampling designs, whenever the finite population observations on the study variable are generated from some heavy-tailed distributions. This chapter is based on [32].

In the case of i.i.d. sample observations, [46], [39], [50], [51], [59], [33], [21], [49], [42], etc. studied several asymptotic properties of the estimators obtained from LS, ALS, TLS, LAD, QR, and other well-known regression methods. However, asymptotic behavior of the above-mentioned estimators have not been studied much, when the sample observations are drawn from a finite population using some sampling design. It becomes challenging to show Bahadur type representations and asymptotic normality of these estimators, when the sample observations may neither be independent nor identical.

In this thesis, several asymptotic results (e.g., central limit theorems for several estimators of the finite population mean, weak convergence of various empirical and quantile processes, etc.) are first derived under rejective sampling designs using consistency and asymptotic normality of the HT estimator under these sampling designs following the ideas in [40] and [4]. Then, these results are derived under high entropy sampling designs using the fact that any high entropy sampling design can be approximated by a rejective sampling design in Kullback-Liebler divergence. Thus high entropy sampling designs play an important role in the study of the asymptotic behaviour of several estimators, when the sample observations are neither independent nor identical.

Some of the major findings from the above-mentioned chapters are as follows. Given any sampling design, the estimators, which are constructed using the auxiliary information in the estimation stage, often become more efficient than the estimators, which are constructed without using any auxiliary information. However, each of the estimators considered in the above chapters usually becomes more efficient under SRSWOR than under RHC and HEπPS sampling designs, which use the auxiliary information in the design stage. This implies that although the use of the auxiliary information in the estimation stage usually improves the performance of different estimators, the use of the auxiliary information in the design stage often has adverse effect on the performance of these estimators. In practice, SRSWOR is easier to implement than the sampling designs that use the auxiliary information. Thus the above result is significant in view of selecting the appropriate sampling design. Further, for the finite population mean, the estimator constructed based on QR as well as TLS regression becomes more efficient than the GREG estimator constructed based on LS regression under several sampling designs, whenever the population values on the study variable are generated from heavy-tailed distributions.

Chapter 2

A comparison of estimators of mean and its functions in finite populations

Let y be a $\mathbb{R}^d$-valued $(d \geq 1)$ study variable. Throughout this chapter, we assume that the covariate z and the size variable x are the same. Recall from the introduction that (Y_i, X_i) denotes the value of (y, x) for the i^{th} population unit, where i=$1, \ldots, N$, and x is a positive real-valued size variable. Suppose that $\overline{Y}$=$\sum_{i=1}^{N} Y_i/N$ is the finite population mean of y. The HT estimator (see [44])) and the RHC (see [66]) estimator are commonly used design unbiased estimators of $\overline{Y}$. Other well-known estimators of $\overline{Y}$ are the Hájek estimator (see [41], [73], etc.), the ratio estimator (see [24]), the product estimator (see [24]), the GREG estimator (see [22]) and the PEML estimator (see [22]). However, these latter estimators are not always design unbiased. For the expressions of the above estimators, the reader is referred to Table 2.1 in Section 2.1 of this chapter. Now, consider the finite population parameter $g(\sum_{i=1}^{N} h(Y_i)/N)$. Here, h:$\mathbb{R}^d \rightarrow \mathbb{R}^p$ is a function with $p \geq 1$ and g:$\mathbb{R}^p \rightarrow \mathbb{R}$ is a continuously differentiable function. All vectors in Euclidean spaces will be taken as row vectors and superscript T will be used to denote their transpose. Examples of such a parameter are the variance, the correlation coefficient, the regression coefficient, etc. associated with a finite population. For simplicity, we shall often write $h(Y_i)$ as h_i. Then, $g(\overline{h})$=$g(\sum_{i=1}^{N} h_i/N)$ is estimated by plugging in the estimator $\hat{\overline{h}}$ of $\overline{h}$.

In this chapter, our objective is to find asymptotically efficient (in terms of asymptotic MSE) estimator of $g(\overline{h})$. In Section 2.1, based on the asymptotic distribution of the estimator of $g(\overline{h})$ under SRSWOR, LMS, HEπPS and RHC sampling designs (see the introduction), we construct asymptotic-MSE equivalence classes of estimators such that any two estimators in the same class have the same asymptotic MSE. We first consider the special case, when $g(\overline{h})=\overline{Y}$, and compare equivalence classes of estimators under superpopulations satisfying linear models in Section 2.2. Among different estimators under different sampling designs considered in this chapter, the PEML estimator of the population mean under SRSWOR turns out to be the estimator with the lowest asymptotic MSE. Also, the PEML estimator has the same asymptotic MSE under SRSWOR and LMS sampling design. Interestingly, we observe that the performance of the PEML estimator under RHC and any HEπPS sampling designs, which use auxiliary information, is worse than its performance under SRSWOR. Earlier, it was shown that the GREG estimator is asymptotically at least as efficient as the HT, the ratio and the product estimators under SRSWOR (see [24]). It follows from our analysis that the PEML estimator is asymptotically equivalent to the GREG estimator under all the sampling designs considered in this chapter.

[74] proved that the ratio estimator has the same asymptotic distribution under SRSWOR and LMS sampling design. [22] showed that under some conditions on the sampling design, the difference between the PEML and the GREG estimators is asymptotically negligible in probability, i.e., the PEML estimator is asymptotically equivalent to the GREG estimator. Among different sampling designs, SRSWOR and RHC sampling design satisfy these conditions. The result that the difference between two estimators is asymptotically negligible in probability is a stronger result than the result that the asymptotic distributions of these estimators are the same. However, none of the earlier authors constructed asymptotic-MSE equivalence classes, which consist of several estimators of a function of the population means under several sampling designs.

[64] compared the sample mean under the simple random sampling with replacement with the usual unbiased estimator of the population mean under the probability proportional to size sampling with replacement, when the study variable and the size variable are exactly linearly related. [2] compared the ratio estimator of the population mean under SRSWOR with the RHC estimator under RHC sampling design, when an approximate linear relationship holds between the study variable and the size variable. [1] carried out the comparison of the ratio estimator of the population mean under LMS sampling design and the RHC estimator under RHC sampling design, when the study variable and the size variable are approximately linearly related. It was shown that the GREG estimator of the population mean is asymptotically at least as efficient as the HT, the ratio and the product estimators under SRSWOR (see [24]). However, the above comparisons included neither the PEML estimator nor HEπPS sampling designs.

In Section 2.2, we also consider the cases, when $g(\overline{h})$ is the variance, the correlation coefficient and the regression coefficient in the population. Note that if the estimators of the population variance are constructed by plugging in the HT, the ratio, the product or the GREG estimators of the population means, then the estimators of the variance may become negative. One also faces problem with the plug-in estimators of the correlation coefficient and the regression coefficient as these estimators require estimators of population variances. On the other hand, if the estimators of the above-mentioned parameters are constructed by plugging in the Hájek or the PEML estimators of the population means, such a problem does not occur. Therefore, for these parameters, we compare only those equivalence classes, which contain the plug-in estimators based on the Hájek and the PEML estimators. From this comparison under superpopulations satisfying linear models, we once again conclude that for any of these parameters, the plug-in estimator based on the PEML estimator has asymptotically the lowest MSE among all the estimators considered in this chapter under SRSWOR as well as LMS sampling design. Moreover, under any HEπPS sampling design, which use the auxiliary information, the plug-in estimator based on the Hájek estimator has asymptotically the lowest MSE among all the estimators considered in this chapter.

Some empirical studies carried out in Section 2.3 using synthetic and real data demonstrate that the numerical and the theoretical results corroborate each other. In Section 2.4, the biased estimators considered in this chapter are compared empirically with their bias-corrected versions based on jackknifing in terms of MSE. We make some remarks on our major findings in Section 2.5. Proofs of the results are given in Sections 2.6 and 2.7.

2.1. Comparison of different estimators of $g(\overline{h})$

In this section, we first provide the expressions (see Table 2.1 below) of those estimators of $\overline{Y}$, which are considered in this chapter. In Table 2.1, π_i=$\sum_{s \ni U_i} P(s)$ is the inclusion probability of the i^{th} population unit, and G_i is the total of the x values of that randomly formed group from which the i^{th} population unit is selected in the sample by RHC sampling design (see [66] and the introduction). In the case of the GREG estimator, $\hat{\overline{Y}}_*$=$\sum_{i \in s} d(i,s) Y_i / \sum_{i \in s} d(i,s)$, $\hat{\overline{X}}_*$=$\sum_{i \in s} d(i,s) X_i / \sum_{i \in s} d(i,s)$ and $\hat{\beta}$=$\sum_{i \in s} d(i,s)(Y_i - \hat{\overline{Y}}_*)(X_i - \hat{\overline{X}}_*) / \sum_{i \in s} d(i,s)(X_i - \hat{\overline{X}}_*)^2$, where $\{d(i,s) : i \in s\}$ are sampling design weights. Finally, the c_i's (> 0) in the PEML estimator are obtained by maximizing $\sum_{i \in s} d(i,s) \log(c_i)$ subject to $\sum_{i \in s} c_i$=1 and $\sum_{i \in s} c_i (X_i - \overline{X})$=0. Following [22], we consider both the GREG and the PEML estimators

TABLE 2.1: Estimators of $\overline{Y}$.

Estimator	Expression
HT	$\hat{\overline{Y}}_{HT}$=$\sum_{i\in s}(N\pi_i)^{-1}Y_i$
RHC	$\hat{\overline{Y}}_{RHC}$=$\sum_{i\in s}(NX_i)^{-1}G_iY_i$
Hájek	$\hat{\overline{Y}}_{H}$=$\sum_{i\in s}\pi_i^{-1}Y_i/\sum_{i\in s}\pi_i^{-1}$
Ratio	$\hat{\overline{Y}}_{RA}$=$(\sum_{i\in s}\pi_i^{-1}Y_i/\sum_{i\in s}\pi_i^{-1}X_i)\overline{X}$
Product	$\hat{\overline{Y}}_{PR}$=$\sum_{i\in s}(N\pi_i)^{-1}Y_i\sum_{i\in s}(N\pi_i)^{-1}X_i/\overline{X}$
GREG	$\hat{\overline{Y}}_{GREG}$=$\hat{\overline{Y}}_* + \hat{\beta}(\overline{X}-\hat{\overline{X}}_*)$
PEML	$\hat{\overline{Y}}_{PEML}$=$\sum_{i\in s}c_iY_i$

with $d(i,s)$=$(N\pi_i)^{-1}$ under SRSWOR, LMS and any HEπPS sampling designs, and with $d(i,s)$=$(NX_i)^{-1}G_i$ under RHC sampling design.

We compare the estimators of $g(\overline{h})$, which are obtained by plugging in the estimators of $\overline{h}$ mentioned in Table 2.2 below. The expressions of these estimators of $\bar{h}$ are the same as the expressions of the estimators of $\overline{Y}$ (see Table 2.1) with Y_i replaced by $h(Y_i)$. First, we find

TABLE 2.2: Estimators of $\overline{h}$.

Sampling designs	Estimators
SRSWOR	HT (which coincides with Hájek estimator), ratio, product, GREG and PEML estimators
LMS	HT, Hájek, ratio, product, GREG and PEML estimators
HEπPS	HT (which coincides with ratio and product estimators), Hájek, GREG and PEML estimators
RHC	RHC, GREG and PEML estimators

equivalence classes of estimators of $g(\overline{h})$ such that any two estimators in the same class are asymptotically normal with the same mean $g(\overline{h})$ and same variance.

We define our asymptotic framework as follows. Let $\{\mathcal{P}_\nu\}$ be a sequence of populations with N_ν, $n_\nu \to \infty$ as $\nu \to \infty$ (see [48], [85], [26], [7], [43] and references therein), where N_ν and n_ν are, respectively, the population size and the sample size corresponding to the ν^{th} population. Henceforth, we shall suppress the subscript ν that tends to ∞ for the sake of simplicity. Throughout this chapter, we consider the following assumption (cf. Assumption 1 in [12], A4 in [25], A1 in [16] A4 in [26] and (HT3) in [7])

Assumption 2.1.1. *$n/N \to \lambda$ as $\nu \to \infty$, where $0 \leq \lambda < 1$.*

Before we state the main results, let us discuss some assumptions on $\{(X_i, h_i) : 1 \leq i \leq N\}$ (recall that h_i=$h(Y_i)$). Note that in any finite dimensional Euclidean space, we consider the Euclidean norm and denote it by $||\cdot||$.

Assumption 2.1.2. *$\{P_\nu\}$ is such that $\sum_{i=1}^N ||h_i||^4/N$=$O(1)$ and $\sum_{i=1}^N X_i^4/N$ =$O(1)$ as $\nu \to \infty$. Further, $\lim_{\nu\to\infty} \overline{h}$ exists, and $\overline{X}$=$\sum_{i=1}^N X_i/N$ and S_x^2= $\sum_{i=1}^N(X_i - \overline{X})^2/N$ are bounded away from 0 as $\nu \to \infty$. Moreover, $\nabla g(\mu_0) \neq 0$, where μ_0=$\lim_{\nu\to\infty} \overline{h}$ and ∇g is the gradient of g.*

Assumption 2.1.3. *$\max_{1\leq i\leq N} X_i/\min_{1\leq i\leq N} X_i$=$O(1)$ as $\nu \to \infty$.*

Let $\mathbf{V}_i$ be one of h_i, $h_i - \overline{h}$, $h_i - \overline{h}X_i/\overline{X}$, $h_i + \overline{h}X_i/\overline{X}$ and $h_i - \overline{h} - S_{xh}(X_i - \overline{X})/S_x^2$ for i=$1,\ldots,N$, $\overline{h}$=$\sum_{i=1}^N h_i/N$ and S_{xh}=$\sum_{i=1}^N X_i h_i/N - \overline{h}\ \overline{X}$. Define $\mathbf{T}_V$=$\sum_{i=1}^N \mathbf{V}_i(1-\pi_i)/\sum_{i=1}^N \pi_i(1-\pi_i)$, where π_i is the inclusion probability of the i^{th} population unit. Also, in the case of RHC sampling design, define $\overline{\mathbf{V}}$=$\sum_{i=1}^N \mathbf{V}_i/N$, $\overline{X}$=$\sum_{i=1}^N X_i/N$ and γ=$\sum_{r=1}^n \tilde{N}_r(\tilde{N}_r - 1)/N(N-1)$, where $\tilde{N}_r$ is the size of the r^{th} group formed randomly in RHC sampling design, r=$1,\ldots,n$. It follows from Lemma 2.7.5 in Section 2.7 that $n\gamma \to c$ as $\nu \to \infty$ for some $c \geq 1-\lambda$. Now, we state the following assumptions on the population values and the sampling designs.

Assumption 2.1.4. *$P(s)$ is such that $nN^{-2}\sum_{i=1}^N(\boldsymbol{V}_i - \boldsymbol{T}_V\pi_i)^T(\boldsymbol{V}_i - \boldsymbol{T}_V\pi_i)(\pi_i^{-1} - 1)$ converges to some positive definite (p.d.) matrix as $\nu \to \infty$.*

Assumption 2.1.5. *$n\gamma\overline{X}N^{-1}\sum_{i=1}^N(\boldsymbol{V}_i - X_i\overline{\boldsymbol{V}}/\overline{X})^T(\boldsymbol{V}_i - X_i\overline{\boldsymbol{V}}/\overline{X})/X_i$ converges to some p.d. matrix as $\nu \to \infty$.*

Similar assumptions like Assumptions 2.1.2, 2.1.4 and 2.1.5 are often used in sample survey literature (see Assumption 3 in [12], A3 and A6 in both [25] and [26], (HT2) in [7], and F2 and F3 in [43]). Assumptions 2.1.2 and 2.1.5 hold (*almost surely*), whenever $\{(X_i, h_i) : 1 \leq i \leq N\}$ are generated from a superpopulation model satisfying appropriate moment conditions (see Lemma 2.7.8 in Section 2.7). The condition $\sum_{i=1}^N ||h_i||^4/N$=$O(1)$ holds, when h is a bounded function (e.g., $h(y)$=y and y is a binary study variable). Assumption 2.1.3 implies that the variation in the population values $X_1,\ldots,X_N$ cannot be too large. Under any πPS sampling design, Assumption

2.1.3 is equivalent to the condition that $L \leq N\pi_i/n \leq L'$ for some constants $L, L' > 0$, any i=$1, \ldots, N$ and all sufficiently large $\nu \geq 1$. This latter condition was considered earlier in the literature (see (C1) in [7] and Assumption 2–*(i)* in [85]). Assumption 2.1.3 holds (*almost surely*), when $\{X_i\}_{i=1}^N$ are generated from a superpopulation distribution and the support of the distribution of X_i is bounded away from 0 and ∞. Assumption 2.1.4 holds (*almost surely*) for SRSWOR, LMS and any πPS sampling designs under appropriate superpopulation models (see Lemma 2.7.8 in Section 2.7). In the context of the RHC sampling design, we also consider the following assumption.

Assumption 2.1.6. *For the RHC sampling design, $\{\tilde{N}_r\}_{r=1}^n$ are such that*

$$\tilde{N}_r = \begin{cases} N/n, \textit{ for } r = 1, \cdots, n, \textit{ when } N/n \textit{ is an integer}, \\ \lfloor N/n \rfloor, \textit{ for } r = 1, \cdots, k, \textit{ and} \\ \lfloor N/n \rfloor + 1, \textit{ for } r = k+1, \cdots, n, \textit{ when } N/n \textit{ is not an integer}, \end{cases} \tag{2.1.1}$$

where k is such that $\sum_{r=1}^n \tilde{N}_r$=N. Here, $\lfloor N/n \rfloor$ is the integer part of N/n.

[66] showed that this choice of $\{\tilde{N}_r\}_{r=1}^n$ minimizes the variance of the RHC estimator. Assumptions 2.1.1–2.1.6 are used to prove some technical results (see Lemmas 2.7.1–2.7.7 in Section 2.7) under LMS, HEπPS and RHC sampling designs, which will be required to construct asymptotic-MSE equivalence classes of estimators for $g(\overline{h})$ under different sampling designs considered in this chapter. Now, we state the following theorems.

Theorem 2.1.1. *Suppose that Assumptions 2.1.1–2.1.4 hold. Then, classes $1, 2, 3$ and 4 in Table 2.3 describe asymptotic-MSE equivalence classes of estimators for $g(\overline{h})$ under SRSWOR and LMS sampling design.*

For next two theorems, we assume that $n \max_{1 \leq i \leq N} X_i / \sum_{i=1}^N X_i < 1$. Note that this condition is required to hold for any without replacement πPS sampling design.

Theorem 2.1.2. *(i) If Assumptions 2.1.1–2.1.4 hold, then classes $5, 6$ and 7 in Table 2.3 describe asymptotic-MSE equivalence classes of estimators for $g(\overline{h})$ under any HEπPS sampling design. (ii) Under RHC sampling design, if Assumptions 2.1.1–2.1.3, 2.1.5 and 2.1.6 hold, then classes 8 and 9 in Table 2.3 describe asymptotic-MSE equivalence classes of estimators for $g(\overline{h})$.*

TABLE 2.3: Estimators of $\overline{h}$ based on which asymptotic-MSE equivalence classes of estimators for $g(\overline{h})$ are formed.

	Sampling design			
Equivalence classes	SRSWOR	LMS	HEπPS	RHC
class 1	GREG and PEML	GREG and PEML		
class 2	[1] HT	HT and Hájek		
class 3	Ratio	Ratio		
class 4	Product	Product		
class 5			GREG and PEML	
class 6			[2] HT	
class 7			Hájek	
class 8				GREG and PEML
class 9				RHC

[1] The HT and the Hájek estimators coincide under SRSWOR.
[2] The HT, the ratio and the product estimators coincide under HEπPS sampling designs.

Remark 2.1.1. *It is to be noted that if Assumptions 2.1.2–2.1.4 hold, and 2.1.1 holds with* λ=0, *then in Table 2.3, class* 8 *is merged with class* 5*, and class* 9 *is merged with class* 6*. For details, see Section 2.6.*

Next, suppose that W_i=$\nabla g(\overline{h})h_i^T$ for i=$1,\ldots,N$, $\overline{W}$=$\sum_{i=1}^N W_i/N$, S_{xw}= $\sum_{i=1}^N W_iX_i/N -\overline{W}\ \overline{X}$, S_w^2= $\sum_{i=1}^N W_i^2/N -\overline{W}^2$, S_x^2=$\sum_{i=1}^N X_i^2/N - \overline{X}^2$ and ϕ=$\overline{X} - (n/N)\sum_{i=1}^N X_i^2/N\overline{X}$. Now, we state the following theorem.

Theorem 2.1.3. *Suppose that the assumptions of Theorems 2.1.1 and 2.1.2 hold. Then, Table 2.4 gives the expressions of asymptotic MSEs,* $\Delta_1^2,\ldots,\Delta_9^2$*, of estimators in asymptotic-MSE equivalence classes* $1,\ldots,9$ *in Table 2.3, respectively.*

Remark 2.1.2. *It can be shown in a straightforward way from Table 2.4 that* $\Delta_1^2 \leq \Delta_i^2$ *for* i=$2,3$ *and* 4*. Thus, both the plug-in estimators of* $g(\overline{h})$ *that are based on the GREG and the PEML estimators are asymptotically as good as, if not better than, the plug-in estimators based on the HT (which coincides with the Hájek estimator), the ratio and the product estimators under*

SRSWOR, and the plug-in estimators based on the HT, the Hájek, the ratio and the product estimators under LMS sampling design.

TABLE 2.4: Asymptotic variances of estimators for $g(\overline{h})$ (note that for simplifying notations, the subscript ν is dropped from the expressions on which limits are taken).

$\Delta_1^2=(1-\lambda)\lim_{\nu\to\infty}\left(S_w^2-(S_{xw}/S_x)^2\right)$
$\Delta_2^2=(1-\lambda)\lim_{\nu\to\infty} S_w^2$
$\Delta_3^2=(1-\lambda)\lim_{\nu\to\infty}\left(S_w^2-2\overline{W}S_{xw}/\overline{X}+\left(\overline{W}/\overline{X}\right)^2S_x^2\right)$
$\Delta_4^2=(1-\lambda)\lim_{\nu\to\infty}\left(S_w^2+2\overline{W}S_{xw}/\overline{X}+\left(\overline{W}/\overline{X}\right)^2S_x^2\right)$
$\Delta_5^2=\lim_{\nu\to\infty}(1/N)\sum_{i=1}^N\left(W_i-\overline{W}-(S_{xw}/S_x^2)(X_i-\overline{X})\right)^2\times$ $\left((\overline{X}/X_i)-(n/N)\right)$
$\Delta_6^2=\lim_{\nu\to\infty}(1/N)\sum_{i=1}^N\left\{W_i+\phi^{-1}\overline{X}^{-1}X_i\left((n/N)\sum_{i=1}^N W_iX_i/N-\overline{W}\,\overline{X}\right)\right\}^2\times$ $\left\{(\overline{X}/X_i)-(n/N)\right\}$
$\Delta_7^2=\lim_{\nu\to\infty}(1/N)\sum_{i=1}^N\left(W_i-\overline{W}+(n/N\phi\overline{X})X_iS_{xw}\right)^2\times$ $\left((\overline{X}/X_i)-(n/N)\right)$
$\Delta_8^2=\lim_{\nu\to\infty} n\gamma(\overline{X}/N)\sum_{i=1}^N\left(W_i-\overline{W}-(S_{xw}/S_x^2)(X_i-\overline{X})\right)^2/X_i$
$\Delta_9^2=\lim_{\nu\to\infty} n\gamma\left((\overline{X}/N)\sum_{i=1}^N W_i^2/X_i-\overline{W}^2\right)$

Let us now consider some examples of $g(\overline{h})$ in Table 2.5 below. Conclusions of Theorems

TABLE 2.5: Examples of $g(\overline{h})$.

Parameter	d	p	h	g
Mean	1	1	$h(y)=y$	$g(s)=s$
Variance	1	2	$h(y)=(y^2,y)$	$g(s_1,s_2)=s_1-s_2^2$
Correlation coefficient	2	5	$h(z_1,z_2)=(z_1,z_2,$ $z_1^2,z_2^2,z_1z_2)$	$g(s_1,s_2,s_3,s_4,s_5)=(s_5-s_1s_2)/$ $((s_3-s_1^2)(s_4-s_2^2))^{1/2}$
Regression coefficient	2	4	$h(z_1,z_2)=(z_1,z_2,$ $z_2^2,z_1z_2)$	$g(s_1,s_2,s_3,s_4,s_5)=$ $(s_4-s_1s_2)/(s_3-s_2^2)$

2.1.1–2.1.3, and Remarks 2.1.1 and 2.1.2 hold for all of the above parameters. Here, we recall from the 5^{th} paragraph in the beginning of this chapter that for the variance, the correlation coefficient and the regression coefficient, we consider only the plug-in estimators that are based on the Hájek and the PEML estimators.

2.2. Comparison of estimators under superpopulation models

In this section, we derive asymptotically efficient estimators for the mean, the variance, the correlation coefficient and the regression coefficient under superpopulations satisfying linear regression models. Earlier, [64] [58], [2], [1] and [24] used the linear relationship between the Y_i's and the X_i's for comparing different estimators of the mean. However, they did not use any probability distribution for the (Y_i, X_i)'s. Subsequently, [65], [36], [19], [7], [63], etc. considered the linear relationship between the Y_i's and the X_i's and a probability distribution for the (Y_i, X_i)'s for constructing different estimators and studying their behavior. However, the problem of finding asymptotically the most efficient estimator for the mean among a large class of estimators as considered in this chapter was not done earlier in the literature. Also, large sample comparisons of the plug-in estimators of the variance, the correlation coefficient and the regression coefficient considered in this chapter were not carried out in the earlier literature. As mentioned in the introduction, let us assume that $\{(Y_i, X_i) : 1 \leq i \leq N\}$ are i.i.d. random vectors defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Without any loss of generality, for convenience, we take σ_x^2=$E_{\mathbf{P}}(X_i - E_{\mathbf{P}}(X_i))^2$ =1. This might require rescaling the variable x. Here, $E_{\mathbf{P}}$ denotes the expectation with respect to the probability measure $\mathbf{P}$. Recall that the population values $X_1, \ldots, X_N$ are used to implement some of the sampling designs like LMS, RHC, HEπPS, etc. In such a case, we consider a function $P(s, \omega)$ on $\mathcal{S} \times \Omega$ so that $P(s, \cdot)$ is a random variable on Ω for each $s \in \mathcal{S}$, and $P(\cdot, \omega)$ is a probability distribution on $\mathcal{S}$ for each $\omega \in \Omega$ (see [7]). Note that $P(s, \omega)$ is the sampling design for any fixed ω in this case. Then, the Δ_j^2's in Table 2.4 can be expressed in terms of superpopulation moments of $(h(Y_i), X_i)$ by strong law of large numbers (SLLN). In that case, we can easily compare different classes of estimators in Table 2.3 under linear models. Let us first state the following assumption on superpopulation distribution $\mathbf{P}$.

Assumption 2.2.1. *$X_i \leq b$ a.s. $[\mathbb{P}]$ for some $b > 0$, $E_{\mathbb{P}}(X_i)^{-2} < \infty$, and $\max_{1 \leq i \leq N} X_i / \min_{1 \leq i \leq N} X_i$=$O(1)$ as $\nu \to \infty$ a.s. $[\mathbb{P}]$. Also, the support of the distribution of $(h(Y_i), X_i)$ is not a subset of a hyperplane in $\mathbb{R}^{p+1}$.*

The condition, $X_i \leq b$ *a.s.* $[\mathbb{P}]$ for some $b > 0$, in Assumption 2.2.1 and Assumption 2.1.1 along with $0 \leq \lambda < E_{\mathbb{P}}(X_i)/b$ ensure that $n \max_{1 \leq i \leq N} X_i / \sum_{i=1}^{N} X_i < 1$ for all sufficiently large ν *a.s.* $[\mathbb{P}]$, which is required for implementing a πPS sampling design. On the other hand, the condition, $\max_{1 \leq i \leq N} X_i / \min_{1 \leq i \leq N} X_i$ =$O(1)$ as $\nu \to \infty$ *a.s.* $[\mathbb{P}]$, in Assumption 2.2.1 implies that Assumption 2.1.3 holds *a.s.* $[\mathbb{P}]$. Further, Assumption 2.2.1 ensures that Assumption 2.1.5 holds *a.s.* $[\mathbb{P}]$ (see Lemma 2.7.3 in Section 2.7). Assumption 2.2.1 also ensures that

Assumption 2.1.4 holds under LMS and any πPS sampling designs *a.s.* $[\mathbb{P}]$ (see Lemma 2.7.3 in Section 2.7).

Let us first consider the case, when $g(\overline{h})$ is the mean of y (see the 2^{nd} row in Table 2.5) Further, suppose that Y_i=$\alpha + \beta X_i + \epsilon_i$ for $\alpha, \beta \in \mathbb{R}$ and i=$1, \ldots, N$, where $\{\epsilon_i\}_{i=1}^{N}$ are i.i.d. random variables and are independent of $\{X_i\}_{i=1}^{N}$ with $E_{\mathbf{P}}(\epsilon_i)$=0 and $E_{\mathbf{P}}(\epsilon_i)^4 < \infty$. Then, we have the following theorem.

Theorem 2.2.1. *Suppose that Assumption 2.1.1 holds with $0 \leq \lambda < E_{\mathbb{P}}(X_i)/b$, and Assumptions 2.1.6 and 2.2.1 hold. Then, a.s. $[\mathbb{P}]$, the PEML estimator under SRSWOR as well as LMS sampling design has the lowest asymptotic MSE among all the estimators of the population mean under different sampling designs considered in this chapter.*

Remark 2.2.1. *Note that for SRSWOR, the PEML estimator of the population mean has the lowest asymptotic MSE among all the estimators considered in this chapter a.s. $[\mathbb{P}]$, when Assumption 2.1.1 holds with $0 \leq \lambda < 1$, and Assumptions 2.1.6 and 2.2.1 hold (see the proof of Theorem 2.2.1).*

Theorem 2.2.2. *Suppose that Assumption 2.1.1 holds with $0 \leq \lambda < E_{\mathbb{P}}(X_i)/b$, and Assumptions 2.1.6 and 2.2.1 hold. Then, a.s. [**P**], the performance of the PEML estimator of the population mean under RHC and any HEπPS sampling designs, which use auxiliary information is worse than its performance under SRSWOR.*

Recall from the 5^{th} paragraph in the beginning of this chapter that for the variance, the correlation coefficient and the regression coefficient, we compare only those equivalence classes, which contain the plug-in estimators based on the Hájek and the PEML estimators. We first state the following assumption.

Assumption 2.2.2. $\xi > 2\max\{\mu_1, \mu_{-1}/(\mu_1\mu_{-1} - 1)\}$, *where ξ=$\mu_3 - \mu_2\mu_1$ is the covariance between X_i^2 and X_i, and μ_j=$E_{\boldsymbol{P}}(X_i)^j$, j=$-1, 1, 2, 3$.*

The above assumption is used to prove part *(ii)* in each of Theorems 2.2.3 and 2.2.4. This condition holds when the X_i's follow well-known distributions like Gamma (with shape parameter value larger than 1 and any scale parameter value), Beta (with the second shape parameter value greater than the first shape parameter value and the first shape parameter value larger than 1), Pareto (with shape parameter value lying in the interval $(3, (5+\sqrt{17})/2)$ and any scale parameter value), Log-normal (with any parameter value) and Weibull (with shape parameter value lying in the interval $(1, 3.6)$ and any scale parameter value). Now, consider the case, when $g(\overline{h})$ is the variance of y (see the 3^{rd} row in Table 2.5). Recall the linear model Y_i=$\alpha + \beta X_i + \epsilon_i$ from above and assume that $E_{\mathbb{P}}(\epsilon_i)^8 < \infty$. Then, we have the following theorem. Now, consider the case, when $g(\overline{h})$ is the variance of y, i.e., d=1, p=2, $h(y)$=(y, y^2), and $g(s_1, s_2)$=$s_2 - s_1^2$. Recall the linear model Y_i=$\alpha + \beta X_i + \epsilon_i$ from above and assume that $E_{\mathbf{P}}(\epsilon_i)^8 < \infty$. Then, we have the following theorem.

Theorem 2.2.3. *(i) Let us first consider SRSWOR and LMS sampling design and suppose that Assumptions 2.1.1 and 2.2.1 hold. Then, a.s.* $[\mathbb{P}]$*, the plug-in estimator of the population variance based on the PEML estimator has the lowest asymptotic MSE among all the estimators considered in this chapter.*
(ii) Next consider any HEπPS sampling design and suppose that Assumption 2.1.1 holds with $0 \leq \lambda < E_{\mathbb{P}}(X_i)/b$*, and Assumptions 2.2.1 and 2.2.2 hold. Then, a.s.* $[\mathbb{P}]$*, the plug-in estimator of the population variance based on the Hájek estimator has the lowest asymptotic MSE among all the estimators considered in this chapter.*

Next, suppose that y=$(z_1, z_2) \in \mathbb{R}^2$ and consider the case, when $g(\overline{h})$ is the correlation coefficient between z_1 and z_2 (see the 4^{th} row in Table 2.5). Let us also consider the case, when $g(\overline{h})$ is the regression coefficient of z_1 on z_2 (see the 5^{th} row in Table 2.5). Further, suppose that Y_i=$\alpha + \beta X_i + \epsilon_i$ for Y_i=(Z_{1i}, Z_{2i}), $\alpha, \beta \in \mathbb{R}^2$ and i=$1, \ldots, N$, where $\{\epsilon_i\}_{i=1}^N$ are i.i.d. random vectors in $\mathbb{R}^2$ independent of $\{X_i\}_{i=1}^N$ with $E_{\mathbf{P}}(\epsilon_i)$=0 and $E_{\mathbf{P}}||\epsilon_i||^8 < \infty$. Then, we have the following theorem.

Theorem 2.2.4. *(i) Let us first consider SRSWOR and LMS sampling design and suppose that Assumptions 2.1.1 and 2.2.1 hold. Then, a.s.* $[\mathbb{P}]$*, the plug-in estimator of each of the correlation and the regression coefficients in the population based on the PEML estimator has the lowest asymptotic MSE among all the estimators considered in this chapter.*
(ii) Next consider any HEπPS sampling design and suppose that Assumption 2.1.1 holds with $0 \leq \lambda < E_{\mathbb{P}}(X_i)/b$*, and Assumptions 2.2.1 and 2.2.2 hold. Then, a.s.* $[\mathbb{P}]$*, the plug-in estimator of each of the above parameters based on the Hájek estimator has the lowest asymptotic MSE among all the estimators considered in this chapter.*

2.3. Data analysis

In this section, we intend to carry out an empirical comparison of the estimators of the mean, the variance, the correlation coefficient and the regression coefficient, which are discussed in this chapter, based on both real and synthetic data. Recall that for the above parameters, we have considered several estimators and sampling designs, and conducted a theoretical comparison of those estimators in Sections 2.1 and 2.2. For empirical comparison, we exclude some of the estimators considered in theoretical comparison so that the results of the comparison become concise and comprehensive. The reasons for excluding those estimators are given below.

(i) Since the GREG estimator is well-known to be asymptotically better than the HT, the ratio and the product estimators under SRSWOR (see [24]), we exclude these latter estimators under SRSWOR.

(ii) Since the MSEs of the estimators under LMS sampling design become very close to the MSEs of the same estimators under SRSWOR as expected from Theorem 2.1.1, we do not report these results under LMS sampling design. Moreover, SRSWOR is a simpler and more commonly used sampling design than LMS sampling design.

Thus we consider the estimators mentioned in Table 2.6 below for the empirical comparison. Recall from Table 2.2 that the HT, the ratio and the product estimators of the mean coincide

TABLE 2.6: Estimators considered for the empirical comparison.

Parameters	Estimators
Mean	GREG and PEML estimators under SRSWOR; HT, Hájek, GREG and PEML estimators under [3]RS sampling design; and RHC and GREG estimators under RHC sampling design
Variance, correlation coefficient and regression coefficient	Obtained by plugging in Hájek and PEML estimators under each of SRSWOR and [1]RS sampling design, and PEML estimator under RHC sampling design

[3] We consider RS sampling design since it is a HEπPS sampling design, and it is easier to implement than other HEπPS sampling designs.

under any HEπPS sampling design. We draw I=1000 samples each of sizes n=75, 100 and 125 using sampling designs mentioned in Table 2.6. We use the R software for drawing samples as well as computing different estimators. For RS sampling design, we use the 'pps' package in R, and for the PEML estimator, we use R codes in [87]. Two estimators $g(\hat{\overline{h}}_1)$ and $g(\hat{\overline{h}}_2)$ of $g(\overline{h})$ under sampling designs $P_1(s)$ and $P_2(s)$, respectively, are compared empirically by means of the relative efficiency defined as

$$RE(g(\hat{\overline{h}}_1), P_1|g(\hat{\overline{h}}_2), P_2) = MSE_{P_2}(g(\hat{\overline{h}}_2))/MSE_{P_1}(g(\hat{\overline{h}}_1)),$$

where $MSE_{P_j}(g(\hat{\overline{h}}_j))$=$I^{-1}\sum_{l=1}^{I}(g(\hat{\overline{h}}_{jl}) - g(\overline{h}_0))^2$ is the empirical MSE of $g(\hat{\overline{h}}_j)$ under $P_j(s)$, j=1, 2. Here, $\hat{\overline{h}}_{jl}$ is the estimate of $\overline{h}$ based on the j^{th} estimator and the l^{th} sample, and $g(\overline{h}_0)$ is the true value of the prameter $g(\overline{h})$, j=1, 2, l=1, . . . , I. $g(\hat{\overline{h}}_1)$ under $P_1(s)$ will be more efficient than $g(\hat{\overline{h}}_2)$ under $P_2(s)$ if $RE(g(\hat{\overline{h}}_1), P_1|g(\hat{\overline{h}}_2), P_2) > 1$.

Next, for each of the parameters considered in this section, we compare average lengths of asymptotically 95% confidence intervals (CIs) constructed based on several estimators used in this section. In order to construct asymptotically 95% CIs, we need an estimator of the asymptotic MSE of $\sqrt{n}(g(\hat{\overline{h}}) - g(\overline{h}))$. If we consider SRSWOR or RS sampling design, it follows from the proofs of Theorems 2.1.1 and 2.1.2 that the asymptotic MSE of $\sqrt{n}(g(\hat{\overline{h}}) - g(\overline{h}))$ is $\tilde{\Delta}_1^2$=$\lim_{\nu\to\infty} nN^{-2}\nabla g(\overline{h})\sum_{i=1}^{N}(\mathbf{V}_i - \mathbf{T}_V\pi_i)^T(\mathbf{V}_i - \mathbf{T}_V\pi_i)(\pi_i^{-1} - 1)\nabla g(\overline{h})^T$, where $\mathbf{T}_V$=$\sum_{i=1}^{N}\mathbf{V}_i(1-\pi_i)/\sum_{i=1}^{N}\pi_i(1-\pi_i)$. Moreover, $\mathbf{V}_i$ is h_i or $h_i-\overline{h}$ or $h_i-\overline{h}-S_{xh}(X_i-\overline{X})/S_x^2$ if $\hat{\overline{h}}$ is $\hat{\overline{h}}_{HT}$ or $\hat{\overline{h}}_H$ or $\hat{\overline{h}}_{PEML}$ (as well as $\hat{\overline{h}}_{GREG}$) with $d(i,s)$=$(N\pi_i)^{-1}$, respectively. Recall from the paragraph following Assumption 2.1.3 that S_{xh}=$\sum_{i=1}^{N}X_ih_i/N - \overline{X}\,\overline{h}$. Following the idea of [16], we estimate $\tilde{\Delta}_1^2$ by

$$\hat{\Delta}_1^2 = nN^{-2}\nabla g(\hat{\overline{h}})\sum_{i\in s}(\hat{\mathbf{V}}_i - \hat{T}_v\pi_i)^T(\hat{\mathbf{V}}_i - \hat{T}_v\pi_i)(\pi_i^{-1} - 1)\pi_i^{-1}\nabla g(\hat{\overline{h}})^T, \tag{2.3.1}$$

where $\hat{T}_v$=$\sum_{i\in s}\hat{\mathbf{V}}_i(\pi_i^{-1} - 1)/\sum_{i\in s}(1 - \pi_i)$, $\hat{\overline{h}}$=$\hat{\overline{h}}_{HT}$ in the case of the mean, the variance and the regression coefficient, and $\hat{\overline{h}}$=$\hat{\overline{h}}_H$ in the case of the correlation coefficient. Here, $\hat{\mathbf{V}}_i$ is h_i or $h_i - \hat{\overline{h}}_{HT}$ or $h_i - \hat{\overline{h}}_{HT} - \hat{S}_{xh,1}(X_i - \hat{\overline{X}}_{HT})/\hat{S}_{x,1}^2$ if $\hat{\overline{h}}$ is $\hat{\overline{h}}_{HT}$ or $\hat{\overline{h}}_H$ or $\hat{\overline{h}}_{PEML}$ (as well as $\hat{\overline{h}}_{GREG}$) with $d(i,s)$=$(N\pi_i)^{-1}$. Further, $\hat{S}_{xh,1}$=$\sum_{i\in s}(N\pi_i)^{-1}X_ih_i - \hat{\overline{X}}_{HT}\hat{\overline{h}}_{HT}$ and $\hat{S}_{x,1}^2$=$\sum_{i\in s}(N\pi_i)^{-1}X_i^2 - \hat{\overline{X}}_{HT}^2$. We estimate $\overline{h}$ in $\nabla g(\overline{h})$ by $\hat{\overline{h}}_{HT}$ in the case of the mean, the variance and the regression coefficient because $\hat{\overline{h}}_{HT}$ is an unbiased estimator and it is easier to compute than the other estimators of $\overline{h}$ considered in this chapter. On the other hand, different estimators of the correlation coefficient that are considered in this chapter may become undefined if we estimate $\overline{h}$ by any estimator other than $\hat{\overline{h}}_H$ and $\hat{\overline{h}}_{PEML}$ (see the 5^{th} paragraph in the beginning of this chapter). In this case, we choose $\hat{\overline{h}}_H$ because it is easier to compute than $\hat{\overline{h}}_{PEML}$.

Next, if we consider RHC sampling design, it follows from the proof of Theorem 2.1.2 that the asymptotic MSE of $\sqrt{n}(g(\overline{h}) - g(\hat{\overline{h}}))$ is $\tilde{\Delta}_2^2$=$\lim_{\nu\to\infty} n\gamma\overline{X}N^{-1}\nabla g(\overline{h})\sum_{i=1}^{N}(\mathbf{V}_i - X_i\overline{\mathbf{V}}/\overline{X})^T(\mathbf{V}_i - X_i\overline{\mathbf{V}}/\overline{X})X_i^{-1}\nabla g(\overline{h})^T$, where γ and $\overline{\mathbf{V}}$ are as in the paragraph following Assumption 2.1.3. Moreover, $\mathbf{V}_i$ is h_i or $h_i - \overline{h} - S_{xh}(X_i - \overline{X})/S_x^2$ if $\hat{\overline{h}}$ is $\hat{\overline{h}}_{RHC}$ or $\hat{\overline{h}}_{PEML}$ (as well as $\hat{\overline{h}}_{GREG}$) with $d(i,s)$=$(NX_i)^{-1}G_i$, respectively. Here, G_i is the total of the x values of that randomly formed group from which the i^{th} population unit is selected in the sample by RHC sampling design (cf. [20]). We estimate $\tilde{\Delta}_2^2$ by

$$\hat{\Delta}_2^2 = n\gamma\overline{X}N^{-1}\nabla g(\hat{\overline{h}})\sum_{i\in s}(\hat{\mathbf{V}}_i - X_i\hat{\overline{\mathbf{V}}}_{RHC}/\overline{X})^T\times \\ (\hat{\mathbf{V}}_i - X_i\hat{\overline{\mathbf{V}}}_{RHC}/\overline{X})(G_iX_i^{-2})\nabla g(\hat{\overline{h}})^T, \tag{2.3.2}$$

where $\hat{\overline{\mathbf{V}}}_{RHC}$=$\sum_{i\in s}(NX_i)^{-1}G_i\hat{\mathbf{V}}_i$, $\hat{\overline{h}}$=$\hat{\overline{h}}_{RHC}$ in the case of the mean, the variance and the regression coefficient, and $\hat{\overline{h}}$=$\hat{\overline{h}}_{PEML}$ in the case of the correlation coefficient. Here, $\hat{\mathbf{V}}_i$ is h_i or $h_i - \hat{\overline{h}}_{RHC} - \hat{S}_{xh,2}(X_i - \overline{X})/\hat{S}_{x,2}^2$ if $\hat{\overline{h}}$ is $\hat{\overline{h}}_{RHC}$ or $\hat{\overline{h}}_{PEML}$ (as well as $\hat{\overline{h}}_{GREG}$) with $d(i,s)$=$(NX_i)^{-1}G_i$. Further, $\hat{S}_{xh,2}$=$\sum_{i\in s}N^{-1}G_ih_i - \overline{X}\,\hat{\overline{h}}_{RHC}$ and $\hat{S}_{x,1}^2$=$\sum_{i\in s}N^{-1}G_iX_i - \overline{X}^2$. In the case of the mean, the variance and the regression coefficient, we estimate $\overline{h}$ in $\nabla g(\overline{h})$ by $\hat{\overline{h}}_{RHC}$ for the same reason as discussed in the preceding paragraph, where we discuss the estimation of $\overline{h}$ by $\hat{\overline{h}}_{HT}$ under SRSWOR and RS sampling design. On the other hand, in the case of the correlation coefficient, we estimate $\overline{h}$ in $\nabla g(\overline{h})$ by $\hat{\overline{h}}_{PEML}$ under RHC sampling design so that the estimator of the correlation coefficient appeared in the expression of $\nabla g(\overline{h})$ in this case becomes well defined.

We draw I=1000 samples each of sizes n=75, 100 and 125 using sampling designs mentioned in Table 2.6. Then, for each of the parameters, the sampling designs and the estimators mentioned in Table 2.6, we construct I many asymptotically 95% CIs based on these samples and compute the average and the standard deviation (s.d.) of their lengths.

2.3.1 Analysis based on synthetic data

In this section, we consider the population values $\{(Y_i, X_i) : 1 \leq i \leq N\}$ on (y, x) generated from a linear model as follows. We choose N=5000 and generate the X_i's from a gamma distribution with mean 1000 and s.d. 200. Then, Y_i is generated from the linear model Y_i=500 +

X_i+ϵ_i for i=1, . . . , N, where ϵ_i is generated independently of $\{X_i\}_{i=1}^N$ from a normal distribution with mean 0 and s.d. 100. We also generate the population values $\{(Y_i, X_i) : 1 \leq i \leq N\}$ from a linear model, when y=(z_1, z_2) is a bivariate study variable. The population values $\{X_i\}_{i=1}^N$ are generated in the same way as in the earlier case. Then, Y_i=(Z_{1i}, Z_{2i}) is generated from the linear model Z_{ji}=$\alpha_j + X_i + \epsilon_{ji}$ for i=1, . . . , N, where α_1=500 and α_2=1000. The ϵ_{1i}'s are generated independently of the X_i's from a normal distribution with mean 0 and s.d. 100, and the ϵ_{2i}'s are generated independently of the X_i's and the ϵ_{1i}'s from a normal distribution with mean 0 and s.d. 200. We consider the estimation of the mean and the variance of y for the first data set and the correlation and the regression coefficients between z_1 and z_2 for the second data set.

The results of the empirical comparison based on synthetic data are summarized as follows. For each of the mean, the variance, the correlation coefficient and the regression coefficient, the plug-in estimator based on the PEML estimator under SRSWOR turns out to be more efficient than any other estimator under any other sampling design (see Tables 2.7–2.11) considered in Table 2.6 when compared in terms of relative efficiencies. Also, for each of the above parameters, asymptotically 95% CI based on the PEML estimator under SRSWOR has the least average length (see Tables 2.12–2.16). Thus the empirical results stated here corroborate the theoretical results stated in Theorems 2.2.1–2.2.4.

TABLE 2.7: Relative efficiencies of estimators for mean of y.

Relative efficiency \ Sample size	n=75	n=100	n=125
RE($\hat{\overline{Y}}_{PEML}$, SRSWOR | $\hat{\overline{Y}}_{GREG}$, SRSWOR)	1.049985	1.020252	1.035038
RE($\hat{\overline{Y}}_{PEML}$, SRSWOR | $\hat{\overline{Y}}_{H}$, RS)	4.870516	5.370899	4.987635
RE($\hat{\overline{Y}}_{PEML}$, SRSWOR | $\hat{\overline{Y}}_{HT}$, RS)	2.026734	2.061607	2.027386
RE($\hat{\overline{Y}}_{PEML}$, SRSWOR | $\hat{\overline{Y}}_{PEML}$, RS)	1.144439	1.124697	1.170224
RE($\hat{\overline{Y}}_{PEML}$, SRSWOR | $\hat{\overline{Y}}_{GREG}$, RS)	1.144455	1.124975	1.170267
RE($\hat{\overline{Y}}_{PEML}$, SRSWOR| $\hat{\overline{Y}}_{RHC}$, RHC)	2.022378	1.978623	2.143015
RE($\hat{\overline{Y}}_{PEML}$, SRSWOR | $\hat{\overline{Y}}_{PEML}$, RHC)	1.089837	1.030332	1.094067
RE($\hat{\overline{Y}}_{PEML}$, SRSWOR | $\hat{\overline{Y}}_{GREG}$, RHC)	1.089853	1.030587	1.094108

TABLE 2.8: Relative efficiencies of estimators for variance of y. Recall from Table 2.5 in Section 2.1 that for variance of y, $h(y)$=(y^2, y) and $g(s_1, s_2)$=$s_1 - s_2^2$.

Relative efficiency \ Sample size	n=75	n=100	n=125
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | $g(\hat{\bar{h}}_{H})$, SRSWOR)	1.0926	1.0848	1.0419
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | $g(\hat{\bar{h}}_{H})$, RS)	1.0367	1.0435	1.0226
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | $g(\hat{\bar{h}}_{PEML})$, RS)	1.15067	1.136	1.1635
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | $g(\hat{\bar{h}}_{PEML})$, RHC)	1.141	1.1849	1.1631

TABLE 2.9: Relative efficiencies of estimators for correlation coefficient between z_1 and z_2. Recall from Table 2.5 in Section 2.1 that for correlation coefficient between z_1 and z_2, $h(z_1, z_2)$=$(z_1, z_2, z_1^2, z_2^2, z_1 z_2)$ and $g(s_1, s_2, s_3, s_4, s_5)$=$(s_5 - s_1 s_2)/((s_3 - s_1^2)(s_4 - s_2^2))^{1/2}$.

Relative efficiency \ Sample size	n=75	n=100	n=125
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | $g(\hat{\bar{h}}_{H})$, SRSWOR)	1.0304	1.0274	1.0385
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | $g(\hat{\bar{h}}_{H})$, RS)	1.0307	1.0838	1.0515
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | $g(\hat{\bar{h}}_{PEML})$, RS)	1.0573	1.1862	1.1081
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | $g(\hat{\bar{h}}_{PEML})$, RHC)	1.0847	1.1459	1.0911

TABLE 2.10: Relative efficiencies of estimators for regression coefficient of z_1 on z_2. Recall from Table 2.5 in Section 2.1 that for regression coefficient of z_1 on z_2, $h(z_1, z_2)$=$(z_1, z_2, z_2^2, z_1 z_2)$ and $g(s_1, s_2, s_3, s_4)$=$(s_4 - s_1 s_2)/(s_3 - s_2^2)$.

Relative efficiency \ Sample size	n=75	n=100	n=125
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | $g(\hat{\bar{h}}_{H})$, SRSWOR)	1.0389	1.0473	1.0218
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | $g(\hat{\bar{h}}_{H})$, RS)	1.0589	1.0829	1.0827
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | $g(\hat{\bar{h}}_{PEML})$, RS)	1.1219	1.1334	1.2137
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | $g(\hat{\bar{h}}_{PEML})$, RHC)	1.2037	1.1307	1.1399

TABLE 2.11: Relative efficiencies of estimators for regression coefficient of z_2 on z_1. Recall from Table 2.5 in Section 2.1 that for regression coefficient of z_2 on z_1, $h(z_1, z_2)$=$(z_2, z_1, z_1^2, z_1 z_2)$ and $g(s_1, s_2, s_3, s_4)$=$(s_4 - s_1 s_2)/(s_3 - s_2^2)$.

Relative efficiency \ Sample size	n=75	n=100	n=125
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | $g(\hat{\bar{h}}_H)$, SRSWOR)	1.0498	1.04	1.0301
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | $g(\hat{\bar{h}}_H)$, RS)	1.0655	1.0652	1.0548
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | $g(\hat{\bar{h}}_{PEML})$, RS)	1.1073	1.1153	1.1135
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | $g(\hat{\bar{h}}_{PEML})$, RHC)	1.0762	1.0905	1.1108

TABLE 2.12: Average and s.d. of lengths of asymptotically 95% CIs for mean of y.

	Average length (s.d.)		
Estimator and sampling design based on which CI is constructed \ Sample size	n=75	n=100	n=125
$\hat{\bar{Y}}_H$, SRSWOR	536.821 (11.357)	538.177 (9.0784)	539.218 (6.8211)
[4] $\hat{\bar{Y}}_{PEML}$, SRSWOR	44.824 (3.7002)	38.81 (2.7727)	34.648 (2.2055)
$\hat{\bar{Y}}_{HT}$, RS	689.123 (7.8452)	597.999 (5.7176)	535.951 (4.8422)
$\hat{\bar{Y}}_H$, RS	102.611 (10.969)	87.915 (8.453)	59.98307 (6.5828)
[4] $\hat{\bar{Y}}_{PEML}$, RS	345.956 (654.77)	115.944 (265.93)	78.711 (1041.2)
$\hat{\bar{Y}}_{RHC}$, RHC	848.033 (6.8489)	624.881 (4.9609)	541.421 (4.0927)
[4] $\hat{\bar{Y}}_{PEML}$, RHC	64.573 (715.16)	56.531 (275.11)	50.601 (651.31)

[4] It is to be noted that in the cases of PEML and GREG estimators under any given sampling design, we have the same asymptotic MSE and hence the same asymptotic CI. Therefore, the average and the s.d. of lengths of CIs are not reported for the GREG estimator.

TABLE 2.13: Average and s.d. of lengths of asymptotically 95% CIs for variance of y. Recall from Table 2.5 in Section 2.1 that for variance of y, $h(y_1)$=(y^2, y) and $g(s_1, s_2)$=$s_1 - s_2^2$.

Estimator and sampling design based on which CI is constructed \ Sample size	Average length (s.d.) n=75	n=100	n=125
$g(\hat{\overline{h}}_H)$, SRSWOR	1010775 (34245.5)	878689.4 (26373.9)	786228 (20414.5)
$g(\hat{\overline{h}}_{PEML})$, SRSWOR	29432.4 (6076.97)	25929 (4441.2)	23422 (3526.8)
$g(\hat{\overline{h}}_H)$, RS	444594.4 (44701.7)	434160.7 (31965.2)	239065 (26739.6)
$g(\hat{\overline{h}}_{PEML})$, RS	1152403 (9083944)	1290084 (869339.1)	235909.1 (1183961)
$g(\hat{\overline{h}}_{PEML})$, RHC	1031407 (7311193)	895639 (1530759)	801178.9 (417582.9)

TABLE 2.14: Average and s.d. of lengths of asymptotically 95% CIs for correlation coefficient between z_1 and z_2. Recall from Table 2.5 in Section 2.1 that for correlation coefficient between z_1 and z_2, $h(z_1, z_2)$=$(z_1, z_2, z_1^2, z_2^2, z_1 z_2)$ and $g(s_1, s_2, s_3, s_4, s_5)$=$(s_5 - s_1 s_2)/((s_3 - s_1^2)(s_4 - s_2^2))^{1/2}$.

Estimator and sampling design based on which CI is constructed \ Sample size	Average length (s.d.) n=75	n=100	n=125
$g(\hat{\overline{h}}_H)$, SRSWOR	8.2191 (2.429)	8.0909 (1.889)	8.0897 (1.449)
$g(\hat{\overline{h}}_{PEML})$, SRSWOR	0.2542 (0.0467)	0.2575 (0.0365)	0.2583 (0.0294)
$g(\hat{\overline{h}}_H)$, RS	4.6847 (2.555)	3.3135 (1.884)	1.3942 (1.421)
$g(\hat{\overline{h}}_{PEML})$, RS	5.0473 (162.9)	4.3229 (17.19)	3.1306 (21.04)
$g(\hat{\overline{h}}_{PEML})$, RHC	8.3174 (15.82)	8.3898 (41.88)	8.3514 (19.62)

TABLE 2.15: Average and s.d. of lengths of asymptotically 95% CIs for regression coefficient of z_1 on z_2. Recall from Table 2.5 in Section 2.1 that for regression coefficient of z_1 on z_2, $h(z_1, z_2)$=$(z_1, z_2, z_2^2, z_1 z_2)$ and $g(s_1, s_2, s_3, s_4)$=$(s_4 - s_1 s_2)/(s_3 - s_2^2)$.

	Average length (s.d.)		
Sample size / Estimator and sampling design based on which CI is constructed	n=75	n=100	n=125
$g(\hat{\overline{h}}_H)$, SRSWOR	5.9565 (2.013)	5.068 (1.514)	4.4818 (1.135)
$g(\hat{\overline{h}}_{PEML})$, SRSWOR	0.2596 (0.0429)	0.2251 (0.0324)	0.2032 (0.025)
$g(\hat{\overline{h}}_H)$, RS	3.0488 (2.178)	1.469 (1.517)	1.1532 (1.171)
$g(\hat{\overline{h}}_{PEML})$, RS	3.6477 (19.09)	1.8558 (4.697)	1.4023 (4.672)
$g(\hat{\overline{h}}_{PEML})$, RHC	6.111 (25.16)	5.1324 (38.36)	4.6658 (11.17)

TABLE 2.16: Average and s.d. of lengths of asymptotically 95% CIs for regression coefficient of z_2 on z_1. Recall from Table 2.5 in Section 2.1 that for regression coefficient of z_2 on z_1, $h(z_1, z_2)$=$(z_2, z_1, z_1^2, z_1 z_2)$ and $g(s_1, s_2, s_3, s_4)$=$(s_4 - s_1 s_2)/(s_3 - s_2^2)$.

	Average length (s.d.)		
Sample size / Estimator and sampling design based on which CI is constructed	n=75	n=100	n=125
$g(\hat{\overline{h}}_H)$, SRSWOR	11.2173 (3.238)	9.6463 (2.418)	8.5885 (1.877)
$g(\hat{\overline{h}}_{PEML})$, SRSWOR	0.4198 (0.0661)	0.3652 (0.0531)	0.3307 (0.0405)
$g(\hat{\overline{h}}_H)$, RS	6.7247 (3.546)	3.3547 (2.539)	1.7421 (1.921)
$g(\hat{\overline{h}}_{PEML})$, RS	11.3373 (151.9)	9.988 (31.83)	8.7889 (7.405)
$g(\hat{\overline{h}}_{PEML})$, RHC	19.9049 (28.77)	3.5595 (321.7)	1.8327 (8.164)

2.3.2 Analysis based on real data

In this section, we consider a data set on the village amenities in the state of West Bengal in India obtained from the Office of the Registrar General & Census Commissioner, India (*https://censusindia.gov.in*). Relevant study variables for this data set are described in Table 2.17 below. We consider the following estimation problems for a population consisting of 37478

TABLE 2.17: Description of study variables.

y_1	Number of primary schools in village
y_2	Scheduled castes population size in village
y_3	Number of secondary schools in village
y_4	Scheduled tribes population size in village

villages. For these estimation problems, we use the number of people living in village x as the size variable.

(i) First, we consider the estimation of the mean and the variance of each of y_1 and y_2. It can be shown from the scatter plot and the least square regression line in Figure 2.1 below that y_1 and x have an approximate linear relationship. Also, the correlation coefficient between y_1 and x is 0.72. On the other hand, y_2 and x do not seem to have a linear relationship (see the scatter plot and the least square regression line in Figure 2.2 below).

(ii) Next, we consider the estimation of the correlation and the regression coefficients of y_1 and y_3 as well as of y_2 and y_4. The scatter plot and the least square regression line in Figure 2.3 below show that y_3 does not seem to be dependent on x. Further, we see from the scatter plot and the least square regression line of y_4 and x (see Figure 2.4 below) that y_4 and x do not seem to have a linear relationship.

The results of the empirical comparison based on real data are summarized in Table 2.18 below. For further details see Tables 2.19–2.38 below. The approximate linear relationship between y_1 and x (see the scatter plot and the least square regression line in Figure 2.1 below) could be a possible reason why the plug-in estimator based on the PEML estimator under SRSWOR becomes the most efficient for each of the mean and the variance of y_1 among all the estimators under different sampling designs considered in this section. Also, possibly for the same reason, the plug-in estimators of the correlation and the regression coefficients between y_1 and y_3 based on the PEML estimator under SRSWOR become the most efficient among all the estimators under different sampling designs considered in this section.

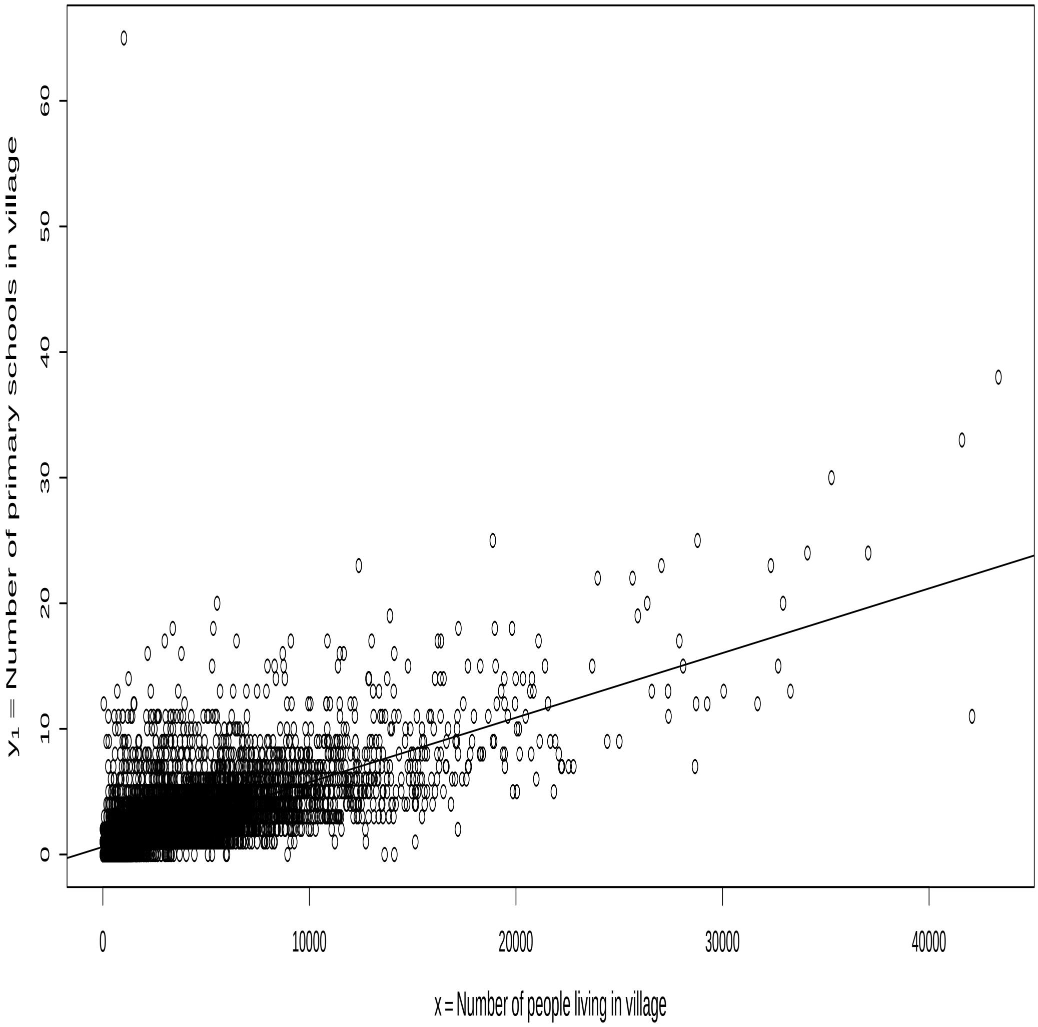

FIGURE 2.1: Scatter plot and least square regression line for variables y_1 and x.

On the other hand, any of y_2, and y_4 does not seem to have a linear relationship with x (see the scatter plots and the least square regression lines in Figures 2.2 and 2.4 below). Possibly, because of this reason, the plug-in estimators of the parameters related to y_2 and y_4 based on the PEML estimator are not able to outperform the the plug-in estimators of those parameters based on the HT and the Hájek estimators. Next, we observe that there are substantial correlation present between y_2 and x (correlation coefficient=0.67), and y_4 and x (correlation coefficient=0.25).

Possibly, because of this, under RS sampling design, which uses the auxiliary information, the plug-in estimators of the parameters related to y_2 and y_4 based on the HT and the Hájek estimators become the most efficient among all the estimators under different sampling designs considered in this section.

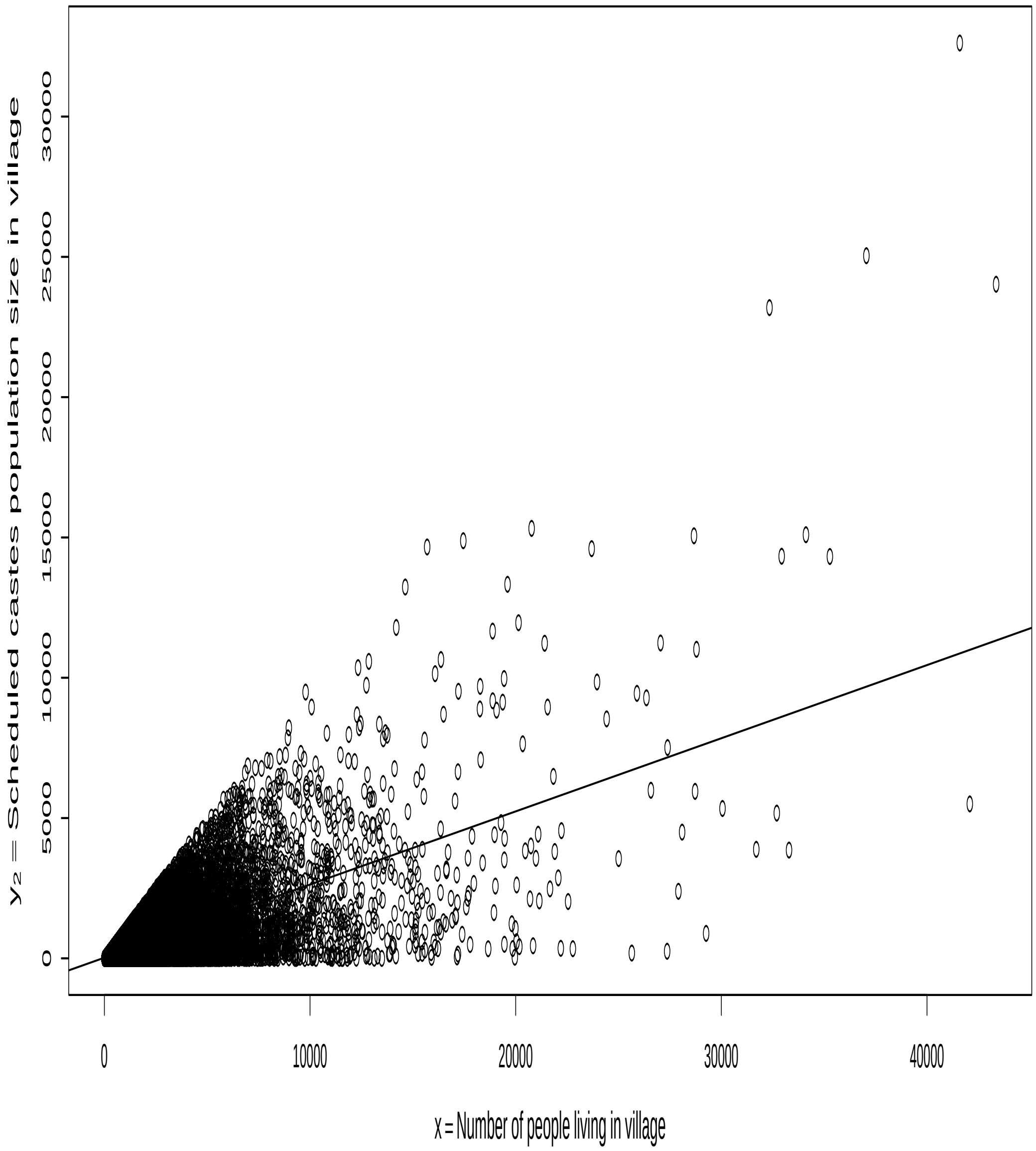

FIGURE 2.2: Scatter plot and least square regression line for variables y_2 and x.

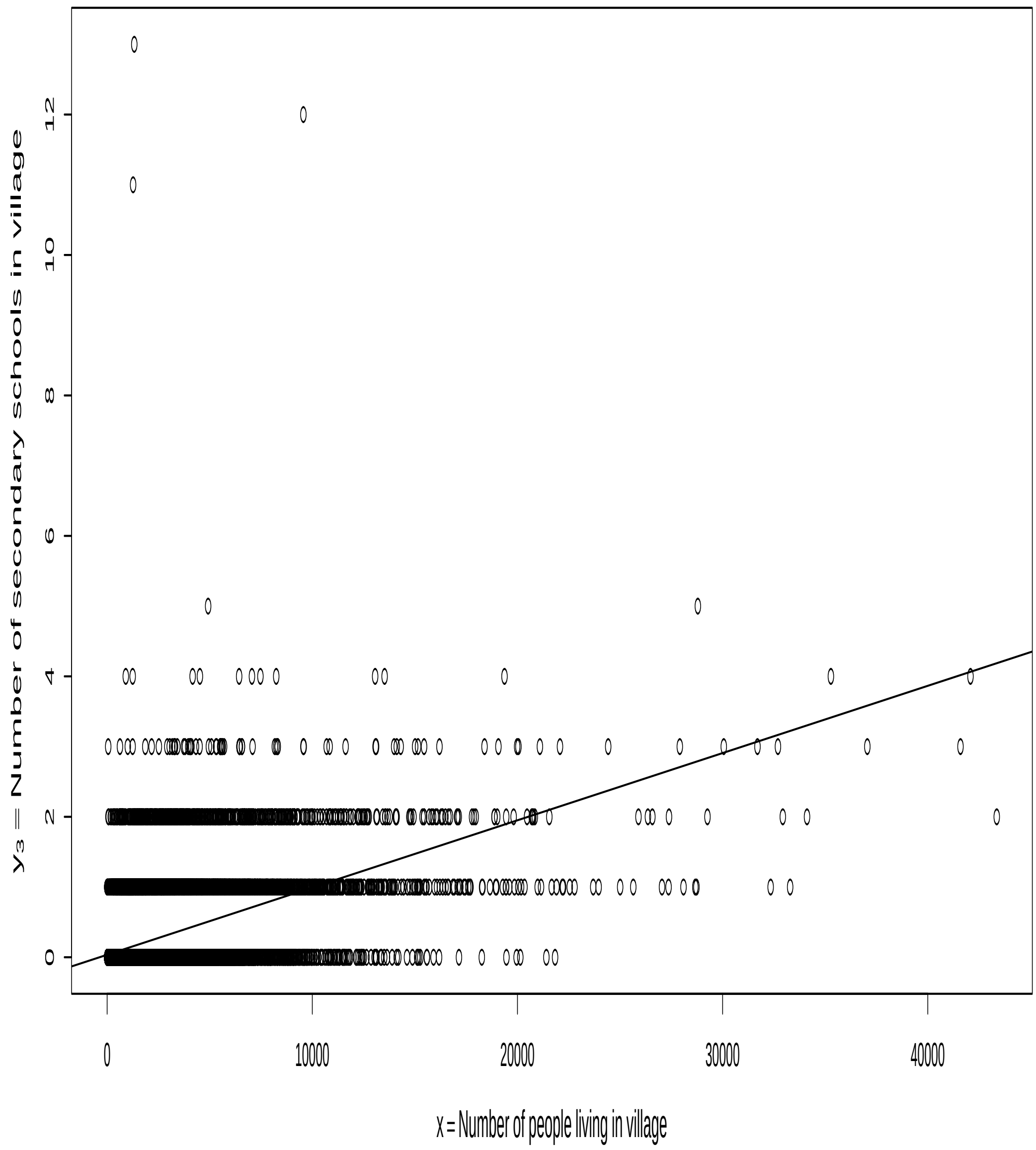

FIGURE 2.3: Scatter plot and least square regression line for variables y_3 and x.

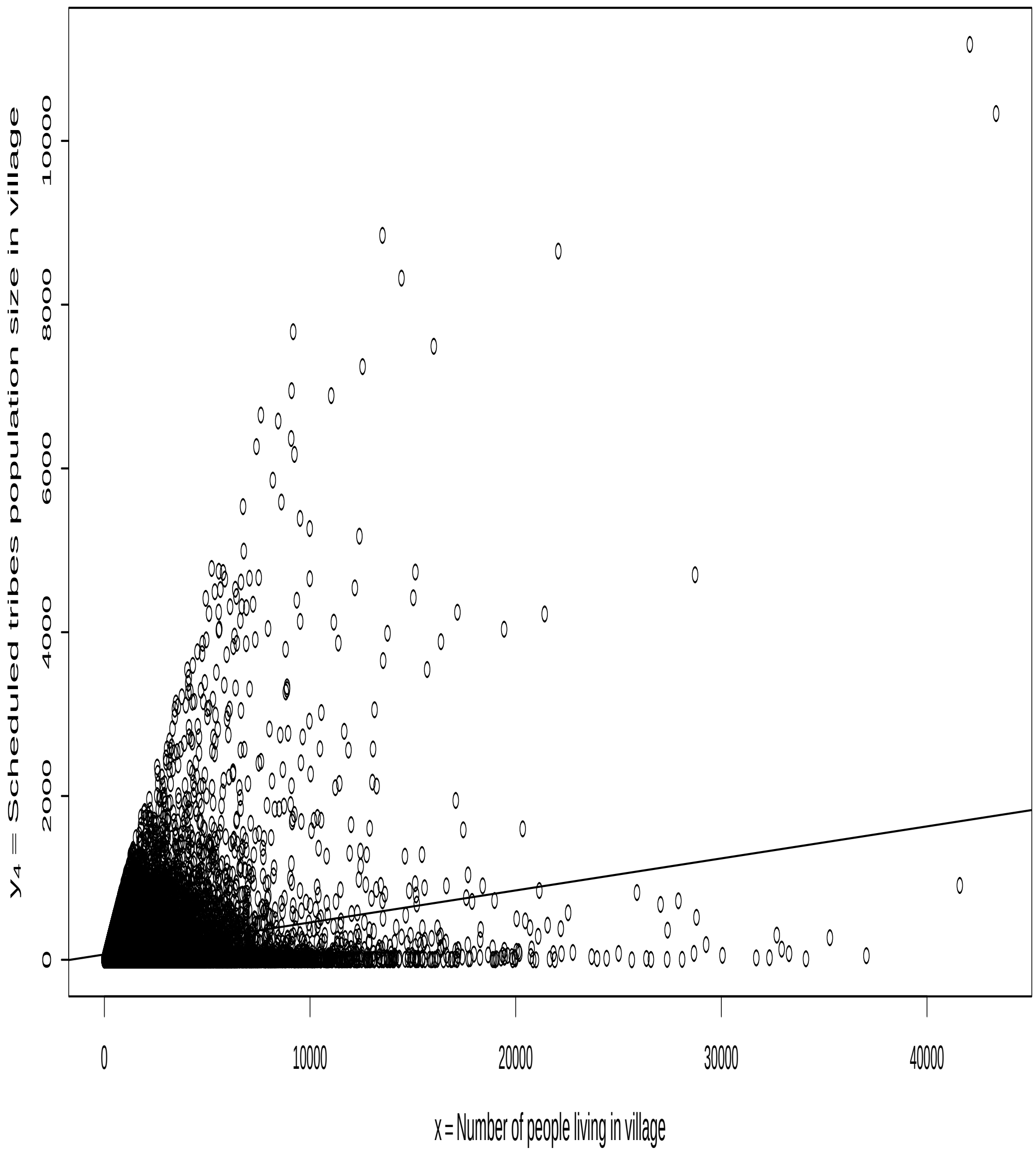

FIGURE 2.4: Scatter plot and least square regression line for variables y_4 and x.

TABLE 2.18: Most efficient estimators in terms of relative efficiencies (it follows from Tables 2.29–2.38 that asymptotically 95% CIs based on most efficient estimators have least average lengths).

Parameters	Most efficient estimators
Mean and variance of y_1	The plug-in estimator based on the the PEML estimator under SRSWOR
Mean of y_2	The HT estimator under RS sampling design
Variance of y_2	the plug-in estimator based on the Hájek estimator under RS sampling design
Correlation and regression coefficients of y_1 and y_3	The plug-in estimator based on the PEML estimator under SRSWOR
Correlation and regression coefficients of y_2 and y_4	The plug-in estimator based on the Hájek estimator under RS sampling design

TABLE 2.19: Relative efficiencies of estimators for mean of y_1.

Relative efficiency \ Sample size	n=75	n=100	n=125
RE($\hat{\overline{Y}}_{PEML}$, SRSWOR | $\hat{\overline{Y}}_{GREG}$, SRSWOR)	1.008215	1.005233	1.020408
RE($\hat{\overline{Y}}_{PEML}$, SRSWOR | $\hat{\overline{Y}}_{H}$, RS)	3.503939	3.880443	4.175886
RE($\hat{\overline{Y}}_{PEML}$, SRSWOR | $\hat{\overline{Y}}_{HT}$, RS)	1.796937	2.182675	1.8311
RE($\hat{\overline{Y}}_{PEML}$, SRSWOR | $\hat{\overline{Y}}_{PEML}$, RS)	1.20961	1.228022	1.50233
RE($\hat{\overline{Y}}_{PEML}$, SRSWOR | $\hat{\overline{Y}}_{GREG}$, RS)	1.21831	1.237737	1.553863
RE($\hat{\overline{Y}}_{PEML}$, SRSWOR | $\hat{\overline{Y}}_{RHC}$, RHC)	3.274031	2.059141	2.030995
RE($\hat{\overline{Y}}_{PEML}$, SRSWOR | $\hat{\overline{Y}}_{PEML}$, RHC)	1.088166	1.388563	1.51547
RE($\hat{\overline{Y}}_{PEML}$, SRSWOR | $\hat{\overline{Y}}_{GREG}$, RHC)	1.097934	1.398241	1.567545

TABLE 2.20: Relative efficiencies of estimators for variance of y_1. Recall from Table 2.5 in Section 2.1 that for variance of y_1, $h(y_1)$=(y_1^2, y_1) and $g(s_1, s_2)$=$s_1 - s_2^2$.

Relative efficiency \ Sample size	n=75	n=100	n=125
RE($g(\hat{\overline{h}}_{PEML})$, SRSWOR | $g(\hat{\overline{h}}_{H})$, SRSWOR)	1.3294	1.2413	1.1476
RE($g(\hat{\overline{h}}_{PEML})$, SRSWOR | $g(\hat{\overline{h}}_{H})$, RS)	2.5303	1.6656	1.5374
RE($g(\hat{\overline{h}}_{PEML})$, SRSWOR | $g(\hat{\overline{h}}_{PEML})$, RS)	3.1642	2.4051	2.5831
RE($g(\hat{\overline{h}}_{PEML})$, SRSWOR | $g(\hat{\overline{h}}_{PEML})$, RHC)	2.5499	4.7704	3.0985

TABLE 2.21: Relative efficiencies of estimators for mean of y_2.

Relative efficiency \ Sample size	n=75	n=100	n=125
RE($\hat{\overline{Y}}_{HT}$, RS | $\hat{\overline{Y}}_{H}$, RS)	4.367712	4.008655	4.463214
RE($\hat{\overline{Y}}_{HT}$, RS | $\hat{\overline{Y}}_{PEML}$, RS)	1.148074	1.082488	1.088804
RE($\hat{\overline{Y}}_{HT}$, RS | $\hat{\overline{Y}}_{GREG}$, RS)	1.216958	1.115967	1.154132
RE($\hat{\overline{Y}}_{HT}$, RS | $\hat{\overline{Y}}_{RHC}$, RHC)	1.073138	1.03213	1.07484
RE($\hat{\overline{Y}}_{HT}$, RS | $\hat{\overline{Y}}_{PEML}$, RHC)	1.230884	1.0937	1.207308
RE($\hat{\overline{Y}}_{HT}$, RS | $\hat{\overline{Y}}_{GREG}$, RHC)	1.304737	1.127526	1.279746
RE($\hat{\overline{Y}}_{HT}$, RS | $\hat{\overline{Y}}_{PEML}$, SRSWOR)	2.440441	2.305339	2.350916
RE($\hat{\overline{Y}}_{HT}$, RS | $\hat{\overline{Y}}_{GREG}$, SRSWOR)	2.58687	2.376638	2.49197

TABLE 2.22: Relative efficiencies of estimators for variance of y_2. Recall from Table 2.5 in Section 2.1 that for variance of y_2, $h(y_2)$=(y_2^2, y_2) and $g(s_1, s_2)$=$s_1 - s_2^2$.

Relative efficiency \ Sample size	n=75	n=100	n=125
RE($g(\hat{\overline{h}}_H)$, RS | $g(\hat{\overline{h}}_{PEML})$, RS)	11.893	6.967	34.691
RE($g(\hat{\overline{h}}_H)$, RS | $g(\hat{\overline{h}}_{PEML})$, RHC)	5.0093	19.456	21.919
RE($g(\hat{\overline{h}}_H)$, RS | $g(\hat{\overline{h}}_{H})$, SRSWOR)	9.8232	10.27	16.763
RE($g(\hat{\overline{h}}_H)$, RS | $g(\hat{\overline{h}}_{PEML})$, SRSWOR)	2.4768	4.8093	6.2264

TABLE 2.23: Relative efficiencies of estimators for correlation coefficient between y_1 and y_3. Recall from Table 2.5 in Section 2.1 that for correlation coefficient between y_1 and y_3, $h(y_1, y_3)$=($y_1, y_3, y_1^2, y_3^2, y_1 y_3$) and $g(s_1, s_2, s_3, s_4, s_5)$=$(s_5 - s_1 s_2)/((s_3 - s_1^2)(s_4 - s_2^2))^{1/2}$.

Relative efficiency \ Sample size	n=75	n=100	n=125
RE($g(\hat{\overline{h}}_{PEML})$, SRSWOR | $g(\hat{\overline{h}}_H)$, SRSWOR)	1.0967	1.0369	1.0374
RE($g(\hat{\overline{h}}_{PEML})$, SRSWOR | $g(\hat{\overline{h}}_H)$, RS)	1.317	1.4831	1.2561
RE($g(\hat{\overline{h}}_{PEML})$, SRSWOR | $g(\hat{\overline{h}}_{PEML})$, RS)	1.9803	1.9874	1.8441
RE($g(\hat{\overline{h}}_{PEML})$, SRSWOR | $g(\hat{\overline{h}}_{PEML})$, RHC)	2.0562	1.9651	1.8541

TABLE 2.24: Relative efficiencies of estimators for regression coefficient of y_1 on y_3. Recall from Table 2.5 in Section 2.1 that for regression coefficient of y_1 on y_3, $h(y_1, y_3)$=$(y_1, y_3, y_3^2, y_1y_3)$ and $g(s_1, s_2, s_3, s_4)$=$(s_4 - s_1s_2)/(s_3 - s_2^2)$.

Relative efficiency \ Sample size	n=75	n=100	n=125
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | $g(\hat{\bar{h}}_H)$, SRSWOR)	1.0298	1.0504	1.0423
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | $g(\hat{\bar{h}}_H)$, RS)	1.8046	1.2304	1.3482
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | $g(\hat{\bar{h}}_{PEML})$, RS)	2.2709	1.5949	1.854
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | $g(\hat{\bar{h}}_{PEML})$, RHC)	1.8719	1.5069	1.5626

TABLE 2.25: Relative efficiencies of estimators for regression coefficient of y_3 on y_1. Recall from Table 2.5 in Section 2.1 that for regression coefficient of y_3 on y_1, $h(y_1, y_3)$=$(y_3, y_1, y_1^2, y_1y_3)$ and $g(s_1, s_2, s_3, s_4)$=$(s_4 - s_1s_2)/(s_3 - s_2^2)$.

Relative efficiency \ Sample size	n=75	n=100	n=125
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | $g(\hat{\bar{h}}_H)$, SRSWOR)	1.0997	1.2329	1.1529
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | $g(\hat{\bar{h}}_H)$, RS)	1.3948	1.3329	1.368
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | $g(\hat{\bar{h}}_{PEML})$, RS)	3.6069	1.5532	1.8035
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | $g(\hat{\bar{h}}_{PEML})$, RHC)	2.5567	1.4867	1.5335

TABLE 2.26: Relative efficiencies of estimators for correlation coefficient between y_2 and y_4. Recall from Table 2.5 in Section 2.1 that for correlation coefficient between y_2 and y_4, $h(y_2, y_4)$=$(y_2, y_4, y_2^2, y_4^2, y_2y_4)$ and $g(s_1, s_2, s_3, s_4, s_5)$=$(s_5 - s_1s_2)/((s_3 - s_1^2)(s_4 - s_2^2))^{1/2}$.

Relative efficiency \ Sample size	n=75	n=100	n=125
RE($g(\hat{\bar{h}}_H)$, RS | $g(\hat{\bar{h}}_{PEML})$, RS)	1.448	1.696	2.027
RE($g(\hat{\bar{h}}_H)$, RS | $g(\hat{\bar{h}}_{PEML})$, RHC)	1.491	2.135	2.27
RE($g(\hat{\bar{h}}_H)$, RS | $g(\hat{\bar{h}}_H)$, SRSWOR)	2.39	2.521	2.849
RE($g(\hat{\bar{h}}_H)$, RS | $g(\hat{\bar{h}}_{PEML})$, SRSWOR)	2.185	2.396	2.594

TABLE 2.27: Relative efficiencies of estimators for regression coefficient of y_2 on y_4. Recall from Table 2.5 in Section 2.1 that for regression coefficient of y_2 on y_4, $h(y_2, y_4)$=$(y_2, y_4, y_4^2, y_2y_4)$ and $g(s_1, s_2, s_3, s_4)$=$(s_4 - s_1s_2)/(s_3 - s_2^2)$.

Relative efficiency \ Sample size	n=75	n=100	n=125
RE($g(\hat{\bar{h}}_H)$, RS | $g(\hat{\bar{h}}_{PEML})$, RS)	1.8158	2.3771	3.2021
RE($g(\hat{\bar{h}}_H)$, RS | $g(\hat{\bar{h}}_{PEML})$, RHC)	2.5985	2.6002	3.4744
RE($g(\hat{\bar{h}}_H)$, RS | $g(\hat{\bar{h}}_H)$, SRSWOR)	3.3278	4.5041	6.312
RE($g(\hat{\bar{h}}_H)$, RS | $g(\hat{\bar{h}}_{PEML})$, SRSWOR)	2.9788	3.9417	6.0391

TABLE 2.28: Relative efficiencies of estimators for regression coefficient of y_4 on y_2. Recall from Table 2.5 in Section 2.1 that for regression coefficient of y_4 on y_2, $h(y_2, y_4)$=$(y_4, y_2, y_2^2, y_2y_4)$ and $g(s_1, s_2, s_3, s_4)$=$(s_4 - s_1s_2)/(s_3 - s_2^2)$.

Relative efficiency \ Sample size	n=75	n=100	n=125
RE($g(\hat{\bar{h}}_H)$, RS | $g(\hat{\bar{h}}_{PEML})$, RS)	1.3146	1.6055	1.937
RE($g(\hat{\bar{h}}_H)$, RS | $g(\hat{\bar{h}}_{PEML})$, RHC)	1.652	2.7715	2.0362
RE($g(\hat{\bar{h}}_H)$, RS | $g(\hat{\bar{h}}_H)$, SRSWOR)	3.8248	2.4388	3.4371
RE($g(\hat{\bar{h}}_H)$, RS | $g(\hat{\bar{h}}_{PEML})$, SRSWOR)	3.1843	2.3399	3.038

TABLE 2.29: Average and s.d. of lengths of asymptotically 95% CIs for mean of y_1.

	Average length (s.d.)		
Estimator and sampling design based on which CI is constructed \ Sample size	n=75	n=100	n=125
$\hat{\bar{Y}}_H$, SRSWOR	0.7233 (0.2304)	0.7303 (0.1885)	0.7333 (0.1431)
${}^4\hat{\bar{Y}}_{PEML}$, SRSWOR	0.3703 (0.1608)	0.3734 (0.1534)	0.3847 (0.1074)
$\hat{\bar{Y}}_{HT}$, RS	0.7738 (0.2724)	0.7735 (1.071)	0.8271 (0.2001)
$\hat{\bar{Y}}_H$, RS	0.4345 (0.8312)	0.455 (8.807)	0.5414 (0.5479)
${}^4\hat{\bar{Y}}_{PEML}$, RS	0.6784 (0.3945)	0.7207 (12.176)	0.7896 (0.2694)
$\hat{\bar{Y}}_{RHC}$, RHC	0.7415 (0.4007)	0.7716 (0.6359)	0.8014 (0.2931)
${}^4\hat{\bar{Y}}_{PEML}$, RHC	0.4911 (0.9865)	0.5078 (0.4992)	0.5289 (0.3594)

TABLE 2.30: Average and s.d. of lengths of asymptotically 95% CIs for variance of y_1. Recall from Table 2.5 in Section 2.1 that for variance of y_1, $h(y_1)$=(y_1^2, y_1) and $g(s_1, s_2)$=$s_1 - s_2^2$.

Estimator and sampling design based on which CI is constructed \ Sample size	Average length (s.d.) n=75	n=100	n=125
$g(\hat{\overline{h}}_H)$, SRSWOR	5.2879 (8.762)	4.2111 (9.309)	4.4304 (6.856)
$g(\hat{\overline{h}}_{PEML})$, SRSWOR	2.7519 (7.181)	2.9935 (8.622)	3.0013 (5.952)
$g(\hat{\overline{h}}_H)$, RS	3.5121 (1.345)	3.1177 (11.37)	3.1095 (10.88)
$g(\hat{\overline{h}}_{PEML})$, RS	3.7475 (4.041)	3.939 (16.14)	3.792 (11.08)
$g(\hat{\overline{h}}_{PEML})$, RHC	3.6365 (14.99)	3.4972 (8.278)	3.4158 (10.95)

TABLE 2.31: Average and s.d. of lengths of asymptotically 95% CIs for mean of y_2.

Estimator and sampling design based on which CI is constructed \ Sample size	Average length (s.d.) n=75	n=100	n=125
$\hat{\overline{Y}}_H$, SRSWOR	312.1 (150.08)	322.48 (121.86)	326.36 (93.707)
${}^4\hat{\overline{Y}}_{PEML}$, SRSWOR	243.23 (65.059)	216.42 (55.256)	198.11 (44.972)
$\hat{\overline{Y}}_{HT}$, RS	184.98 (24.336)	160.79 (17.942)	144.43 (13.89)
$\hat{\overline{Y}}_H$, RS	189.49 (314.18)	163.19 (209.6)	145.82 (164.32)
${}^4\hat{\overline{Y}}_{PEML}$, RS	343.6 (60.804)	300.14 (20.411)	272.63 (21.998)
$\hat{\overline{Y}}_{RHC}$, RHC	277.91 (16.039)	240.09 (12.042)	214.78 (9.2784)
${}^4\hat{\overline{Y}}_{PEML}$, RHC	279.97 (52.788)	242.43 (58.394)	217.09 (21.356)

TABLE 2.32: Average and s.d. of lengths of asymptotically 95% CIs for variance of y_2. Recall from Table 2.5 in Section 2.1 that for variance of y_2, $h(y_2)$=(y_2^2, y_2) and $g(s_1, s_2)$=$s_1 - s_2^2$.

Estimator and sampling design based on which CI is constructed \ Sample size	Average length (s.d.) n=75	n=100	n=125
$g(\hat{\bar{h}}_H)$, SRSWOR	1498664 (3236118)	1588740 (2694726)	2418155 (3205532)
$g(\hat{\bar{h}}_{PEML})$, SRSWOR	1035032 (1472036)	1077345 (1376947)	1002397 (1573834)
$g(\hat{\bar{h}}_H)$, RS	887813.9 (464853)	764055.6 (377760)	684218.5 (298552)
$g(\hat{\bar{h}}_{PEML})$, RS	1385778 (1584677)	1168689 (1339377)	1055339 (1177054)
$g(\hat{\bar{h}}_{PEML})$, RHC	1319413 (1473379)	1134532 (1384754)	1072290 (1472584)

TABLE 2.33: Average and s.d. of lengths of asymptotically 95% CIs for correlation coefficient between y_1 and y_3. Recall from Table 2.5 in Section 2.1 that for correlation coefficient between y_1 and y_3, $h(y_1, y_3)$=$(y_1, y_3, y_1^2, y_3^2, y_1y_3)$ and $g(s_1, s_2, s_3, s_4, s_5)$=$(s_5 - s_1s_2)/((s_3 - s_1^2)(s_4 - s_2^2))^{1/2}$.

Estimator and sampling design based on which CI is constructed \ Sample size	Average length (s.d.) n=75	n=100	n=125
$g(\hat{\bar{h}}_H)$, SRSWOR	0.3682 (0.1138)	0.3753 (0.1039)	0.3893 (0.0936)
$g(\hat{\bar{h}}_{PEML})$, SRSWOR	0.2747 (0.1095)	0.2881 (0.1008)	0.2884 (0.0879)
$g(\hat{\bar{h}}_H)$, RS	0.3351 (0.1652)	0.3453 (0.0938)	0.3587 (0.1034)
$g(\hat{\bar{h}}_{PEML})$, RS	592.48 (0.2859)	260.44 (0.3441)	469.36 (2.738)
$g(\hat{\bar{h}}_{PEML})$, RHC	3838.4 (1.2271)	2740.5 (0.1467)	2238.3 (0.1104)

TABLE 2.34: Average and s.d. of lengths of asymptotically 95% CIs for regression coefficient of y_1 on y_3. Recall from Table 2.5 in Section 2.1 that for regression coefficient of y_1 on y_3, $h(y_1, y_3)$=$(y_1, y_3, y_3^2, y_1 y_3)$ and $g(s_1, s_2, s_3, s_4)$=$(s_4 - s_1 s_2)/(s_3 - s_2^2)$.

	Average length (s.d.)		
Sample size / Estimator and sampling design based on which CI is constructed	n=75	n=100	n=125
$g(\hat{\bar{h}}_H)$, SRSWOR	1.6443 (1.223)	1.781 (1.127)	1.8077 (0.8849)
$g(\hat{\bar{h}}_{PEML})$, SRSWOR	1.3984 (0.8867)	1.4239 (0.7898)	1.491 (0.6645)
$g(\hat{\bar{h}}_H)$, RS	1.4072 (0.6463)	1.5299 (0.4833)	1.5449 (0.4883)
$g(\hat{\bar{h}}_{PEML})$, RS	3240.4 (4.3202)	4938.4 (1.659)	1705.3 (2.017)
$g(\hat{\bar{h}}_{PEML})$, RHC	50701.7 (2.659)	17291.2 (3.93)	22245.7 (1.51)

TABLE 2.35: Average and s.d. of lengths of asymptotically 95% CIs for regression coefficient of y_3 on y_1. Recall from Table 2.5 in Section 2.1 that for regression coefficient of y_3 on y_1, $h(y_1, y_3)$=$(y_3, y_1, y_1^2, y_1 y_3)$ and $g(s_1, s_2, s_3, s_4)$=$(s_4 - s_1 s_2)/(s_3 - s_2^2)$.

	Average length (s.d.)		
Sample size / Estimator and sampling design based on which CI is constructed	n=75	n=100	n=125
$g(\hat{\bar{h}}_H)$, SRSWOR	0.1387 (0.091)	0.1449 (0.072)	0.1508 (0.0616)
$g(\hat{\bar{h}}_{PEML})$, SRSWOR	0.1015 (0.0868)	0.0994 (0.0692)	0.1002 (0.0593)
$g(\hat{\bar{h}}_H)$, RS	0.1305 (0.0919)	0.1379 (0.0438)	0.1447 (0.0357)
$g(\hat{\bar{h}}_{PEML})$, RS	113.4 (0.1712)	263.23 (0.0725)	78.782 (0.0545)
$g(\hat{\bar{h}}_{PEML})$, RHC	798.95 (0.6227)	490.91 (0.0862)	286.92 (0.1107)

TABLE 2.36: Average and s.d. of lengths of asymptotically 95% CIs for correlation coefficient between y_2 and y_4. Recall from Table 2.5 in Section 2.1 that for correlation coefficient between y_2 and y_4, $h(y_2, y_4)$=$(y_2, y_4, y_2^2, y_4^2, y_2y_4)$ and $g(s_1, s_2, s_3, s_4, s_5)$=$(s_5 - s_1s_2)/((s_3 - s_1^2)(s_4 - s_2^2))^{1/2}$.

	Average length (s.d.)		
Sample size / Estimator and sampling design based on which CI is constructed	n=75	n=100	n=125
$g(\hat{\overline{h}}_H)$, SRSWOR	0.3428 (0.191)	0.359 (0.1783)	0.3821 (0.1844)
$g(\hat{\overline{h}}_{PEML})$, SRSWOR	0.3088 (0.1886)	0.3279 (0.171)	0.3537 (0.1773)
$g(\hat{\overline{h}}_H)$, RS	0.2924 (0.1561)	0.2926 (0.1491)	0.298 (0.1568)
$g(\hat{\overline{h}}_{PEML})$, RS	833.87 (0.5226)	300.13 (0.4406)	242.51 (0.8658)
$g(\hat{\overline{h}}_{PEML})$, RHC	7593.1 (0.4385)	3526.1 (0.4869)	2390.9 (0.2661)

TABLE 2.37: Average and s.d. of lengths of asymptotically 95% CIs for regression coefficient of y_2 on y_4. Recall from Table 2.5 in Section 2.1 that for regression coefficient of y_2 on y_4, $h(y_2, y_4)$=$(y_2, y_4, y_4^2, y_2y_4)$ and $g(s_1, s_2, s_3, s_4)$=$(s_4 - s_1s_2)/(s_3 - s_2^2)$.

	Average length (s.d.)		
Sample size / Estimator and sampling design based on which CI is constructed	n=75	n=100	n=125
$g(\hat{\overline{h}}_H)$, SRSWOR	1.1188 (1.251)	1.1117 (1.061)	1.1566 (1.171)
$g(\hat{\overline{h}}_{PEML})$, SRSWOR	0.9865 (0.9935)	1.0005 (0.8784)	1.0534 (0.8758)
$g(\hat{\overline{h}}_H)$, RS	0.8575 (0.6472)	0.847 (0.5219)	0.8427 (0.4524)
$g(\hat{\overline{h}}_{PEML})$, RS	1583.8 (1.733)	1647.2 (1.822)	1533.9 (1.302)
$g(\hat{\overline{h}}_{PEML})$, RHC	24127.4 (2.05)	10798.8 (1.468)	5076.1 (2.385)

TABLE 2.38: Average and s.d. of lengths of asymptotically 95% CIs for regression coefficient of y_4 on y_2. Recall from Table 2.5 in Section 2.1 that for regression coefficient of y_4 on y_2, $h(y_2, y_4)=(y_4, y_2, y_2^2, y_2 y_4)$ and $g(s_1, s_2, s_3, s_4)=(s_4 - s_1 s_2)/(s_3 - s_2^2)$.

Estimator and sampling design based on which CI is constructed \ Sample size	Average length (s.d.)		
	n=75	n=100	n=125
$g(\hat{\bar{h}}_H)$, SRSWOR	0.1607 (0.2236)	0.1727 (0.2175)	0.1682 (0.1744)
$g(\hat{\bar{h}}_{PEML})$, SRSWOR	0.1456 (0.2018)	0.1586 (0.1868)	0.1577 (0.1616)
$g(\hat{\bar{h}}_H)$, RS	0.1219 (0.0798)	0.1232 (0.0663)	0.1273 (0.0615)
$g(\hat{\bar{h}}_{PEML})$, RS	236.81 (0.3529)	108.3 (0.1879)	85.466 (0.3227)
$g(\hat{\bar{h}}_{PEML})$, RHC	1568.1 (0.4045)	2215.1 (0.197)	659.3 (0.1416)

2.4. Comparison of estimators with their bias-corrected versions

In this section, we empirically compare the biased estimators considered in Table 2.6 in Section 2.3 with their bias-corrected versions based on both synthetic and real data used in Section 2.3. Following the idea in [80], we compute the bias-corrected jackknife estimator corresponding to each of the biased estimators considered in Table 2.6. For the mean, we compute the bias-corrected jackknife estimators corresponding to the GREG and the PEML estimators under each of SRSWOR, RS and RHC sampling designs, and the Hájek estimator under RS sampling design. On the other hand, for each of the variance, the correlation coefficient and the regression coefficient, we consider the bias-corrected jackknife estimators corresponding to the estimators that are obtained by plugging in the Hájek and the PEML estimators under each of SRSWOR and RS sampling design, and the PEML estimator under RHC sampling design.

Suppose that s is a sample of size n drawn using one of the sampling designs given in Table 2.6. Further, suppose that s_{-i} is the subset of s, which excludes the i^{th} unit for any given $i \in s$. Now, for any $i \in s$, let us denote the estimator $g(\hat{\bar{h}})$ constructed based on s_{-i} by $g(\hat{\bar{h}}_{-i})$. Then, we compute the bias-corrected jackknife estimator of $g(\bar{h})$ corresponding to $g(\hat{\bar{h}})$ as $ng(\hat{\bar{h}}) - (n-1)\sum_{i \in s} g(\hat{\bar{h}}_{-i})/n$ (cf. [80]). Recall from Section 2.3 that we draw I=1000 samples each of sizes n=75, 100 and 125 from some synthetic as well as real data sets using sampling designs mentioned in Table 2.6 and compute MSEs of the estimators considered in Table 2.6 based on these samples. Here, we compute MSEs of the above-mentioned bias-corrected jackknife estimators using the same procedure and compare them with the original biased estimators in terms of MSE. We observe from the above analyses that for all the parameters considered in Section 2.3, the bias-corrected jackknife estimators become worse than the original biased estimators in the cases of both the synthetic and the real data (see Tables 2.39–2.53 below). Despite reducing the biases of the original biased estimators, bias-correction increases the variances of these estimators significantly. This is the reason why the bias-corrected jackknife estimators have larger MSEs than the original biased estimators in the cases of both the synthetic and the real data.

TABLE 2.39: Relative efficiencies of estimators for mean of y in the case of synthetic data.

Relative efficiency \ Sample size	n=75	n=100	n=125
RE($\hat{\bar{Y}}_{PEML}$, SRSWOR | [5]$\hat{\bar{Y}}_{BCPEML}$, SRSWOR)	1.050461	1.021275	1.038282
RE($\hat{\bar{Y}}_{GREG}$, SRSWOR | [5]$\hat{\bar{Y}}_{BCGREG}$, SRSWOR)	1.002649	1.003156	1.005397
RE($\hat{\bar{Y}}_{H}$, RS | [5]$\hat{\bar{Y}}_{BCH}$, RS)	1.036379	1.006945	1.12841
RE($\hat{\bar{Y}}_{PEML}$, RS | [5]$\hat{\bar{Y}}_{BCPEML}$, RS)	1.016953	1.013402	1.011762
RE($\hat{\bar{Y}}_{GREG}$, RS | [5]$\hat{\bar{Y}}_{BCGREG}$, RS)	1.016692	1.011597	1.011493
RE($\hat{\bar{Y}}_{PEML}$, RHC | [5]$\hat{\bar{Y}}_{BCPEML}$, RHC)	1.01914	1.02292	1.024689
RE($\hat{\bar{Y}}_{GREG}$, RHC | [5]$\hat{\bar{Y}}_{BCGREG}$, RHC)	1.011583	1.052311	1.023058

[5] BCPEML=Bias-corrected PEML estimator, BCH=Bias-corrected Hájek estimator, and BCGREG=Bias-corrected GREG estimator.

TABLE 2.40: Relative efficiencies of estimators for variance of y in the case of synthetic data. Recall from Table 2.5 in Section 2.1 that for variance of y, $h(y)$=(y^2, y) and $g(s_1, s_2)$=$s_1 - s_2^2$.

Relative efficiency \ Sample size	n=75	n=100	n=125
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | [6]BC $g(\hat{\bar{h}}_{PEML})$, SRSWOR)	1.0208	1.01	1.0669
RE($g(\hat{\bar{h}}_{H})$, SRSWOR | [6]BC $g(\hat{\bar{h}}_{H})$, SRSWOR)	38.642	50.009	65.398
RE($g(\hat{\bar{h}}_{H})$, RS | [6]BC $g(\hat{\bar{h}}_{H})$, RS)	1.0029	1.0117	1.074
RE($g(\hat{\bar{h}}_{PEML})$, RS | [6]BC $g(\hat{\bar{h}}_{PEML})$, RS)	1.0112	1.023	1.0377
RE($g(\hat{\bar{h}}_{PEML})$, RHC | [6]BC $g(\hat{\bar{h}}_{PEML})$, RHC)	1.0141	1.015	1.0126

[6] BC=Bias-corrected.

TABLE 2.41: Relative efficiencies of estimators for correlation coefficient between z_1 and z_2 in the case of synthetic data. Recall from Table 2.5 in Section 2.1 that for correlation coefficient between z_1 and z_2, $h(z_1, z_2)$=$(z_1, z_2, z_1^2, z_2^2, z_1 z_2)$ and $g(s_1, s_2, s_3, s_4, s_5)$=$(s_5 - s_1 s_2)/((s_3 - s_1^2)(s_4 - s_2^2))^{1/2}$.

Relative efficiency \ Sample size	n=75	n=100	n=125
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | [6]BC $g(\hat{\bar{h}}_{PEML})$, SRSWOR)	89.989	95.299	123.89
RE($g(\hat{\bar{h}}_{H})$, SRSWOR | [6]BC $g(\hat{\bar{h}}_{H})$, SRSWOR)	90.407	96.79	141.989
RE($g(\hat{\bar{h}}_{H})$, RS | [6]BC $g(\hat{\bar{h}}_{H})$, RS)	90.037	102.914	152.993
RE($g(\hat{\bar{h}}_{PEML})$, RS | [6]BC $g(\hat{\bar{h}}_{PEML})$, RS)	95.68	98.758	158.832
RE($g(\hat{\bar{h}}_{PEML})$, RHC | [6]BC $g(\hat{\bar{h}}_{PEML})$, RHC)	86.27	120.582	125.374

TABLE 2.42: Relative efficiencies of estimators for regression coefficient of z_1 on z_2 in the case of synthetic data. Recall from Table 2.5 in Section 2.1 that for regression coefficient of z_1 on z_2, $h(z_1, z_2)$=$(z_1, z_2, z_2^2, z_1 z_2)$ and $g(s_1, s_2, s_3, s_4)$=$(s_4 - s_1 s_2)/(s_3 - s_2^2)$.

Relative efficiency \ Sample size	n=75	n=100	n=125
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | [6]BC $g(\hat{\bar{h}}_{PEML})$, SRSWOR)	80.64	91.707	124.476
RE($g(\hat{\bar{h}}_{H})$, SRSWOR | [6]BC $g(\hat{\bar{h}}_{H})$, SRSWOR)	79.298	89.105	123.042
RE($g(\hat{\bar{h}}_{RS})$, RS | [6]BC $g(\hat{\bar{h}}_{H})$, RS)	85.97	96.22	135.449
RE($g(\hat{\bar{h}}_{PEML})$, RS | [6]BC $g(\hat{\bar{h}}_{PEML})$, RS)	83.331	97.583	125.657
RE($g(\hat{\bar{h}}_{PEML})$, RHC | [6]BC $g(\hat{\bar{h}}_{PEML})$, RHC)	75.343	112.619	115.594

TABLE 2.43: Relative efficiencies of estimators for regression coefficient of z_2 on z_1 in the case of synthetic data. Recall from Table 2.5 in Section 2.1 that for regression coefficient of z_2 on z_1, $h(z_1, z_2)$=$(z_2, z_1, z_1^2, z_1 z_2)$ and $g(s_1, s_2, s_3, s_4)$=$(s_4 - s_1 s_2)/(s_3 - s_2^2)$.

Relative efficiency \ Sample size	n=75	n=100	n=125
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | [6]BC $g(\hat{\bar{h}}_{PEML})$, SRSWOR)	72.061	105.389	111.124
RE($g(\hat{\bar{h}}_{H})$, SRSWOR | [6]BC $g(\hat{\bar{h}}_{H})$, SRSWOR)	69.114	108.837	118.675
RE($g(\hat{\bar{h}}_{H})$, RS | [6]BC $g(\hat{\bar{h}}_{H})$, RS)	69.16	115.113	144.811
RE($g(\hat{\bar{h}}_{PEML})$, RS | [6]BC $g(\hat{\bar{h}}_{PEML})$, RS)	72.448	127.387	131.558
RE($g(\hat{\bar{h}}_{PEML})$, RHC | [6]BC $g(\hat{\bar{h}}_{PEML})$, RHC)	90.132	104.121	148.139

TABLE 2.44: Relative efficiencies of estimators for mean of y_1 in the case of real data.

Sample size / Relative efficiency	n=75	n=100	n=125
RE($\hat{\bar{Y}}_{PEML}$, SRSWOR | [5]$\hat{\bar{Y}}_{BCPEML}$, SRSWOR)	1.070226	1.019958	1.007533
RE($\hat{\bar{Y}}_{GREG}$, SRSWOR | [5]$\hat{\bar{Y}}_{BCGREG}$, SRSWOR)	1.146007	1.116225	1.117507
RE($\hat{\bar{Y}}_{H}$, RS | [5]$\hat{\bar{Y}}_{BCH}$, RS)	1.240493	1.012969	1.155246
RE($\hat{\bar{Y}}_{PEML}$, RS | [5]$\hat{\bar{Y}}_{BCPEML}$, RS)	1.374578	1.046986	1.055930
RE($\hat{\bar{Y}}_{GREG}$, RS | [5]$\hat{\bar{Y}}_{BCGREG}$, RS)	1.466647	1.138300	1.205053
RE($\hat{\bar{Y}}_{PEML}$, RHC | [5]$\hat{\bar{Y}}_{BCPEML}$, RHC)	1.566827	1.083589	1.132790
RE($\hat{\bar{Y}}_{GREG}$, RHC | [5]$\hat{\bar{Y}}_{BCGREG}$, RHC)	1.460886	1.037045	1.028358

TABLE 2.45: Relative efficiencies of estimators for variance of y_1 in the case of real data. Recall from Table 2.5 in Section 2.1 that for variance of y_1, $h(y_1)$=(y_1^2, y_1) and $g(s_1, s_2)$=$s_1 - s_2^2$.

Sample size / Relative efficiency	n=75	n=100	n=125
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | [6]BC $g(\hat{\bar{h}}_{PEML})$, SRSWOR)	1.1812	1.2736	1.8669
RE($g(\hat{\bar{h}}_{H})$, SRSWOR | [6]BC $g(\hat{\bar{h}}_{H})$, SRSWOR)	4.3526	4.8948	6.0349
RE($g(\hat{\bar{h}}_{H})$, RS | [6]BC $g(\hat{\bar{h}}_{H})$, RS)	1.115	1.1239	1.2269
RE($g(\hat{\bar{h}}_{PEML})$, RS | [6]BC $g(\hat{\bar{h}}_{PEML})$, RS)	1.4373	1.1739	1.6481
RE($g(\hat{\bar{h}}_{PEML})$, RHC | [6]BC $g(\hat{\bar{h}}_{PEML})$, RHC)	1.8502	1.0186	1.0384

TABLE 2.46: Relative efficiencies of estimators for mean of y_2 in the case of real data.

Sample size / Relative efficiency	n=75	n=100	n=125
RE($\hat{\bar{Y}}_{H}$, RS | [5]$\hat{\bar{Y}}_{BCH}$, RS)	1.252123	1.325047	1.241809
RE($\hat{\bar{Y}}_{PEML}$, RS | [5]$\hat{\bar{Y}}_{BCPEML}$, RS)	1.988105	2.146357	2.260343
RE($\hat{\bar{Y}}_{GREG}$, RS | [5]$\hat{\bar{Y}}_{BCGREG}$, RS)	2.055588	2.018015	2.287817
RE($\hat{\bar{Y}}_{PEML}$, RHC | [5]$\hat{\bar{Y}}_{BCPEML}$, RHC)	1.831377	2.083210	2.006134
RE($\hat{\bar{Y}}_{GREG}$, RHC | [5]$\hat{\bar{Y}}_{BCGREG}$, RHC)	1.925938	1.983984	2.091003
RE($\hat{\bar{Y}}_{PEML}$, SRSWOR | [5]$\hat{\bar{Y}}_{BCPEML}$, SRSWOR)	1.001786	1.004973	1.060588
RE($\hat{\bar{Y}}_{GREG}$, SRSWOR | [5]$\hat{\bar{Y}}_{BCGREG}$, SRSWOR)	1.021103	1.008525	1.003390

TABLE 2.47: Relative efficiencies of estimators for variance of y_2 in the case of real data. Recall from Table 2.5 in Section 2.1 that for variance of y_2, $h(y_2)$=(y_2^2, y_2) and $g(s_1, s_2)$=$s_1 - s_2^2$.

Relative efficiency \ Sample size	n=75	n=100	n=125
RE($g(\hat{\bar{h}}_H)$, RS | ^{6}BC $g(\hat{\bar{h}}_H)$, RS)	13.301	6.3589	33.579
RE($g(\hat{\bar{h}}_{PEML})$, RS | ^{6}BC $g(\hat{\bar{h}}_{PEML})$, RS)	4.448	7.4621	7.989
RE($g(\hat{\bar{h}}_{PEML})$, RHC | ^{6}BC $g(\hat{\bar{h}}_{PEML})$, RHC)	21.855	3.0076	11.368
RE($g(\hat{\bar{h}}_H)$, SRSWOR | ^{6}BC $g(\hat{\bar{h}}_H)$, SRSWOR)	8.7641	5.6119	13.7
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | ^{6}BC $g(\hat{\bar{h}}_{PEML})$, SRSWOR)	6.2655	2.0015	6.959

TABLE 2.48: Relative efficiencies of estimators for correlation coefficient between y_1 and y_3 in the case of real data. Recall from Table 2.5 in Section 2.1 that for correlation coefficient between y_1 and y_3, $h(y_1, y_3)$=$(y_1, y_3, y_1^2, y_3^2, y_1y_3)$ and $g(s_1, s_2, s_3, s_4, s_5)$=$(s_5 - s_1s_2)/((s_3 - s_1^2)(s_4 - s_2^2))^{1/2}$.

Relative efficiency \ Sample size	n=75	n=100	n=125
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | ^{6}BC $g(\hat{\bar{h}}_{PEML})$, SRSWOR)	23.149	51.887	45.976
RE($g(\hat{\bar{h}}_H)$, SRSWOR | ^{6}BC $g(\hat{\bar{h}}_H)$, SRSWOR)	90.769	163.74	154.97
RE($g(\hat{\bar{h}}_H)$, RS | ^{6}BC $g(\hat{\bar{h}}_H)$, RS)	72.604	79.355	163.03
RE($g(\hat{\bar{h}}_{PEML})$, RS | ^{6}BC $g(\hat{\bar{h}}_{PEML})$, RS)	24.483	35.874	43.164
RE($g(\hat{\bar{h}}_{PEML})$, RHC | ^{6}BC $g(\hat{\bar{h}}_{PEML})$, RHC)	29.189	65.949	43.13

TABLE 2.49: Relative efficiencies of estimators for regression coefficient of y_1 on y_3 in the case of real data. Recall from Table 2.5 in Section 2.1 that for regression coefficient of y_1 on y_3, $h(y_1, y_3)$=$(y_1, y_3, y_3^2, y_1y_3)$ and $g(s_1, s_2, s_3, s_4)$=$(s_4 - s_1s_2)/(s_3 - s_2^2)$.

Relative efficiency \ Sample size	n=75	n=100	n=125
RE($g(\hat{\bar{h}}_{PEML})$, SRSWOR | ^{6}BC $g(\hat{\bar{h}}_{PEML})$, SRSWOR)	31.789	50.26	50.107
RE($g(\hat{\bar{h}}_H)$, SRSWOR | ^{6}BC $g(\hat{\bar{h}}_H)$, SRSWOR)	236.49	119.88	222.23
RE($g(\hat{\bar{h}}_H)$, RS | ^{6}BC $g(\hat{\bar{h}}_H)$, RS)	63.933	77.049	184.45
RE($g(\hat{\bar{h}}_{PEML})$, RS | ^{6}BC $g(\hat{\bar{h}}_{PEML})$, RS)	31.503	44.945	263.5
RE($g(\hat{\bar{h}}_{PEML})$, RHC | ^{6}BC $g(\hat{\bar{h}}_{PEML})$, RHC)	65.145	76.533	90.413

TABLE 2.50: Relative efficiencies of estimators for regression coefficient of y_3 on y_1 in the case of real data. Recall from Table 2.5 in Section 2.1 that for regression coefficient of y_3 on y_1, $h(y_1, y_3)$=$(y_3, y_1, y_1^2, y_1y_3)$ and $g(s_1, s_2, s_3, s_4)$=$(s_4 - s_1s_2)/(s_3 - s_2^2)$.

Relative efficiency \ Sample size	n=75	n=100	n=125
RE($g(\hat{\overline{h}}_{PEML})$, SRSWOR | ^{6}BC $g(\hat{\overline{h}}_{PEML})$, SRSWOR)	26.09	29.557	32.345
RE($g(\hat{\overline{h}}_{H})$, SRSWOR | ^{6}BC $g(\hat{\overline{h}}_{H})$, SRSWOR)	98.43	104.19	165.95
RE($g(\hat{\overline{h}}_{H})$, RS | ^{6}BC $g(\hat{\overline{h}}_{H})$, RS)	100.3	110.15	196.34
RE($g(\hat{\overline{h}}_{PEML})$, RS | ^{6}BC $g(\hat{\overline{h}}_{PEML})$, RS)	11.416	71.664	23.433
RE($g(\hat{\overline{h}}_{PEML})$, RHC | ^{6}BC $g(\hat{\overline{h}}_{PEML})$, RHC)	13.268	28.198	50.571

TABLE 2.51: Relative efficiencies of estimators for correlation coefficient between y_2 and y_4 in the case of real data. Recall from Table 2.5 in Section 2.1 that for correlation coefficient between y_2 and y_4, $h(y_2, y_4)$=$(y_2, y_4, y_2^2, y_4^2, y_2y_4)$ and $g(s_1, s_2, s_3, s_4, s_5)$=$(s_5 - s_1s_2)/((s_3 - s_1^2)(s_4 - s_2^2))^{1/2}$.

Relative efficiency \ Sample size	n=75	n=100	n=125
RE($g(\hat{\overline{h}}_{H})$, RS | ^{6}BC $g(\hat{\overline{h}}_{PEML})$, RS)	79.092	58.241	120.229
RE($g(\hat{\overline{h}}_{PEML})$, RS | ^{6}BC $g(\hat{\overline{h}}_{PEML})$, RS)	82.309	61.995	316.929
RE($g(\hat{\overline{h}}_{PEML})$, RHC | ^{6}BC $g(\hat{\overline{h}}_{PEML})$, RHC)	175.22	74.847	220.74
RE($g(\hat{\overline{h}}_{H})$, SRSWOR | ^{6}BC $g(\hat{\overline{h}}_{H})$, SRSWOR)	87.942	36.363	97.432
RE($g(\hat{\overline{h}}_{PEML})$, SRSWOR | ^{6}BC $g(\hat{\overline{h}}_{PEML})$, SRSWOR)	120.02	51.959	121.42

TABLE 2.52: Relative efficiencies of estimators for regression coefficient of y_2 on y_4 in the case of real data. Recall from Table 2.5 in Section 2.1 that for regression coefficient of y_2 on y_4, $h(y_2, y_4)$=$(y_2, y_4, y_4^2, y_2y_4)$ and $g(s_1, s_2, s_3, s_4)$=$(s_4 - s_1s_2)/(s_3 - s_2^2)$.

RE($g(\hat{\overline{h}}_{H})$, RS | ^{6}BC $g(\hat{\overline{h}}_{H})$, RS)	125.17	256.45	260.15
RE($g(\hat{\overline{h}}_{PEML})$, RS | ^{6}BC $g(\hat{\overline{h}}_{PEML})$, RS)	145.1	333.5	135.65
RE($g(\hat{\overline{h}}_{PEML})$, RHC | ^{6}BC $g(\hat{\overline{h}}_{PEML})$, RHC)	86.93	238.32	292.89
RE($g(\hat{\overline{h}}_{PEML})$, SRSWOR | ^{6}BC $g(\hat{\overline{h}}_{PEML})$, SRSWOR)	93.707	101.93	121.44
RE($g(\hat{\overline{h}}_{H})$, SRSWOR | ^{6}BC $g(\hat{\overline{h}}_{H})$, SRSWOR)	115.85	146.16	104.66

TABLE 2.53: Relative efficiencies of estimators for regression coefficient of y_4 on y_2 in the case of real data. Recall from Table 2.5 in Section 2.1 that for regression coefficient of y_4 on y_2, $h(y_2, y_4)$=$(y_4, y_2, y_2^2, y_2y_4)$ and $g(s_1, s_2, s_3, s_4)$=$(s_4 - s_1s_2)/(s_3 - s_2^2)$.

Relative efficiency \ Sample size	n=75	n=100	n=125
RE($g(\hat{\bar{h}}_H)$, RS | ^{6}BC $g(\hat{\bar{h}}_H)$, RS)	47.3317	73.749	52.592
RE($g(\hat{\bar{h}}_H)$, RS | ^{6}BC $g(\hat{\bar{h}}_{PEML})$, RS)	105.87	126.42	323.82
RE($g(\hat{\bar{h}}_H)$, RS | ^{6}BC $g(\hat{\bar{h}}_{PEML})$, RHC)	93.403	79.453	91.347
RE($g(\hat{\bar{h}}_H)$, RS | ^{6}BC $g(\hat{\bar{h}}_{PEML})$, SRSWOR)	530.94	173.19	191.26
RE($g(\hat{\bar{h}}_H)$, RS | ^{6}BC $g(\hat{\bar{h}}_H)$, SRSWOR)	394.29	156.27	164.7

2.5. Concluding discussion and remarks

It follows from Theorem 2.2.1 that the PEML estimator of the mean under SRSWOR becomes asymptotically either more efficient than or equivalent to any other estimator under any other sampling design considered in this chapter. It also follows from Theorems 2.1.1 and 2.1.2 that the GREG estimator of the mean is asymptotically equivalent to the PEML estimator under different sampling designs considered in this chapter. However, our numerical studies (see Section 2.3) based on finite samples indicate that the PEML estimator of the mean performs slightly better than the GREG estimator under all the sampling designs considered in Section 2.3 (see Tables 2.7, 2.19 and 2.21). Moreover, as pointed out in the 5^{th} paragraph in the beginning of this chapter, if the estimators of the variance, the correlation coefficient and the regression coefficient are constructed by plugging in the GREG estimator of the mean, then the estimators of the population variances involved in these parameters may become negative. On the other hand, if the estimators of these parameters are constructed by plugging in the PEML estimator of the mean, then such a problem does not occur. Further, for these parameters, depending on sampling designs, the plug-in estimator based on either the PEML or the Hájek estimator turns out to be asymptotically best among different estimators that we have considered (see Theorems 2.2.3 and 2.2.4).

We see from Theorem 2.2.1 that for the population mean, the PEML estimator, which is not design unbiased, performs better than design unbiased estimators like the HT and the RHC estimators. Further, as pointed out in the beginning of this chapter, the plug-in estimators of the population variance based on the HT and the RHC estimators may become negative. This affects the plug-in estimators of the correlation and the regression coefficients based on the HT and the RHC estimators.

It follows from Table 2.3 that under LMS sampling design, the large sample performances of all the estimators of functions of means considered in this chapter are the same as their large sample performances under SRSWOR. The LMS sampling design was introduced to make the ratio estimator of the mean unbiased. It follows from Remark 2.1.2 in Section 2.1 that the performance of the ratio estimator of the mean is worse than several other estimators that we have considered even under LMS sampling design.

The coefficient of variation is another well-known finite population parameter, which can be expressed as a function of population means $g(\overline{h})$. We have d=1, p=2, $h(y)$=(y^2, y) and $g(s_1, s_2)$=$\sqrt{s_1 - s_2^2}/s_2$ in this case. Among the estimators considered in this chapter, the plug-in estimators of $g(\overline{h})$ that are based on the PEML and the Hájek estimators of the mean can be used for estimating this parameter since it involves the finite population variance (see the 5^{th} paragraph in the beginning of this chapter). We have avoided reporting the comparison of the estimators of the coefficient of variation in this chapter because of complex mathematical expressions. However, the asymptotic results stated in Theorems 2.2.3 and 2.2.4 also hold for this parameter.

In sample survey, sometimes we deal with stratified sampling designs (see [24]) in which the population is divided into H (> 1) strata and a sample is drawn from each stratum by a sampling design independently across the strata. For a stratified population, the population mean of y can be expressed as $\overline{Y}$=$\sum_{l=1}^{H}(N_l/N)\overline{Y}_l$, where N_l is the number of population units in the l^{th} stratum and $\overline{Y}_l$ is the mean of y for the l^{th} stratum. Further, N=$\sum_{l=1}^{H} N_l$. Therefore, an estimator of $\overline{Y}$ under a stratified sampling design is obtained as $\hat{\overline{Y}}$=$\sum_{l=1}^{H}(N_l/N)\hat{\overline{Y}}_l$, where $\hat{\overline{Y}}_l$ is the HT, the RHC, the Hájek, the ratio, the product, the GREG or the PEML estimator of $\overline{Y}_l$ constructed based on the sample drawn from the l^{th} stratum. Also, several plug in estimators of a function of population means $g(\overline{h})$ can be constructed under a stratified sampling design following the approach of this chapter. Suppose that H is fixed as $\nu \to \infty$, the assumptions of Theorems 2.1.1–2.1.3 and Remarks 2.1.1–2.1.2 hold in each stratum, and $\lim_{\nu\to\infty}(N_l/N)$=$\Lambda_l$ for some $0 < \Lambda_l < 1$, l=$1, \ldots, l$. Then, conclusions of the aforementioned results hold for estimators of $g(\overline{h})$ under stratified sampling design.

An empirical comparison of the biased estimators considered in this chapter and their bias-corrected versions are carried out based on jackknifing in Section 2.4 in terms of their MSEs.

It follows from this comparison that for all the parameters considered in this chapter, the bias-corrected estimators become worse than the original biased estimators in the cases of both the synthetic and the real data. This is because, although bias-correction results in reduction of biases in the original biased estimators, the variances of these estimators increase substantially after bias-correction.

2.6. Proofs of the main results

In this section, we give the proofs of Theorems 2.1.1–2.1.3 and 2.2.2–2.2.4, and Remark 2.1.1. Let us denote the HT, the RHC, the Hájek, the ratio, the product, the GREG and the PEML estimators of population means of $h(y)$ by $\hat{\overline{h}}_{HT}$, $\hat{\overline{h}}_{RHC}$, $\hat{\overline{h}}_{H}$, $\hat{\overline{h}}_{RA}$, $\hat{\overline{h}}_{PR}$, $\hat{\overline{h}}_{GREG}$ and $\hat{\overline{h}}_{PEML}$, respectively.

Proof of Theorem 2.1.1. Let us consider SRSWOR and LMS sampling design. It follows from *(i)* in Lemma 2.7.4 in Section 2.7 that $\sqrt{n}(\hat{\overline{h}} - \overline{h}) \xrightarrow{\mathcal{L}} N(0, \Gamma)$ as $\nu \to \infty$ for some p.d. matrix Γ, when $\hat{\overline{h}}$ is one of $\hat{\overline{h}}_{HT}$, $\hat{\overline{h}}_{H}$, $\hat{\overline{h}}_{RA}$, $\hat{\overline{h}}_{PR}$, and $\hat{\overline{h}}_{GREG}$ with $d(i,s)$=$(N\pi_i)^{-1}$ under any of these sampling designs. Now, note that $\max_{i\in s}|X_i - \overline{X}|$=$o_p(\sqrt{n})$, and $\sum_{i\in s}\pi_i^{-1}(X_i - \overline{X})/\sum_{i\in s}\pi_i^{-1}(X_i-\overline{X})^2$=$O_p(1/\sqrt{n})$ as $\nu \to \infty$ under the above sampling designs (see Lemma 2.7.7 in Section 2.7). Then, by applying Theorem 1 of [22] to each real-valued coordinate of $\hat{\overline{h}}_{PEML}$ and $\hat{\overline{h}}_{GREG}$, we get $\sqrt{n}(\hat{\overline{h}}_{PEML} - \hat{\overline{h}}_{GREG})$=$o_p(1)$ as $\nu \to \infty$ for $d(i,s)$=$(N\pi_i)^{-1}$ under these sampling designs. This implies that $\hat{\overline{h}}_{PEML}$ and $\hat{\overline{h}}_{GREG}$ with $d(i,s)$=$(N\pi_i)^{-1}$ have the same asymptotic distribution. Therefore, if $\hat{\overline{h}}$ is one of $\hat{\overline{h}}_{HT}$, $\hat{\overline{h}}_{H}$, $\hat{\overline{h}}_{RA}$, $\hat{\overline{h}}_{PR}$, and $\hat{\overline{h}}_{GREG}$ and $\hat{\overline{h}}_{PEML}$ with $d(i,s)$=$(N\pi_i)^{-1}$, we have

$$\sqrt{n}(g(\hat{\overline{h}}) - g(\overline{h})) \xrightarrow{\mathcal{L}} N(0, \Delta^2) \text{ as } \nu \to \infty \tag{2.6.1}$$

under any of the above mentioned sampling designs for some $\Delta^2 > 0$ by the delta method and the assumption $\nabla g(\mu_0) \neq 0$ at μ_0=$\lim_{\nu\to\infty}\overline{h}$. It can be shown from the proof of *(i)* in Lemma 2.7.4 in Section 2.7 that Δ^2=$\nabla g(\mu_0)\Gamma_1\,(\nabla g(\mu_0)^T$, where Γ_1=$\lim_{\nu\to\infty} nN^{-2}\sum_{i=1}^{N}(\mathbf{V}_i - \mathbf{T}_V\pi_i)^T(\mathbf{V}_i - \mathbf{T}_V\pi_i)(\pi_i^{-1} - 1)$. It can also be shown from Table 2.54 in Section 2.7 that under each of the above sampling designs, $\mathbf{V}_i$ in Γ_1 is h_i or $h_i - \overline{h}$ or $h_i - \overline{h}X_i/\overline{X}$ or $h_i + \overline{h}X_i/\overline{X}$ or $h_i - \overline{h} - S_{xh}(X_i - \overline{X})/S_x^2$ if $\hat{\overline{h}}$ is $\hat{\overline{h}}_{HT}$ or $\hat{\overline{h}}_{H}$ or $\hat{\overline{h}}_{RA}$ or $\hat{\overline{h}}_{PR}$, or $\hat{\overline{h}}_{GREG}$ with $d(i,s)$=$(N\pi_i)^{-1}$, respectively.

Now, by Lemma *(i)* in 2.7.6 in Section 2.7, we have

$$\sigma_1^2 = \sigma_2^2 = (1-\lambda) \lim_{\nu\to\infty} \sum_{i=1}^{N} (A_i - \bar{A})^2/N, \tag{2.6.2}$$

where σ_1^2 and σ_2^2 are as defined in the statement of Lemma 2.7.6 in Section 2.7, and A_i=$\nabla g(\mu_0)\mathbf{V}_i^T$ for different choices of $\mathbf{V}_i$ mentioned in the preceding paragraph. Note that $g(\hat{\overline{h}}_{GREG})$ and $g(\hat{\overline{h}}_{PEML})$ have the same asymptotic distribution under each of SRSWOR and LMS sampling design since $\sqrt{n}(\hat{\overline{h}}_{PEML} - \hat{\overline{h}}_{GREG})$=$o_p(1)$ for $\nu \to \infty$ under these sampling designs as pointed out earlier in this proof. Further, (2.6.2) implies that $g(\hat{\overline{h}}_{GREG})$ with $d(i,s)$=$(N\pi_i)^{-1}$ has the same asymptotic MSE under SRSWOR and LMS sampling design. Thus $g(\hat{\overline{h}}_{GREG})$ and $g(\hat{\overline{h}}_{PEML})$ with $d(i,s)$=$(N\pi_i)^{-1}$ under SRSWOR and LMS sampling design form class 1 in Table 2.3.

Next, (2.6.2) yields that $g(\hat{\overline{h}}_{HT})$ has the same asymptotic MSE under SRSWOR and LMS sampling design. It also follows from (2.6.2) that $g(\hat{\overline{h}}_{H})$ has the same asymptotic MSE under SRSWOR and LMS sampling design. Now, note that $g(\hat{\overline{h}}_{HT})$ and $g(\hat{\overline{h}}_{H})$ coincide under SRSWOR. Thus $g(\hat{\overline{h}}_{HT})$ under SRSWOR, and $g(\hat{\overline{h}}_{HT})$ and $g(\hat{\overline{h}}_{H})$ under LMS sampling design form class 2 in Table 2.3.

Next, (2.6.2) implies that $g(\hat{\overline{h}}_{RA})$ has the same asymptotic MSE under SRSWOR and LMS sampling design. Further, (2.6.2) implies that $g(\hat{\overline{h}}_{PR})$ has the same asymptotic MSE under SRSWOR and LMS sampling design. Thus $g(\hat{\overline{h}}_{RA})$ under SRSWOR and LMS sampling design forms class 3 in Table 2.3, and $g(\hat{\overline{h}}_{PR})$ under those sampling designs forms class 4 in Table 2.3. This completes the proof of Theorem 2.1.1. □

Proof of Theorem 2.1.2. Let us first consider a HEπPS sampling design. Then, it can be shown in the same way as in the 1^{st} paragraph of the proof of Theorem 2.1.1 that $\sqrt{n}(\hat{\overline{h}}_{PEML} - \hat{\overline{h}}_{GREG})$=$o_p(1)$ for $d(i,s)$=$(N\pi_i)^{-1}$ under this sampling design. It can also be shown in the same way as in the 1^{st} paragraph of the proof of Theorem 2.1.1 that if $\hat{\overline{h}}$ is one of $\hat{\overline{h}}_{HT}$, $\hat{\overline{h}}_{H}$, and $\hat{\overline{h}}_{GREG}$ and $\hat{\overline{h}}_{PEML}$ with $d(i,s)$=$(N\pi_i)^{-1}$, then (2.6.1) holds under the above-mentioned sampling design. Here, we recall from Table 2.3 that the HT, the ratio and the product estimators coincide under any HEπPS sampling design. Further, the asymptotic MSE of $\sqrt{n}(g(\hat{\overline{h}}) - g(\overline{h}))$ is $\nabla g(\mu_0)\Gamma_1\ (\nabla g(\mu_0))^T$, where μ_0=$\lim_{\nu\to\infty}\overline{h}$, Γ_1=$\lim_{\nu\to\infty} nN^{-2}\sum_{i=1}^{N}(\mathbf{V}_i - \mathbf{T}_V\pi_i)^T(\mathbf{V}_i - \mathbf{T}_V\pi_i)(\pi_i^{-1}-1)$, and $\mathbf{V}_i$ in Γ_1 is h_i or $h_i - \overline{h}$ or $h_i - \overline{h} - S_{xh}(X_i - \overline{X})/S_x^2$ if $\hat{\overline{h}}$ is $\hat{\overline{h}}_{HT}$ or $\hat{\overline{h}}_{H}$, or $\hat{\overline{h}}_{GREG}$ with $d(i,s)$=$(N\pi_i)^{-1}$, respectively. Now, since $\sqrt{n}(\hat{\overline{h}}_{PEML} - \hat{\overline{h}}_{GREG})$=$o_p(1)$

for $\nu \to \infty$ under any HEπPS sampling design, $g(\hat{\bar{h}}_{GREG})$ and $g(\hat{\bar{h}}_{PEML})$ have the same asymptotic distribution under this sampling design. Thus under any HEπPS sampling design, $g(\hat{\bar{h}}_{GREG})$ and $g(\hat{\bar{h}}_{PEML})$ with $d(i,s)$=$(N\pi_i)^{-1}$ form class 5, $g(\hat{\bar{h}}_{HT})$ forms class 6, and $g(\hat{\bar{h}}_{H})$ forms class 7 in Table 2.3. This completes the proof of *(i)* in Theorem 2.1.2.

Let us now consider the RHC sampling design. We can show from *(ii)* in Lemma 2.7.4 in Section 2.7 that $\sqrt{n}(\hat{\bar{h}} - \bar{h}) \xrightarrow{\mathcal{L}} N(0, \Gamma)$ as $\nu \to \infty$ for some p.d. matrix Γ, when $\hat{\bar{h}}$ is either $\hat{\bar{h}}_{RHC}$ or $\hat{\bar{h}}_{GREG}$ with $d(i,s)$=$(NX_i)^{-1}G_i$ under RHC sampling design. Further, $\sqrt{n}(\hat{\bar{h}}_{PEML} - \hat{\bar{h}}_{GREG})$=$o_p(1)$ as $\nu \to \infty$ for $d(i,s)$=$(NX_i)^{-1}G_i$ under RHC sampling design since Assumption 2.1.3 holds, and S_x^2 is bounded away from 0 as $\nu \to \infty$ (see A2.2 of Appendix 2 in [22]). Therefore, if $\hat{\bar{h}}$ is one of $\hat{\bar{h}}_{RHC}$, and $\hat{\bar{h}}_{GREG}$ and $\hat{\bar{h}}_{PEML}$ with $d(i,s)$=$(NX_i)^{-1}G_i$, then we have

$$\sqrt{n}(g(\hat{\bar{h}}) - g(\bar{h})) \xrightarrow{\mathcal{L}} N(0, \Delta^2) \text{ as } \nu \to \infty \tag{2.6.3}$$

for some $\Delta^2 > 0$ by the delta method and the condition $\nabla g(\mu_0) \neq 0$ at μ_0=$\lim_{\nu\to\infty} \bar{h}$. Moreover, it follows from the proof of *(ii)* in Lemma 2.7.4 in Section 2.7 that Δ^2=$\nabla g(\mu_0)\Gamma_2(\nabla g(\mu_0))^T$, where Γ_2=$\lim_{\nu\to\infty} n\gamma\overline{X}N^{-1}\sum_{i=1}^{N}(\mathbf{V}_i - X_i\overline{\mathbf{V}}/\overline{X})^T(\mathbf{V}_i - X_i\overline{\mathbf{V}}/\overline{X})/X_i$. It further follows from Table 2.54 in Section 2.7 that $\mathbf{V}_i$ in Γ_2 is h_i if $\hat{\bar{h}}$ is $\hat{\bar{h}}_{RHC}$. Also, $\mathbf{V}_i$ in Γ_2 is $h_i - \bar{h} - S_{xh}(X_i - \overline{X})/S_x^2$ if $\hat{\bar{h}}$ is $\hat{\bar{h}}_{GREG}$ with $d(i,s)$=$(NX_i)^{-1}G_i$. Now, $g(\hat{\bar{h}}_{GREG})$ and $g(\hat{\bar{h}}_{PEML})$ have the same asymptotic distribution under RHC sampling design since $\sqrt{n}(\hat{\bar{h}}_{PEML} - \hat{\bar{h}}_{GREG})$=$o_p(1)$ for $\nu \to \infty$ under this sampling design as pointed out earlier in this paragraph. Thus $g(\hat{\bar{h}}_{GREG})$ and $g(\hat{\bar{h}}_{PEML})$ with $d(i,s)$=$(NX_i)^{-1}G_i$ under RHC sampling design form class 8, and $g(\hat{\bar{h}}_{RHC})$ forms class 9 in Table 2.3. This completes the proof of *(ii)* in Theorem 2.1.2. □

Proof of Remark 2.1.1. It follows from *(ii)* in Lemma 2.7.6 in Section 2.7 that in the case of λ=0,

$$\sigma_3^2 = \sigma_4^2 = \lim_{\nu\to\infty}((\overline{X}/N)\sum_{i=1}^{N} A_i^2/X_i - \bar{A}^2), \tag{2.6.4}$$

where σ_1^3 and σ_2^4 are as defined in the statement of Lemma 2.7.6 in Section 2.7, and A_i=$\nabla g(\mu_0)\mathbf{V}_i^T$ for different choices of $\mathbf{V}_i$ mentioned in the proof of Theorem 2.1.2 above. Thus $g(\hat{\bar{h}}_{GREG})$ with $d(i,s)$=$(N\pi_i)^{-1}$ under any HEπPS sampling design, and with $d(i,s)$=$(NX_i)^{-1}G_i$ under RHC sampling design have the same asymptotic MSE. Therefore, class 8 is merged with class 5 in Table 2.3. Further, (2.6.4) implies that $g(\hat{\bar{h}}_{HT})$ under any HEπPS sampling design and $g(\hat{\bar{h}}_{RHC})$ have the same asymptotic MSE. Therefore, class 9 is merged with class 6 in Table 2.3. This completes the proof of Remark 2.1.1. □

Proof of Theorem 2.1.3. Recall the expression of A_i's from the proofs of Theorems 2.1.1 and 2.1.2. Note that $\lim_{\nu\to\infty}\sum(A_i-\bar{A})^2/N$=$\lim_{\nu\to\infty}\sum(B_i-\bar{B})^2/N$, $\lim_{\nu\to\infty} n\gamma\big((\overline{X}/N)\times\sum_{i=1}^{N}A_i^2/X_i-\bar{A}^2\big)$=$\lim_{\nu\to\infty} n\gamma\big((\overline{X}/N)\sum_{i=1}^{N}B_i^2/X_i-\bar{B}^2\big)$ and $\lim_{\nu\to\infty}\big\{(1/N)\sum_{i=1}^{N}A_i^2\times\big((\overline{X}/X_i)-(n/N)\big)-\phi^{-1}\overline{X}^{-1}\big((n/N)\sum_{i=1}^{N}A_iX_i/N-\bar{A}\overline{X}\big)^2\big\}$=$\lim_{\nu\to\infty}\big\{(1/N)\sum_{i=1}^{N}B_i^2\times\big((\overline{X}/X_i)-(n/N)\big)-\phi^{-1}\overline{X}^{-1}\big((n/N)\sum_{i=1}^{N}B_iX_i/N-\bar{B}\,\overline{X}\big)^2\big\}$ for B_i=$\nabla g(\overline{h})\mathbf{V}_i^T$ and $\mathbf{V}_i$ as in Table 2.54 in Section 2.7 since $\nabla g(\overline{h})\to\nabla g(\mu_0)$ as $\nu\to\infty$. Here, ϕ=$\overline{X}-(n/N)\sum_{i=1}^{N}X_i^2/N\overline{X}$. Then, from Lemma 2.7.6 in Section 2.7 and the expressions of asymptotic MSEs of $\sqrt{n}(g(\hat{\overline{h}})-g(\overline{h}))$ discussed in the proofs of Theorems 2.1.1 and 2.1.2, the results in Table 2.4 follow. This completes the proof of Theorem 2.1.3. □

Proof of Theorem 2.2.1. Note that Assumptions 2.1.2 and 2.1.3 hold *a.s.* [**P**] since Assumption 2.2.1 holds and $E_{\mathbf{P}}(\epsilon_i)^4<\infty$. Also, note that Assumption 2.1.4 holds *a.s.* [**P**] under SRSWOR and LMS sampling design (see Lemma 2.7.8 in Section 2.7). Then, under the above sampling designs, conclusions of Theorems 2.1.1 and 2.1.3 hold *a.s.* [**P**] for d=p=1, $h(y)$=y and $g(s)$=s. Note that W_i=$\nabla g(\overline{h})h_i^T$=$Y_i$. Also, note that the Δ_i^2's in Table 2.4 can be expressed in terms of superpopulation moments of (Y_i,X_i) *a.s.* [**P**] by SLLN since $E_{\mathbf{P}}(\epsilon_i)^4<\infty$. Recall from the beginning of Section 2.2 that we have taken σ_x^2=1. Then, we have $\Delta_2^2-\Delta_1^2$=$(1-\lambda)\sigma_{xy}^2$, $\Delta_3^2-\Delta_1^2$=$(1-\lambda)(\sigma_{xy}-E_{\mathbf{P}}(Y_i)/\mu_1)^2$ and $\Delta_4^2-\Delta_1^2$=$(1-\lambda)(\sigma_{xy}+E_{\mathbf{P}}(Y_i)/\mu_1)^2$ *a.s.* [**P**], where μ_1=$E_{\mathbf{P}}(X_i)$ and σ_{xy}=$cov_{\mathbf{P}}(X_i,Y_i)$. Hence, $\Delta_1^2<\Delta_i^2$ *a.s.* [**P**] for i=$2,3,4$. This completes the proof of *(i)* in Theorem 2.2.1.

Next consider the case $0\leq\lambda<E_{\mathbb{P}}(X_i)/b$. Note that $n\gamma\to c$ as $\nu\to\infty$ for some $c\geq 1-\lambda>0$ by Lemma 2.7.5 in Section 2.7. Also, note that *a.s.* [$\mathbb{P}$], Assumption 2.1.5 holds in the case of RHC sampling design and Assumption 2.1.4 holds in the case of any HEπPS sampling design (see Lemma 2.7.8 in Section 2.7). Then, under RHC and any HEπPS sampling designs, conclusions of Theorems 2.1.2 and 2.1.3 hold *a.s.* [**P**] for d=p=1, $h(y)$=y and $g(s)$=s. Further, we have $\Delta_5^2-\Delta_1^2$=$\big\{E_{\mathbb{P}}\big(Y_i-E_{\mathbb{P}}(Y_i)\big)^2\big(\mu_1/X_i-\lambda\big)-\mu_1^2\sigma_{xy}\big(\sigma_{xy}cov_{\mathbb{P}}(X_i,1/X_i)-2cov_{\mathbb{P}}(Y_i,1/X_i)\big)+\lambda\sigma_{xy}^2\big\}-(1-\lambda)\big\{\sigma_y^2-\sigma_{xy}^2\big\}$, $\Delta_6^2-\Delta_5^2$= $E_{\mathbb{P}}\big(Y_i^2\big(\mu_1/X_i-\lambda\big)\big)-\big\{\lambda E_{\mathbb{P}}(Y_iX_i)-E_{\mathbb{P}}(Y_i)\mu_1\big\}^2/\chi\mu_1-\big\{E_{\mathbb{P}}\big(Y_i-E_{\mathbb{P}}(Y_i)-\sigma_{xy}(X_i-\mu_1)\big)^2\big(\mu_1/X_i-\lambda\big)\big\}$, $\Delta_7^2-\Delta_5^2$=$\big\{\mu_1^2\sigma_{xy}\big(\sigma_{xy}cov_{\mathbb{P}}(X_i,1/X_i)-2cov_{\mathbb{P}}(Y_i,1/X_i)\big)-\lambda\sigma_{xy}^2-\lambda^2\sigma_{xy}^2/\mu_1\chi\big\}$, $\Delta_8^2-\Delta_1^2$=$c\big\{\mu_1E_{\mathbb{P}}(Y_i-E_{\mathbb{P}}(Y_i))^2/X_i-\mu_1^2\sigma_{xy}(\sigma_{xy}cov_{\mathbb{P}}(X_i,1/X_i)-2cov_{\mathbb{P}}(Y_i,1/X_i))\big\}-(1-\lambda)\big\{\sigma_y^2-\sigma_{xy}^2\big\}$ and $\Delta_9^2-\Delta_1^2$=$c\big\{\mu_1E_{\mathbb{P}}(Y_i^2/X_i)-E_{\mathbb{P}}^2(Y_i)\big\}-(1-\lambda)\big\{\sigma_y^2-\sigma_{xy}^2\big\}$ *a.s.* [$\mathbb{P}$], where σ_y^2=$var_{\mathbb{P}}(Y_i)$, χ=$\mu_1-\lambda(\mu_2/\mu_1)$ and μ_2=$E_{\mathbb{P}}(X_i)^2$. Here, we note that χ=$E_{\mathbb{P}}\big(X_i^2(\mu_1/X_i-\lambda)\big)/\mu_1>0$ because Assumption 2.2.1 holds and Assumption 2.1.1

holds with $0 \leq \lambda < E_{\mathbb{P}}(X_i)/b$. Moreover, from the linear model set up, we can show that $\Delta_5^2 - \Delta_1^2$=$\sigma^2(\mu_1\mu_{-1} - 1) > 0$, $\Delta_6^2 - \Delta_5^2$=$E_{\mathbb{P}}\big\{(\alpha + \beta X_i) - \chi^{-1}X_i(\alpha + \beta\mu_1 - \lambda\alpha - \lambda\beta\mu_2/\mu_1)\big\}^2\big\{\mu_1/X_i - \lambda\big\} \geq 0$, $\Delta_7^2 - \Delta_5^2$=$\beta^2 E_{\mathbb{P}}\big\{(X_i - \mu_1) - \lambda\chi^{-1}X_i(\mu_1 - \mu_2/\mu_1)\big\}^2\big\{\mu_1/X_i - \lambda\big\} \geq 0$, $\Delta_8^2 - \Delta_1^2$=$\sigma^2\big(c\mu_1\mu_{-1} - (1-\lambda)\big) \geq c\sigma^2(\mu_1\mu_{-1} - 1) > 0$ and $\Delta_9^2 - \Delta_1^2$=$\sigma^2\big(c\mu_1\mu_{-1} - (1-\lambda)\big) + c\alpha^2(\mu_1\mu_{-1} - 1) > 0$ *a.s.* $[\mathbb{P}]$, where σ^2=$E_{\mathbb{P}}(\epsilon_i)^2$. Note that $\Delta_6^2 - \Delta_5^2 \geq 0$ and $\Delta_7^2 - \Delta_5^2 \geq 0$ because Assumption 2.2.1 holds and Assumption 2.1.1 holds with $0 \leq \lambda < E_{\mathbb{P}}(X_i)/b$. Therefore, $\Delta_1^2 < \Delta_i^2$ *a.s.* $[\mathbb{P}]$ for i=$2, \ldots, 9$. This completes the proof of *(ii)* in Theorem 2.2.1. □

Proof of Theorem 2.2.2. The proof follows in a straightforward way from the proof of Theorem 2.2.1. □

Proof of Theorem 2.2.3. Using similar arguments as in the 1^{st} paragraph of proof of Theorem 2.2.1, we can say that under SRSWOR and LMS sampling design, conclusions of Theorems 2.1.1 and 2.1.3 hold *a.s.* $[\mathbf{P}]$ for d=1, p=2, $h(y)$=(y, y^2) and $g(s_1, s_2)$=$s_2 - s_1^2$ in the same way as conclusions of Theorems 2.1.1 and 2.1.3 hold *a.s.* $[\mathbb{P}]$ for d=p=1, $h(y)$=y and $g(s)$=s in the 1^{st} paragraph of the proof of Theorem 2.2.1. Note that W_i=$Y_i^2 - 2Y_i\overline{Y}$ for the above choices of h and g. Further, it follows from SLLN and the assumption $E_{\mathbf{P}}(\epsilon_i)^8 < \infty$ that the Δ_i^2's in Table 2.4 can be expressed in terms of superpopulation moments of (Y_i, X_i) *a.s.* $[\mathbf{P}]$. Note that $\Delta_2^2 - \Delta_1^2$=$cov_{\mathbf{P}}^2(\tilde{W}_i, X_i)$ *a.s.* $[\mathbf{P}]$, where $\tilde{W}_i$=$Y_i^2 - 2Y_iE_{\mathbf{P}}(Y_i)$. Then, $\Delta_1^2 < \Delta_2^2$ *a.s.* $[\mathbf{P}]$. This completes the proof of *(i)* in Theorem 2.2.3.

Next consider the case of $0 \leq \lambda < E_{\mathbb{P}}(X_i)/b$. Using the same line of arguments as in the 2^{nd} paragraph of the proof of Theorem 2.2.1, it can be shown that under RHC and any HEπPS sampling designs, conclusions of Theorems 2.1.2 and 2.1.3 hold *a.s.* $[\mathbb{P}]$ for d=1, p=2, $h(y)$=(y, y^2) and $g(s_1, s_2)$=$s_2 - s_1^2$ in the same way as conclusions of Theorems 2.1.2 and 2.1.3 hold *a.s.* $[\mathbb{P}]$ for d=p=1, $h(y)$=y and $g(s)$=s in the 2^{nd} paragraph of the proof of Theorem 2.2.1. Note that $\Delta_7^2 - \Delta_5^2$=$\big\{\mu_1^2 cov_{\mathbb{P}}(\tilde{W}_i, X_i)\big(cov_{\mathbb{P}}(\tilde{W}_i, X_i)cov_{\mathbb{P}}(X_i, 1/X_i) - 2cov_{\mathbb{P}}(\tilde{W}_i, 1/X_i)\big)\big\} - \lambda^2 cov_{\mathbb{P}}^2(\tilde{W}_i, X_i)/\chi\mu_1 - \lambda cov_{\mathbb{P}}^2(\tilde{W}_i, X_i) \leq \big\{\mu_1^2 \times cov_{\mathbb{P}}(\tilde{W}_i, X_i)\big(cov_{\mathbb{P}}(\tilde{W}_i, X_i)cov_{\mathbb{P}}(X_i, 1/X_i) - 2cov_{\mathbb{P}}(\tilde{W}_i, 1/X_i)\big)\big\}$ *a.s.* $[\mathbb{P}]$ because $\chi > 0$. Recall from Assumption 2.2.2 that ξ=$\mu_3 - \mu_2\mu_1$ and μ_j=$E_{\mathbb{P}}(X_i)^j$ for j=$-1, 1, 2, 3$. Then, from the linear model set up, we have $\big\{\mu_1^2 cov_{\mathbb{P}}(\tilde{W}_i, X_i) \times \big(cov_{\mathbb{P}}(\tilde{W}_i, X_i)cov_{\mathbb{P}}(X_i, 1/X_i) - 2cov_{\mathbb{P}}(\tilde{W}_i, 1/X_i)\big)\big\}$=$(\beta^2\mu_1)^2(\xi - 2\mu_1)((\xi + 2\mu_1)\zeta_1 - 2\zeta_2)$. Here, ζ_1=$1 - \mu_1\mu_{-1}$ and ζ_2=$\mu_1 - \mu_2\mu_{-1}$. Note that $(\xi + 2\mu_1)\zeta_1 - 2\zeta_2$=$\xi\zeta_1 + 2\mu_{-1}$ and $\zeta_1 < 0$. Therefore, $\big\{\mu_1^2 cov_{\mathbb{P}}(\tilde{W}_i, X_i)\big(cov_{\mathbb{P}}(\tilde{W}_i, X_i)cov_{\mathbb{P}}(X_i, 1/X_i) - 2cov_{\mathbb{P}}(\tilde{W}_i, 1/X_i)\big)\big\} < 0$ if $\xi > 2\max\{\mu_1, \mu_{-1}/(\mu_1\mu_{-1} - 1)\}$. Hence, $\Delta_7^2 - \Delta_5^2 < 0$ *a.s.* $[\mathbb{P}]$. This completes the proof of *(ii)* in Theorem 2.2.3. □

Proof of Theorem 2.2.4. Using the same line of arguments as in the 1^{st} paragraph of the proof of Theorem 2.2.1, it can be shown that under SRSWOR and LMS sampling design, conclusions of Theorems 2.1.1 and 2.1.3 hold *a.s.* $[\mathbb{P}]$ for d=2, p=5, $h(z_1, z_2)$=$(z_1, z_2, z_1^2, z_2^2, z_1 z_2)$ and $g(s_1, s_2, s_3, s_4, s_5)$=$(s_5 - s_1 s_2)/((s_3 - s_1^2)(s_4 - s_2^2))^{1/2}$ in the case of the correlation coefficient between z_1 and z_2, and for d=2, p=4, $h(z_1, z_2)$=$(z_1, z_2, z_2^2, z_1 z_2)$ and $g(s_1, s_2, s_3, s_4)$= $(s_4 - s_1 s_2)/(s_3 - s_2^2)$ in the case of the regression coefficient of z_1 on z_2 in the same way as conclusions of Theorems 2.1.1 and 2.1.3 hold *a.s.* $[\mathbb{P}]$ for d=p=1, $h(y)$=y and $g(s)$=s in the case of the mean of y in the 1^{st} paragraph of the proof of Theorem 2.2.1. Further, if Assumption 2.1.1 holds with $0 \leq \lambda < E_{\mathbb{P}}(X_i)/b$, then using similar arguments as in the 2^{nd} paragraph of the proof of Theorem 2.2.1, it can also be shown that under RHC and any HEπPS sampling designs, conclusions of Theorems 2.1.2 and 2.1.3 hold *a.s.* $[\mathbb{P}]$ for d=2, p=5, $h(z_1, z_2)$=$(z_1, z_2, z_1^2, z_2^2, z_1 z_2)$ and $g(s_1, s_2, s_3, s_4, s_5)$=$(s_5 - s_1 s_2)/((s_3 - s_1^2)(s_4 - s_2^2))^{1/2}$ in the case of the correlation coefficient between z_1 and z_2, and for d=2, p=4, $h(z_1, z_2)$=$(z_1, z_2, z_2^2, z_1 z_2)$ and $g(s_1, s_2, s_3, s_4)$=$(s_4 - s_1 s_2)/(s_3 - s_2^2)$ in the case of the regression coefficient of z_1 on z_2 in the same way as conclusions of Theorems 2.1.2 and 2.1.3 hold *a.s.* $[\mathbb{P}]$ for d=p=1, $h(y)$=y and $g(s)$=s in the case of the mean of y in the 2^{nd} paragraph of the proof of Theorem 2.2.1. Note that W_i=$R_{12}[(\overline{Z}_1/S_1^2 - \overline{Z}_2/S_{12})Z_{1i} + (\overline{Z}_2/S_2^2 - \overline{Z}_1/S_{12})Z_{2i} - Z_{1i}^2/2S_1^2 - Z_{2i}^2/2S_2^2 + Z_{1i}Z_{2i}/S_{12}]$ for the correlation coefficient, and W_i=$(1/S_2^2)[-\overline{Z}_2 Z_{1i} - (\overline{Z}_1 - 2S_{12}\overline{Z}_2/S_2^2)Z_{2i} - S_{12}Z_{2i}^2/S_2^2 + Z_{1i}Z_{2i}]$ for the regression coefficient. Here, $\overline{Z}_1$=$\sum_{i=1}^{N} Z_{1i}/N$, $\overline{Z}_2$=$\sum_{i=1}^{N} Z_{2i}/N$, S_1^2=$\sum_{i=1}^{N} Z_{1i}^2/N - \overline{Z}_1^2$, S_2^2=$\sum_{i=1}^{N} Z_{2i}^2/N - \overline{Z}_2^2$, S_{12}=$\sum_{i=1}^{N} Z_{1i} Z_{2i}/N - \overline{Z}_1\overline{Z}_2$ and R_{12}=$S_{12}/S_1 S_2$. Also, note that since $E_{\mathbb{P}}||\epsilon_i||^8 < \infty$, the Δ_i^2's in Table 2.4 can be expressed in terms of superpopulation moments of $(h(Z_{1i}, Z_{2i}), X_i)$ *a.s.* $[\mathbb{P}]$ for both the parameters by SLLN. Further, for the above parameters, we have $\Delta_2^2 - \Delta_1^2$=$cov_{\mathbb{P}}^2(\tilde{W}_i, X_i) > 0$ and $\Delta_7^2 - \Delta_5^2$=$\big\{\mu_1^2 cov_{\mathbb{P}}(\tilde{W}_i, X_i)\big(cov_{\mathbb{P}}(\tilde{W}_i, X_i)cov_{\mathbb{P}}(X_i, 1/X_i) - 2\times cov_{\mathbb{P}}(\tilde{W}_i, 1/X_i)\big)\big\} - \lambda^2 cov_{\mathbb{P}}^2(\tilde{W}_i, X_i)/\chi\mu_1 - \lambda cov_{\mathbb{P}}^2(\tilde{W}_i, X_i) \leq \big\{\mu_1^2 cov_{\mathbb{P}}(\tilde{W}_i, X_i)\big(cov_{\mathbb{P}}(\tilde{W}_i, X_i)\times cov_{\mathbb{P}}(X_i, 1/X_i) - 2cov_{\mathbb{P}}(\tilde{W}_i, 1/X_i)\big)\big\}$ *a.s.* $[\mathbb{P}]$, where $\tilde{W}_i$ is the same as W_i with all finite population moments in the expression of W_i replaced by their corresponding superpopulation moments. Also, from the linear model set up, we have $\big\{\mu_1^2 cov_{\mathbb{P}}(\tilde{W}_i, X_i)\big(cov_{\mathbb{P}}(\tilde{W}_i, X_i)cov_{\mathbb{P}}(X_i, 1/X_i) - 2cov_{\mathbb{P}}(\tilde{W}_i, 1/X_i)\big)\big\}$=$K_1(\xi - 2\mu_1)((\xi + 2\mu_1)\zeta_1 - 2\zeta_2)$ for some constant $K_1 > 0$ in the case of the correlation coefficient, and $\big\{\mu_1^2 cov_{\mathbb{P}}(\tilde{W}_i, X_i)\big(cov_{\mathbb{P}}(\tilde{W}_i, X_i)cov_{\mathbb{P}}(X_i, 1/X_i) - 2cov_{\mathbb{P}}(\tilde{W}_i, 1/X_i)\big)\big\}$ =$K_2(\xi - 2\mu_1)((\xi + 2\mu_1)\zeta_1 - 2\zeta_2)$ for some constant $K_2 > 0$ in the case of the regression coefficient. Thus proofs of both the parts of the theorem follow in the same way as the proof of Theorem 2.2.3. □

2.7. Proofs of additional results required to prove the main results

In this section, we state and prove some lemmas, which are required to prove Theorems 2.1.1–2.1.3 and 2.2.2–2.2.4, and Remark 2.1.1.

Lemma 2.7.1. *Suppose that Assumption 2.1.3 holds. Then, LMS sampling design is a high entropy sampling design. Moreover, under each of SRSWOR, LMS and any HEπPS sampling designs, there exist constants $L, L' > 0$ such that*

$$L \leq \min_{1\leq i\leq N}(N\pi_i/n) \leq \max_{1\leq i\leq N}(N\pi_i/n) \leq L' \tag{2.7.1}$$

for all sufficiently large ν .

The condition (2.7.1) was considered earlier in [85], [7], etc. However, the above authors did not discuss whether LMS and HEπPS sampling designs satisfy (2.7.1) or not.

Proof. Suppose that $P(s)$ and $R(s)$ denote LMS sampling design and SRSWOR, respectively. Note that SRSWOR is a rejective sampling design. Then, $P(s)$=$(\overline{x}/\overline{X})/{}^N C_n$ and $R(s)$=$({}^N C_n)^{-1}$, where $\overline{x}$=$\sum_{i\in s} X_i/n$ and $s \in \mathcal{S}$. By Cauchy-Schwarz inequality, we have

$$D(P||R) = E((\overline{x}/\overline{X})\log(\overline{x}/\overline{X})) \leq K_1 E|\overline{x}/\overline{X} - 1| \leq K_1 E(\overline{x}/\overline{X} - 1)^2 \tag{2.7.2}$$

for some $K_1 > 0$ since Assumption 2.1.3 holds, and $\log(x) \leq |x - 1|$ for $x > 0$. Here E denotes the expectation with respect to $R(s)$. Therefore,

$$\begin{aligned} nD(P||R) &\leq K_1(1 - n/N)(N/(N-1))(S_x^2/\overline{X}^2) \leq 2K_1(\sum_{i=1}^{N} X_i^2/N\overline{X}^2) \\ &\leq 2K_1(\max_{1\leq i\leq N} X_i/\min_{1\leq i\leq N} X_i)^2 = O(1) \end{aligned} \tag{2.7.3}$$

as $\nu \to \infty$. Hence, $D(P||R) \to 0$ as $\nu \to \infty$. Thus LMS sampling design is a high entropy sampling design.

Next, note that (2.7.1) holds trivially under SRSWOR. Now, suppose that $\{\pi_i\}_{i=1}^N$ denote inclusion probabilities of $P(s)$. Then, we have π_i=$(n-1)/(N-1) + (X_i/\sum_{i=1}^N X_i)((N-n)/(N-1))$ and $\pi_i - n/N$=$-(N-n)(N(N-1))^{-1}(X_i/\overline{X} - 1)$. Further,

$$\frac{|\pi_i - n/N|}{n/N} = \frac{N-n}{n(N-1)}\left|\frac{X_i}{\overline{X}} - 1\right| \leq \frac{N-n}{n(N-1)}\left(\frac{\max_{1\leq i\leq N} X_i}{\min_{1\leq i\leq N} X_i} + 1\right). \tag{2.7.4}$$

Therefore, $\max_{1\leq i\leq N}|N\pi_i/n-1|\rightarrow 0$ as $\nu\rightarrow\infty$ by Assumption 2.1.3. Hence, $K_2\leq\min_{1\leq i\leq N}(N\pi_i/n)\leq\max_{1\leq i\leq N}(N\pi_i/n)\leq K_3$ for all sufficiently large ν and some constants $K_2>0$ and $K_3>0$. Thus (2.7.1) holds under LMS sampling design. Further, (2.7.1) holds under any HEπPS sampling design since Assumption 2.1.3 holds. □

Next, consider $\mathbf{V}_i$'s and $\overline{\mathbf{V}}$ as in the paragraph preceding Assumption 2.1.4. Let us define $\hat{\overline{\mathbf{V}}}_1$=$\sum_{i\in s}(N\pi_i)^{-1}\mathbf{V}_i$ and Σ_1=$nN^{-2}\sum_{i=1}^{N}(\mathbf{V}_i-\mathbf{T}_V\pi_i)^T(\mathbf{V}_i-\mathbf{T}_V\pi_i)(\pi_i^{-1}-1)$, where π_i's and $\mathbf{T}_V$ are as in the paragraph preceding Assumption 2.1.4. Let us also define $\hat{\overline{\mathbf{V}}}_2$=$\sum_{i\in s}(NX_i)^{-1}G_i\mathbf{V}_i$ and Σ_2=$n\gamma\overline{X}N^{-1}\sum_{i=1}^{N}(\mathbf{V}_i-X_i\overline{\mathbf{V}}/\overline{X})^T(\mathbf{V}_i-X_i\overline{\mathbf{V}}/\overline{X})/X_i$, where G_i's are as in the paragraph containing Table 2.1, and γ is as in the paragraph preceding Assumption 2.1.4. Now, we state the following Lemma.

Lemma 2.7.2. *Suppose that Assumptions 2.1.1–2.1.4 hold. Then, under SRSWOR, LMS and any HEπPS sampling designs, we have $\sqrt{n}(\hat{\overline{\boldsymbol{V}}}_1-\overline{\boldsymbol{V}})\xrightarrow{\mathcal{L}}N(0,\Gamma_1)$ as $\nu\rightarrow\infty$, where Γ_1=$\lim_{\nu\rightarrow\infty}\Sigma_1$. Further, suppose that Assumptions 2.1.1–2.1.3, 2.1.5 and 2.1.6 hold. Then, we have $\sqrt{n}(\hat{\overline{\boldsymbol{V}}}_2-\overline{\boldsymbol{V}})\xrightarrow{\mathcal{L}}N(0,\Gamma_2)$ as $\nu\rightarrow\infty$ under RHC sampling, where Γ_2=$\lim_{\nu\rightarrow\infty}\Sigma_2$.*

Proof. Note that SRSWOR is a high entropy sampling design since it is a rejective sampling design. It follows from Lemma 2.7.1 that (2.7.1) in Lemma 2.7.1 holds under SRSWOR and any HEπPS sampling design. It also follows from Lemma 2.7.1 that LMS sampling design is a high entropy sampling design, and (2.7.1) holds under this sampling design. Now, fix $\epsilon>0$ and $\mathbf{m}_1\in\mathbb{R}^p$. Suppose that $L(\epsilon,\mathbf{m}_1)$=$(n^{-1}N^2\mathbf{m}_1\Sigma_1\mathbf{m}_1^T)^{-1}\sum_{i\in G(\epsilon,\mathbf{m}_1)}(\mathbf{m}_1(\mathbf{V}_i-\mathbf{T}_V\pi_i)^T)^2(\pi_i^{-1}-1)$ for $G(\epsilon,\mathbf{m}_1)$=$\{1\leq i\leq N:|\mathbf{m}_1(\mathbf{V}_i-\mathbf{T}_V\pi_i)^T|>\epsilon\pi_iN(n^{-1}\mathbf{m}_1\Sigma_1\mathbf{m}_1^T)^{1/2}\}$, $\mathbf{T}_V$=$\sum_{i=1}^{N}\mathbf{V}_i(1-\pi_i)/\sum_{i=1}^{N}\pi_i(1-\pi_i)$ and $\tilde{\mathbf{V}}_i$=$(n/N\pi_i)\mathbf{V}_i-(n/N)\mathbf{T}_V$, i=$1,\ldots,N$. Then, given any $\delta>0$,

$$L(\epsilon,\mathbf{m}_1)\leq(\mathbf{m}_1\Sigma_1\mathbf{m}_1^T)^{-(1+\delta/2)}n^{-\delta/2}\epsilon^{-\delta}N^{-1}\sum_{i=1}^{N}(||\mathbf{m}_1||||\tilde{\mathbf{V}}_i||)^{2+\delta}(N\pi_i/n) \tag{2.7.5}$$

since $|\mathbf{m}_1\tilde{\mathbf{V}}_i^T|/(\sqrt{n}\epsilon(\mathbf{m}_1\Sigma_1\mathbf{m}_1^T)^{1/2})>1$ for any $i\in G(\epsilon,\mathbf{m}_1)$. It follows from Jensen's inequality that

$$\begin{aligned}N^{-1}\sum_{i=1}^{N}||\tilde{\mathbf{V}}_i||^{2+\delta}(N\pi_i/n)\leq 2^{1+\delta}\big(N^{-1}\sum_{i=1}^{N}||\mathbf{V}_i(n/N\pi_i)||^{2+\delta}(N\pi_i/n)+\\ ||(n/N)\mathbf{T}_V||^{2+\delta}\big)\end{aligned} \tag{2.7.6}$$

since $\sum_{i=1}^{N}\pi_i$=n. It also follows from Assumptions 2.1.2 and 2.1.3, and Jensen's inequality that $\sum_{i=1}^{N}||\mathbf{V}_i||^{2+\delta}/N$=$O(1)$ as $\nu\rightarrow\infty$ for any $0<\delta\leq 2$. Further, $\sum_{i=1}^{N}\pi_i(1-\pi_i)/n$ is bounded away from 0 as $\nu\rightarrow\infty$ under SRSWOR, LMS and any HEπPS sampling designs because (2.7.1) holds under these sampling designs, and Assumption 2.1.1 holds. Therefore,

$$N^{-1}\sum_{i=1}^{N}||\mathbf{V}_i(n/N\pi_i)||^{2+\delta}(N\pi_i/n)=O(1) \text{ and } ||(n/N)\mathbf{T}_V||^{2+\delta}=O(1), \tag{2.7.7}$$

and hence $N^{-1}\sum_{i=1}^{N}||\tilde{\mathbf{V}}_i||^{2+\delta}(N\pi_i/n)$=$O(1)$ as $\nu\rightarrow\infty$ under the above sampling designs. Then, $L(\epsilon,\mathbf{m}_1)\rightarrow 0$ as $\nu\rightarrow\infty$ for any $\epsilon>0$ under all of these sampling designs since Assumption 2.1.4 holds. Therefore, $\inf\{\epsilon>0: L(\epsilon,\mathbf{m}_1)\leq\epsilon\}\rightarrow 0$ as $\nu\rightarrow\infty$, and consequently the Hájek-Lindeberg condition holds for $\{\mathbf{m}_1\mathbf{V}_i^T\}_{i=1}^{N}$ under each of the above sampling designs. Also, $\sum_{i=1}^{N}\pi_i(1-\pi_i)\rightarrow\infty$ as $\nu\rightarrow\infty$ under these sampling designs. Then, from Theorem 5 in [4], $\sqrt{n}\mathbf{m}_1(\hat{\overline{\mathbf{V}}}_1-\overline{\mathbf{V}})^T\xrightarrow{\mathcal{L}}N(0,\mathbf{m}_1\Gamma_1\mathbf{m}_1^T)$ as $\nu\rightarrow\infty$ under each of the above sampling designs for any $\mathbf{m}_1\in\mathbb{R}^p$ and Γ_1=$\lim_{\nu\rightarrow\infty}\Sigma_1$. Hence, $\sqrt{n}(\hat{\overline{\mathbf{V}}}_1-\overline{\mathbf{V}})\xrightarrow{\mathcal{L}}N(0,\Gamma_1)$ as $\nu\rightarrow\infty$ under the above-mentioned sampling designs.

Next, define

$$L(\mathbf{m}_1)=n\gamma\Big(\max_{1\leq i\leq N}X_i\Big)\Big(N^{-1}\sum_{r=1}^{n}\tilde{N}_r^3(\tilde{N}_r-1)\sum_{i=1}^{N}(\mathbf{m}_1(\mathbf{V}_i\overline{X}/X_i-\overline{\mathbf{V}})^T)^4\times X_i)^{1/2}(\overline{X}^{3/2}\sum_{r=1}^{n}\tilde{N}_r(\tilde{N}_r-1)\mathbf{m}_1\Sigma_2\mathbf{m}_1^T\Big)^{-1}, \tag{2.7.8}$$

where γ=$\sum_{r=1}^{n}\tilde{N}_r(\tilde{N}_r-1)/N(N-1)$ as before. Note that as $\nu\rightarrow\infty$,

$$\Big(N^{-1}\sum_{i=1}^{N}(\mathbf{m}_1(\mathbf{V}_i\overline{X}/X_i-\overline{\mathbf{V}})^T)^4(X_i/\overline{X})\Big)^{1/2}=O(1) \text{ and } \overline{X}^{-1}\max_{1\leq i\leq N}X_i=O(1) \tag{2.7.9}$$

since Assumptions 2.1.2 and 2.1.3 hold. Now, under Assumptions 2.1.1 and 2.1.6, we have $(\sum_{r=1}^{n}\tilde{N}_r^3(\tilde{N}_r-1))^{1/2}(\sum_{r=1}^{n}\tilde{N}_r(\tilde{N}_r-1))^{-1}$=$O(1/\sqrt{n})$ and $n\gamma$=$O(1)$ as $\nu\rightarrow\infty$. Therefore, $L(\mathbf{m}_1)\rightarrow 0$ as $\nu\rightarrow\infty$ since Assumption 2.1.5 holds. This implies that the condition C1 in [61] holds for $\{\mathbf{m}_1\mathbf{V}_i^T\}_{i=1}^{N}$. Therefore, by Theorem 2.1 in [61], $\sqrt{n}\mathbf{m}_1(\hat{\overline{\mathbf{V}}}_2-\overline{\mathbf{V}})^T\xrightarrow{\mathcal{L}}N(0,\mathbf{m}_1\Gamma_2\mathbf{m}_1^T)$ as $\nu\rightarrow\infty$ under RHC sampling design for any $\mathbf{m}_1\in\mathbb{R}^p$ and Γ_2=$\lim_{\nu\rightarrow\infty}\Sigma_2$. Hence, $\sqrt{n}(\hat{\overline{\mathbf{V}}}_2-\overline{\mathbf{V}})\xrightarrow{\mathcal{L}}N(0,\Gamma_2)$ as $\nu\rightarrow\infty$ under RHC sampling design. □

Next, suppose that $\overline{\mathbf{C}}$=$\sum_{i=1}^{N}\mathbf{C}_i/N$, $\hat{\overline{\mathbf{C}}}_1$=$\sum_{i\in s}(N\pi_i)^{-1}\mathbf{C}_i$ and $\hat{\overline{\mathbf{C}}}_2$= $\sum_{i\in s}(NX_i)^{-1}G_i\mathbf{C}_i$ for $\mathbf{C}_i$=(h_i, X_ih_i, X_i^2), i=$1,\ldots,N$. Let us also define $\hat{\overline{X}}_1$=$\sum_{i\in s}(N\pi_i)^{-1}X_i$. Now, we state the following lemma.

Lemma 2.7.3. *Suppose that Assumptions 2.1.1–2.1.3 and 2.1.6 hold. Then, under SRSWOR, LMS and any HEπPS sampling designs, we have* $\hat{\overline{\mathbf{C}}}_1-\overline{\mathbf{C}}=o_p(1)$, $\sqrt{n}(\hat{\overline{X}}_1-\overline{X})=O_p(1)$ *and* $\sqrt{n}(\sum_{i\in s}(N\pi_i)^{-1}-1)=O_p(1)$ *as* $\nu\to\infty$*. Moreover, under RHC sampling design, we have* $\hat{\overline{\mathbf{C}}}_2-\overline{\mathbf{C}}=o_p(1)$ *and* $\sqrt{n}(\sum_{i\in s}(NX_i)^{-1}G_i-1)=O_p(1)$ *as* $\nu\to\infty$*.*

Proof. We first show that as $\nu\to\infty$, $\hat{\overline{\mathbf{C}}}_1-\overline{\mathbf{C}}=o_p(1)$, $\sqrt{n}(\hat{\overline{X}}_1-\overline{X})=O_p(1)$ and $\sqrt{n}(\sum_{i\in s}(N\pi_i)^{-1}-1)=O_p(1)$ under a high entropy sampling design $P(s)$ satisfying (2.7.1) in Lemma 2.7.1. Fix $\mathbf{m}_2\in\mathbb{R}^{2p+1}$. Suppose that $Q(s)$ is a rejective sampling design with inclusion probabilities equal to those of $P(s)$ (cf. [4]). Under $Q(s)$, $var(\mathbf{m}_2(\sqrt{n}(\hat{\overline{\mathbf{C}}}_1-\overline{\mathbf{C}})^T))$=$\mathbf{m}_2(nN^{-2}\sum_{i=1}^N(\mathbf{C}_i-\mathbf{T}_C\pi_i)^T(\mathbf{C}_i-\mathbf{T}_C\pi_i)(\pi_i^{-1}-1))\mathbf{m}_2^T(1+e)$ (see Theorem 6.1 in [40]), where $\mathbf{T}_C$=$\sum_{i=1}^N\mathbf{C}_i\,(1-\pi_i)/\sum_{i=1}^N\pi_i(1-\pi_i)$, and $e\to 0$ as $\nu\to\infty$ whenever $\sum_{i=1}^N\pi_i(1-\pi_i)\to\infty$ as $\nu\to\infty$. Note that (2.7.1) holds under $Q(s)$, and hence $\sum_{i=1}^N\pi_i(1-\pi_i)\to\infty$ as $\nu\to\infty$ under $Q(s)$ because (2.7.1) holds under $P(s)$, and Assumption 2.1.1 holds. Then, $\mathbf{m}_2(nN^{-2}\sum_{i=1}^N(\mathbf{C}_i-\mathbf{T}_C\pi_i)^T(\mathbf{C}_i-\mathbf{T}_C\pi_i)(\pi_i^{-1}-1))\mathbf{m}_2^T\leq nN^{-2}\sum_{i=1}^N(\mathbf{m}_2\mathbf{C}_i^T)^2/\pi_i$=$O(1)$ under $Q(s)$ since Assumption 2.1.2 holds. Therefore, $\sqrt{n}(\hat{\overline{\mathbf{C}}}_1-\overline{\mathbf{C}})$=$O_p(1)$ as $\nu\to\infty$ under $Q(s)$ since $var(\mathbf{m}_2(\sqrt{n}(\hat{\overline{\mathbf{C}}}_1-\overline{\mathbf{C}})^T))$=$O(1)$ as $\nu\to\infty$ for any $\mathbf{m}_2\in\mathbb{R}^{2p+1}$ under $Q(s)$. Now, $\sum_{s\in E}P(s)\leq\sum_{s\in E}Q(s)+\sum_{s\in\mathcal{S}}|P(s)-Q(s)|\leq\sum_{s\in E}Q(s)+(2D(P||Q))^{1/2}\leq\sum_{s\in E}Q(s)+(2D(P||R))^{1/2}$ (see Lemmas 2 and 3 in [4]), where E=$\{s\in\mathcal{S}:||\sqrt{n}(\hat{\overline{\mathbf{C}}}_1-\overline{\mathbf{C}})||>\delta\}$ for $\delta>0$ and $R(s)$ is any other rejective sampling design. Let us consider a rejective sampling design $R(s)$ such that $D(P||R)\to 0$ as $\nu\to\infty$. Therefore, given any $\epsilon>0$, there exists a $\delta>0$ such that $\sum_{s\in E}P(s)\leq\epsilon$ for all sufficiently large ν. Hence, as $\nu\to\infty$, $\sqrt{n}(\hat{\overline{\mathbf{C}}}_1-\overline{\mathbf{C}})$=$O_p(1)$ and $\hat{\overline{\mathbf{C}}}_1-\overline{\mathbf{C}}$=$o_p(1)$ under $P(s)$. Similarly, we can show that as $\nu\to\infty$, $\sqrt{n}(\hat{\overline{X}}_1-\overline{X})$=$O_p(1)$ and $\sqrt{n}(\sum_{i\in s}(N\pi_i)^{-1}-1)$=$O_p(1)$ under $P(s)$. Now, recall from the proof of Lemma 2.7.2 that SRSWOR and LMS sampling design are high entropy sampling designs, and they satisfy (2.7.1). Also, any HEπPS sampling design satisfies (2.7.1). Therefore, as $\nu\to\infty$, $\hat{\overline{\mathbf{C}}}_1-\overline{\mathbf{C}}$=$o_p(1)$, $\sqrt{n}(\hat{\overline{X}}_1-\overline{X})$=$O_p(1)$ and $\sqrt{n}(\sum_{i\in s}(N\pi_i)^{-1}-1)$=$O_p(1)$ under the above-mentioned sampling designs.

Under RHC sampling design, $var(\mathbf{m}_2(\sqrt{n}(\hat{\overline{\mathbf{C}}}_2-\overline{\mathbf{C}})^T))$=$\mathbf{m}_2(n\gamma\overline{X}N^{-1}\sum_{i=1}^N(\mathbf{C}_i-X_i\overline{\mathbf{C}}/\overline{X})^T(\mathbf{C}_i-X_i\overline{\mathbf{C}}/\overline{X})/X_i)\mathbf{m}_2^T$ (see [61]). Recall from the proof of Lemma 2.7.2 that $n\gamma$=$O(1)$ as $\nu\to\infty$. Then, $var(\mathbf{m}_2(\sqrt{n}(\hat{\overline{\mathbf{C}}}_2-\overline{\mathbf{C}})^T))\leq n\gamma(\overline{X}/N)\sum_{i=1}^N(\mathbf{m}_2\mathbf{C}_i^T)^2/X_i$=$O(1)$ as $\nu\to\infty$ since Assumptions 2.1.2, 2.1.3 and 2.1.6 hold. Hence, as $\nu\to\infty$, $\sqrt{n}(\hat{\overline{\mathbf{C}}}_2-\overline{\mathbf{C}})$=$O_p(1)$

and $\hat{\overline{\mathbf{C}}}_2-\overline{\mathbf{C}}=o_p(1)$ under RHC sampling design. Similarly, we can show that as $\nu\rightarrow\infty$, $\sqrt{n}(\sum_{i\in s}(NX_i)^{-1}G_i-1)=O_p(1)$ under RHC sampling design. $\square$

Recall from the 1^{st} paragraph in Section 2.6 that we denote the HT, the RHC, the Hájek, the ratio, the product, the GREG and the PEML estimators of population means of $h(y)$ by $\hat{\overline{h}}_{HT}$, $\hat{\overline{h}}_{RHC}$, $\hat{\overline{h}}_H$, $\hat{\overline{h}}_{RA}$, $\hat{\overline{h}}_{PR}$, $\hat{\overline{h}}_{GREG}$ and $\hat{\overline{h}}_{PEML}$, respectively. Suppose that $\hat{\overline{h}}$ denotes one of $\hat{\overline{h}}_{HT}$, $\hat{\overline{h}}_H$, $\hat{\overline{h}}_{RA}$, $\hat{\overline{h}}_{PR}$, and $\hat{\overline{h}}_{GREG}$ with $d(i,s)=(N\pi_i)^{-1}$. Then, a Taylor type expansion of $\hat{\overline{h}}-\overline{h}$ can be obtained as $\hat{\overline{h}}-\overline{h}=\Theta(\hat{\overline{\mathbf{V}}}_1-\overline{\mathbf{V}})+\mathbf{R}$, where $\hat{\overline{\mathbf{V}}}_1=\sum_{i\in s}(N\pi_i)^{-1}\mathbf{V}_i$, $\overline{\mathbf{V}}=\sum_{i=1}^N\mathbf{V}_i/N$, and $\mathbf{V}_i$'s, Θ and $\mathbf{R}$ are as described in Table 2.54 below. On the other hand, if $\hat{\overline{h}}$ is either $\hat{\overline{h}}_{RHC}$ or $\hat{\overline{h}}_{GREG}$ with

TABLE 2.54: Expressions of $\mathbf{V}_i$, Θ and $\mathbf{R}$ for different $\hat{\overline{h}}$'s.

$\hat{\overline{h}}$	$\mathbf{V}_i$	Θ	$\mathbf{R}$
$\hat{\overline{h}}_{HT}$	h_i	1	0
$\hat{\overline{h}}_H$	$h_i-\overline{h}$	$(\sum_{i\in s}(N\pi_i)^{-1})^{-1}$	0
$\hat{\overline{h}}_{RA}$	$h_i-\overline{h}X_i/\overline{X}$	$\overline{X}/\hat{\overline{X}}_1$	0
$\hat{\overline{h}}_{PR}$	$h_i+\overline{h}X_i/\overline{X}$	$\hat{\overline{X}}_1/\overline{X}$	$-(1-\hat{\overline{X}}_1/\overline{X})^2\overline{h}$
$\hat{\overline{h}}_{GREG}$ with $d(i,s)=(N\pi_i)^{-1}$	$h_i-\overline{h}-$ $S_{xh}(X_i-\overline{X})/S_x^2$	$(\sum_{i\in s}(N\pi_i)^{-1})^{-1}$	$(\hat{\overline{X}}_2-\overline{X})\times$ $(S_{xh}/S_x^2-\hat{\beta}_1)$
$\hat{\overline{h}}_{RHC}$	h_i	1	0
$\hat{\overline{h}}_{GREG}$ with $d(i,s)=(NX_i)^{-1}G_i$	$h_i-\overline{h}-$ $S_{xh}(X_i-\overline{X})/S_x^2$	$(\sum_{i\in s}(NX_i)^{-1}G_i)^{-1}$	$\overline{X}((\sum_{i\in s}(NX_i)^{-1}G_i)^{-1}$ $-1)(S_{xh}/S_x^2-\hat{\beta}_2)$

$d(i,s)=(NX_i)^{-1}G_i$, a Taylor type expansion of $\hat{\overline{h}}-\overline{h}$ can be obtained as $\hat{\overline{h}}-\overline{h}=\Theta(\hat{\overline{\mathbf{V}}}_2-\overline{\mathbf{V}})+\mathbf{R}$. Here, $\hat{\overline{\mathbf{V}}}_2=\sum_{i\in s}(NX_i)^{-1}G_i\mathbf{V}_i$, G_i's are as in the paragraph containing Table 2.1, and the $\mathbf{V}_i$'s, Θ and $\mathbf{R}$ are once again described in Table 2.54. In Table 2.54, $\hat{\overline{X}}_1=\sum_{i\in s}(N\pi_i)^{-1}X_i$, $\hat{\overline{X}}_2=\hat{\overline{X}}_1/\sum_{i\in s}(N\pi_i)^{-1}$, $\hat{\beta}_1=(\sum_{i\in s}(N\pi_i)^{-1}\sum_{i\in s}(N\pi_i)^{-1}h_iX_i-\hat{\overline{h}}_{HT}\hat{\overline{X}}_1)/(\sum_{i\in s}(N\pi_i)^{-1}\times\sum_{i\in s}(N\pi_i)^{-1}X_i^2-(\hat{\overline{X}}_1)^2)$ and $\hat{\beta}_2=(\sum_{i\in s}((NX_i)^{-1}G_i)\sum_{i\in s}(N^{-1}G_ih_i)-\hat{\overline{h}}_{RHC}\overline{X})/(\sum_{i\in s}((NX_i)^{-1}G_i)\sum_{i\in s}(N^{-1}G_iX_i)-\overline{X}^2)$. Now, we state the following lemma.

Lemma 2.7.4. *(i) Suppose that Assumptions 2.1.1–2.1.4 hold. Further, suppose that $\hat{\overline{h}}$ is one of $\hat{\overline{h}}_{HT}$, $\hat{\overline{h}}_H$, $\hat{\overline{h}}_{RA}$, $\hat{\overline{h}}_{PR}$, and $\hat{\overline{h}}_{GREG}$ with $d(i,s)=(N\pi_i)^{-1}$. Then, under SRSWOR, LMS and any HEπPS sampling designs,*

$$\sqrt{n}(\hat{\overline{h}}-\overline{h})\xrightarrow{\mathcal{L}}N(0,\Gamma)\ as\ \nu\rightarrow\infty \qquad (2.7.10)$$

for some p.d. matrix Γ.

(ii) Further, suppose that Assumptions 2.1.1–2.1.3, 2.1.5 and 2.1.6 hold, and $\hat{\overline{h}}$ *is* $\hat{\overline{h}}_{RHC}$ *or* $\hat{\overline{h}}_{GREG}$ *with* $d(i,s)$=$(NX_i)^{-1}G_i$. *Then,* (2.7.10) *holds under RHC sampling design.*

Proof. It can be shown from Lemma 2.7.2 that $\sqrt{n}(\hat{\overline{\mathbf{V}}}_1 - \overline{\mathbf{V}}) \xrightarrow{\mathcal{L}} N(0,\Gamma_1)$ as $\nu \to \infty$ under SRSWOR, LMS and any HEπPS sampling designs, where Γ_1=$\lim_{\nu\to\infty} nN^{-2}\sum_{i=1}^{N}(\mathbf{V}_i - \mathbf{T}_V\pi_i)^T(\mathbf{V}_i - \mathbf{T}_V\pi_i)(\pi_i^{-1}-1)$ with $\mathbf{T}_V$=$\sum_{i=1}^{N}\mathbf{V}_i(1-\pi_i)/\sum_{i=1}^{N}\pi_i(1-\pi_i)$. Note that Γ_1 is a p.d. matrix under each of the above sampling designs as Assumption 2.1.4 holds under these sampling designs. Let us now consider from Table 2.54 various choices of Θ and **R** corresponding to $\hat{\overline{h}}_{HT}$, $\hat{\overline{h}}_{H}$, $\hat{\overline{h}}_{RA}$, $\hat{\overline{h}}_{PR}$, and $\hat{\overline{h}}_{GREG}$ with $d(i,s)$=$(N\pi_i)^{-1}$. Then, it can be shown from Lemma 2.7.3 that for each of these choices, $\sqrt{n}\mathbf{R}$=$o_p(1)$ and $\Theta - 1$=$o_p(1)$ as $\nu \to \infty$ under the above-mentioned sampling designs. Therefore, (2.7.10) holds under those sampling designs with Γ=Γ_1. This completes the proof of *(i)* in Lemma 2.7.4

We can show from Lemma 2.7.2 that $\sqrt{n}(\hat{\overline{\mathbf{V}}}_2 - \overline{\mathbf{V}}) \xrightarrow{\mathcal{L}} N(0,\Gamma_2)$ as $\nu \to \infty$ under RHC sampling design, where Γ_2=$\lim_{\nu\to\infty} n\gamma\overline{X}N^{-1}\sum_{i=1}^{N}(\mathbf{V}_i - X_i\overline{\mathbf{V}}/\overline{X})^T(\mathbf{V}_i - X_i\overline{\mathbf{V}}/\overline{X})X_i^{-1}$ with γ=$\sum_{r=1}^{n}\tilde{N}_r(\tilde{N}_r-1)/N(N-1)$. Note that Γ_2 is a p.d. matrix since Assumption 2.1.5 holds. Let us now consider from Table 2.54 different choices of Θ and **R** corresponding to $\hat{\overline{h}}_{RHC}$, and $\hat{\overline{h}}_{GREG}$ with $d(i,s)$=$(NX_i)^{-1}G_i$. Then, it follows from Lemma 2.7.3 that for each of these choices, $\sqrt{n}\mathbf{R}$=$o_p(1)$ and $\Theta - 1$=$o_p(1)$ as $\nu \to \infty$ under RHC sampling design. Therefore, (2.7.10) holds under RHC sampling design with Γ=Γ_2. This completes the proof of *(ii)* in Lemma 2.7.4 □

Next, recall from the paragraph following Assumption 2.1.2 that γ=$\sum_{r=1}^{n}\tilde{N}_r(\tilde{N}_r-1)/N(N-1)$ with $\tilde{N}_r$ being the size of the r^{th} group formed randomly in RHC sampling design. Then, we state the following lemma.

Lemma 2.7.5. *Suppose that Assumptions 2.1.1 and 2.1.6 hold. Then,* $n\gamma \to c$ *for some* $c \geq 1-\lambda > 0$ *as* $\nu \to \infty$, *where* λ *is as in Assumption 2.1.1.*

Proof. Let us first consider the case of λ=0. Note that

$$\begin{aligned} &n(N/n-1)(N-n)/(N(N-1)) \leq n\gamma \leq \\ &n(N/n+1)(N-n)/(N(N-1)) \end{aligned} \tag{2.7.11}$$

by Assumption 2.1.6 in Section 2.1. Moreover, $n(N/n{+}1)(N{-}n)/(N(N{-}1))$=$(1{+}n/N)(N{-}n)/(N-1) \to 1$ and $n(N/n-1)(N-n)/(N(N-1))$=$(1-n/N)(N-n)/(N-1) \to 1$ as $\nu \to \infty$ because Assumption 2.1.1 holds and λ=0. Thus we have $n\gamma \to 1$ as $\nu \to \infty$ in this case.

Next, consider the case, when $\lambda > 0$ and λ^{-1} is an integer. Here, we consider the following sub-cases. Let us first consider the sub-case, when N/n is an integer for all sufficiently large ν. Then, by Assumption 2.1.6, we have $n\gamma$=$(N-n)/(N-1)$ for all sufficiently large ν. Now, since Assumption 2.1.1 holds, we have

$$(N-n)/(N-1) \to 1-\lambda \text{ as } \nu \to \infty. \tag{2.7.12}$$

Further, consider the sub-case, when N/n is a non-integer and $N/n - \lambda^{-1} \geq 0$ for all sufficiently large ν. Then by Assumption 2.1.6, we have

$$n\gamma = (N/(N-1))(n/N)\lfloor N/n \rfloor \big(2 - \big((n/N)\lfloor N/n \rfloor\big) - (n/N)\big) \tag{2.7.13}$$

for all sufficiently large ν. Now, since Assumption 2.1.1 holds, we have $0 \leq N/n - \lambda^{-1} < 1$ for all sufficiently large ν. Then, $\lfloor N/n \rfloor$=λ^{-1} for all sufficiently large ν, and hence

$$(N/(N-1))(n/N)\lfloor N/n \rfloor \bigg(2 - \big((n/N)\lfloor N/n \rfloor\big) - (n/N)\bigg) \to 1-\lambda \tag{2.7.14}$$

as $\nu \to \infty$.

Next, consider the sub-case, when N/n is a non-integer and $N/n{-}\lambda^{-1} < 0$ for all sufficiently large ν. Then, the result in (2.7.13) holds by Assumption 2.1.6, and $-1 \leq N/n - \lambda^{-1} < 0$ for all sufficiently large ν by Assumption 2.1.1. Therefore, $\lfloor N/n \rfloor$=$\lambda^{-1} - 1$ for all sufficiently large ν, and hence the result in (2.7.14) holds. Thus in the case of $\lambda > 0$ and λ^{-1} an integer, $n\gamma$ converges to $1-\lambda$ as $\nu \to \infty$ through all the sub-sequences, and hence $n\gamma \to 1-\lambda$ as $\nu \to \infty$. Thus we have c=$1-\lambda$ in this case.

Finally, consider the case, when $\lambda > 0$, and λ^{-1} is a non-integer. Then, N/n must be a non-integer for all sufficiently large ν, and hence $n\gamma$=$(N/(N-1))(n/N)\lfloor N/n \rfloor \big(2 - \big((n/N)\lfloor N/n \rfloor\big) - (n/N)\big)$ for all sufficiently large ν by Assumption 2.1.6. Note that in this case, $N/n - \lfloor \lambda^{-1} \rfloor \to \lambda^{-1} - \lfloor \lambda^{-1} \rfloor \in (0,1)$ as $\nu \to \infty$ by Assumption 2.1.1. Therefore, $\lfloor \lambda^{-1} \rfloor < N/n < \lfloor \lambda^{-1} \rfloor + 1$ for all sufficiently large ν, and hence $\lfloor N/n \rfloor$=$\lfloor \lambda^{-1} \rfloor$ for all sufficiently large ν. Thus $n\gamma \to \lambda\lfloor \lambda^{-1} \rfloor(2 - \lambda\lfloor \lambda^{-1} \rfloor - \lambda)$ as $\nu \to \infty$ by Assumption

2.1.1. Now, if t=$\lfloor\lambda^{-1}\rfloor$ and λ^{-1} is a non-integer, then $(t+1)^{-1} < \lambda < t^{-1}$. Therefore, $\lambda\lfloor\lambda^{-1}\rfloor(2-\lambda\lfloor\lambda^{-1}\rfloor-\lambda)-1+\lambda$=$-\big(1-(2t+1)\lambda+t(t+1)\lambda^2\big)$=$-(1-t\lambda)(1-(t+1)\lambda) > 0$. Thus we have c=$\lambda\lfloor\lambda^{-1}\rfloor(2-\lambda\lfloor\lambda^{-1}\rfloor-\lambda) > 1-\lambda$ in this case. This completes the proof of the Lemma. □

Recall the expressions of Σ_1 and Σ_2 from the paragraph preceding Lemma 2.7.2, and ∇g and μ_0 from Assumption 2.1.2. Note that the expression of Σ_1 remains the same for different HEπPS sampling designs. Also, recall from the paragraph preceding Theorem 2.1.3 that ϕ=$\overline{X}-(n/N)\sum_{i=1}^{N}X_i^2/N\overline{X}$. Now, we state the following lemma.

Lemma 2.7.6. *(i) Suppose that Assumptions 2.1.1–2.1.4 hold. Further, suppose that σ_1^2 and σ_2^2 denote $\lim_{\nu\to\infty}\nabla g(\mu_0)\Sigma_1\nabla g(\mu_0)^T$ under SRSWOR and LMS sampling design, respectively, where μ_0=$\lim_{\nu\to\infty}\overline{h}$. Then, we have σ_1^2=σ_2^2=$(1-\lambda)\lim_{\nu\to\infty}\sum_{i=1}^{N}(A_i-\bar{A})^2/N$ for A_i=$\nabla g(\mu_0)\mathbf{V}_i^T$, i=$1,\ldots,N$.*

(ii) Next, suppose that Assumption 2.1.5 holds, and σ_3^2=$\lim_{\nu\to\infty}\nabla g(\mu_0)\Sigma_2\nabla g(\mu_0)^T$ in the case of RHC sampling design. Then, we have σ_3^2=$\lim_{\nu\to\infty}n\gamma((\overline{X}/N)\sum_{i=1}^{N}A_i^2/X_i-\bar{A}^2)$. On the other hand, if Assumptions 2.1.1–2.1.4 hold, and σ_4^2=$\lim_{\nu\to\infty}\nabla g(\mu_0)\Sigma_1\nabla g(\mu_0)^T$ under any HEπPS sampling design, then we have σ_4^2= $\lim_{\nu\to\infty}\big\{(1/N)\sum_{i=1}^{N}A_i^2\big((\overline{X}/X_i)-(n/N)\big)-\phi^{-1}\overline{X}^{-1}\big((n/N)\sum_{i=1}^{N}A_iX_i/N-\overline{AX}\big)^2\big\}$. Further, if Assumption 2.1.1 holds with λ=0, and Assumptions 2.1.2–2.1.4 and 2.1.6 hold, then we have σ_4^2=σ_3^2=$\lim_{\nu\to\infty}((\overline{X}/N)\sum_{i=1}^{N}A_i^2/X_i-\bar{A}^2)$.

Proof. Let us first note that the limits in the expressions of σ_1^2 and σ_2^2 exist in view of Assumption 2.1.4. Also, note that $\nabla g(\mu_0)\Sigma_1\nabla g(\mu_0)^T$=$nN^{-2}\sum_{i=1}^{N}(A_i-T_a\pi_i)^2(\pi_i^{-1}-1)$=$nN^{-2}[\sum_{i=1}^{N}A_i^2(\pi_i^{-1}-1)-(\sum_{i=1}^{N}A_i(1-\pi_i))^2/\sum_{i=1}^{N}\pi_i(1-\pi_i)]$, where T_a=$\sum_{i=1}^{N}A_i(1-\pi_i)/\sum_{i=1}^{N}\pi_i(1-\pi_i)$ and A_i=$\nabla g(\mu_0)\mathbf{V}_i^T$. Now, substituting π_i=n/N in the above expression for SRSWOR, we get σ_1^2= $\lim_{\nu\to\infty}nN^{-2}[\sum_{i=1}^{N}A_i^2(N/n-1)-(\sum_{i=1}^{N}A_i(1-n/N))^2/n(1-n/N)]$=$\lim_{\nu\to\infty}(1-n/N)\sum_{i=1}^{N}(A_i-\bar{A})^2/N$. Since Assumption 2.1.1 holds, we have σ_1^2=$(1-\lambda)\lim_{\nu\to\infty}\sum_{i=1}^{N}(A_i-\bar{A})^2/N$. Let $\{\pi_i\}_{i=1}^{N}$ be the inclusion probabilities of LMS sampling design. Then, $\sigma_2^2-\sigma_1^2$=$\lim_{\nu\to\infty}nN^{-2}[\sum_{i=1}^{N}A_i^2(\pi_i^{-1}-N/n)-((\sum_{i=1}^{N}A_i(1-\pi_i))^2/\sum_{i=1}^{N}\pi_i(1-\pi_i)-(\sum_{i=1}^{N}A_i(1-n/N))^2/n(1-n/N))]$. Now, it can be shown from the proof of Lemma 2.7.1 that $\max_{1\leq i\leq N}|N\pi_i/n-1|\to 0$ as $\nu\to\infty$. Therefore, using Assumption 2.1.2, we can show that $\lim_{\nu\to\infty}nN^{-2}\sum_{i=1}^{N}A_i^2(\pi_i^{-1}-N/n)$=0 and $\lim_{\nu\to\infty}nN^{-2}[(\sum_{i=1}^{N}A_i(1-\pi_i))^2/\sum_{i=1}^{N}\pi_i(1-\pi_i)-(\sum_{i=1}^{N}A_i(1-n/N))^2/n(1-n/N)]$=0, and consequently σ_1^2=σ_2^2. This completes the proof of *(i)* in Lemma 2.7.6

Next, consider the case of RHC sampling design and note that the limit in the expression of σ_3^2 exists in view of Assumption 2.1.5. Also, note that $\nabla g(\mu_0)\Sigma_2\nabla g(\mu_0)^T$ =$n\gamma(\overline{X}/N)\sum_{i=1}^N(A_i - \bar{A}X_i/\overline{X})^2/X_i$=$n\gamma((\overline{X}/N)\sum_{i=1}^N A_i^2/ X_i - \bar{A}^2)$, where $\bar{A}$=$\sum_{i=1}^N A_i/N$ and γ= $\sum_{r=1}^n \tilde{N}_r(\tilde{N}_r - 1)/N(N-1)$. Thus we have σ_3^2=$\lim_{\nu\to\infty} n\gamma((\overline{X}/N)\sum_{i=1}^N A_i^2/X_i - \bar{A}^2)$=$\lim_{\nu\to\infty}((\overline{X}/N)\times \sum_{i=1}^N A_i^2/X_i - \bar{A}^2)$.

Next, note that the limit in the expression of σ_4^2 exists in view of Assumption 2.1.4. Substituting π_i=$nX_i/\sum_{i=1}^N X_i$ in $\nabla g(\mu_0)\Sigma_1\nabla g(\mu_0)^T$ for any HEπPS sampling design, we get σ_4^2=$\lim_{\nu\to\infty} nN^{-2}[\sum_{i=1}^N A_i^2(\sum_{i=1}^N X_i/nX_i - 1) - (\sum_{i=1}^N A_i(1-nX_i/\sum_{i=1}^N X_i))^2/\sum_{i=1}^N(nX_i/\sum_{i=1}^N X_i)(1-nX_i/\sum_{i=1}^N X_i)]$=$\lim_{\nu\to\infty}\{(1/N)\sum_{i=1}^N A_i^2\big((\overline{X}/X_i)-(n/N)\big) - \phi^{-1}\overline{X}^{-1}\times \big((n/N)\sum_{i=1}^N A_iX_i/N - \overline{A}\,\overline{X}\big)^2\}$. Further, we can show that σ_4^2=$\lim_{\nu\to\infty}((\overline{X}/N)\sum_{i=1}^N A_i^2/X_i - \bar{A}^2)$, when Assumptions 2.1.2 and 2.1.3 hold, and Assumption 2.1.1 holds with λ=0. It also follows from Lemma 2.7.5 that $n\gamma\to 1$ as $\nu\to\infty$, when Assumption 2.1.1 holds with λ=0. Thus we have σ_3^2=σ_4^2=$\lim_{\nu\to\infty}((\overline{X}/N)\sum_{i=1}^N A_i^2/X_i - \bar{A}^2)$. This completes the proof of *(ii)* in Lemma 2.7.6. □

Lemma 2.7.7. *Suppose that Assumptions 2.1.1–2.1.3 hold. Then under SRSWOR, LMS and any HEπPS sampling designs, we have*

$$\text{(i)}\quad u^* = \max_{i\in s}|L_i| = o_p(\sqrt{n}),\ \textit{and}\qquad \text{(ii)}\quad \sum_{i\in s}\pi_i^{-1}L_i\Big/\sum_{i\in s}\pi_i^{-1}L_i^2 = O_p(1/\sqrt{n})$$

as $\nu\to\infty$, where L_i=$X_i-\overline{X}$ for i=$1,\ldots,N$

Proof. Let $P(s)$ be any sampling design and E be the expectation with respect to $P(s)$. Then, $E(u^*/\sqrt{n}) \le (\max_{1\le i\le N} X_i + \overline{X})/\sqrt{n} \le \overline{X}(\max_{1\le i\le N} X_i/\min_{1\le i\le N} X_i + 1)/\sqrt{n}$=$o(1)$ as $\nu\to\infty$ since Assumptions 2.1.2 and 2.1.3 hold. Therefore, *(i)* holds under $P(s)$ by Markov inequality. Thus *(i)* holds under SRSWOR, LMS and any HEπPS sampling designs.

Using similar arguments as in the 1^{st} paragraph of the proof of Lemma 2.7.3, it can be shown that $\sqrt{n}(\sum_{i\in s} L_i/N\pi_i - \bar{L})$=$\sqrt{n}\sum_{i\in s} L_i/N\pi_i$=$O_p(1)$ and $\sum_{i\in s} L_i^2/N\pi_i - \sum_{i=1}^N L_i^2/N$=$o_p(1)$ as $\nu\to\infty$ under a high entropy sampling design $P(s)$ satisfying (2.7.1) in Lemma 2.7.1. Therefore, $1/(\sum_{i\in s} L_i^2/N\pi_i)$ =$O_p(1)$ as $\nu\to\infty$ under $P(s)$ since $\sum_{i=1}^N L_i^2/N$ is bounded away from 0 as $\nu\to\infty$ by Assumption 2.1.2. Thus under $P(s)$, $\sum_{i\in s}\pi_i^{-1}L_i/\sum_{i\in s}\pi_i^{-1}L_i^2$=$O_p(1/\sqrt{n})$ as $\nu\to\infty$.

It follows from Lemma 2.7.1 that SRSWOR and LMS sampling design are high entropy sampling designs and satisfy (2.7.1). It also follows from Lemma 2.7.1 that any HEπPS sampling design satisfies (2.7.1). Therefore, the result in *(ii)* holds under the above-mentioned sampling designs. □

In the following lemma, we demonstrate some situations, when Assumptions 2.1.2–2.1.5 hold. Let us recall $\{\mathbf{V}_i\}_{i=1}^N$ and $\overline{\mathbf{V}}$ from the paragraph preceding Assumption 2.1.4. Let us also recall the expressions of Σ_1 and Σ_2 from the paragraph preceding Lemma 2.7.2 and b from Assumption 2.2.1. Now, we state the following lemma.

Lemma 2.7.8. *(i) Suppose that Assumptions 2.1.1, 2.2.1 and 2.1.6 hold, and* $\{(h(Y_i), X_i) : 1 \leq i \leq N\}$ *are generated from a superpopulation distribution* $\mathbb{P}$ *with* $E_{\mathbb{P}}||h(Y_i)||^4 < \infty$*. Then, Assumptions 2.1.2, 2.1.3 and 2.1.5 hold a.s.* $[\mathbb{P}]$.
(ii) Further, if Assumptions 2.1.1 and 2.2.1 hold, and $E_{\mathbb{P}}||h(Y_i)||^2 < \infty$*, then Assumption 2.1.4 holds a.s.* $[\mathbb{P}]$ *under SRSWOR and LMS sampling design. Moreover, if Assumptions 2.1.1 holds with* $0 \leq \lambda < E_{\mathbb{P}}(X_i)/b$*, Assumption 2.2.1 holds, and* $E_{\mathbb{P}}||h(Y_i)||^2 < \infty$*, then Assumption 2.1.4 holds a.s.* $[\mathbb{P}]$ *under any* π*PS sampling design.*

***Proof*.** As before, for simplicity, let us write $h(Y_i)$ as h_i. Under the conditions Assumption 2.2.1 and $E_{\mathbb{P}}||h(Y_i)||^4 < \infty$, Assumption 2.1.2 holds *a.s.* $[\mathbb{P}]$ by SLLN. Also, under Assumption 2.2.1, Assumption 2.1.3 holds *a.s.* $[\mathbb{P}]$. Next, by SLLN, $\lim_{\nu\to\infty}\Sigma_2 = cE_{\mathbb{P}}(X_i)E_{\mathbb{P}}[(h_i - (E_{\mathbb{P}}(X_i))^{-1}X_i E_{\mathbb{P}}(h_i))^T(h_i - (E_{\mathbb{P}}(X_i))^{-1}X_i E_{\mathbb{P}}(h_i))\, X_i^{-1}]$ *a.s.* $[\mathbb{P}]$ for $\mathbf{V}_i = h_i$, $h_i - \overline{h}X_i/\overline{X}$ and $h_i + \overline{h}X_i/\overline{X}$ because $n\gamma \to c$ as $\nu \to \infty$ by Lemma 2.7.5. Similarly, $\lim_{\nu\to\infty}\Sigma_2 = cE_{\mathbb{P}}(X_i)E_{\mathbb{P}}[(h_i - E_{\mathbb{P}}(h_i))^T(h_i - E_{\mathbb{P}}(h_i))/X_i]$ *a.s.* $[\mathbb{P}]$ for $\mathbf{V}_i = h_i - \overline{h}$, and $\lim_{\nu\to\infty}\Sigma_2 = E_{\mathbb{P}}(X_i)E_{\mathbb{P}}[\,(h_i - E_{\mathbb{P}}(h_i) - C_{xh}(X_i - E_{\mathbb{P}}(X_i)))^T(h_i - E_{\mathbb{P}}(h_i) - C_{xh}(X_i - E_{\mathbb{P}}(X_i)))/X_i]$ *a.s.* $[\mathbb{P}]$ for $\mathbf{V}_i = h_i - \overline{h} - S_{xh}(X_i - \overline{X})/S_x^2$. Here, $C_{xh} = (E_{\mathbb{P}}(h_iX_i) - E_{\mathbb{P}}(h_i)E_{\mathbb{P}}(X_i))/\,(E_{\mathbb{P}}(X_i)^2 - (E_{\mathbb{P}}(X_i))^2)$. Note that the above limits are p.d. matrices because Assumption 2.2.1 holds. Therefore, Assumption 2.1.5 holds *a.s.* $[\mathbb{P}]$. This completes the proof of *(i)* in Lemma 2.7.8

Next, note that $\Sigma_1 = (1 - n/N)(\sum_{i=1}^N \mathbf{V}_i^T\mathbf{V}_i/N - \overline{\mathbf{V}}^T\overline{\mathbf{V}})$ under SRSWOR. Then, Assumption 2.1.4 holds *a.s.* $[\mathbb{P}]$ by directly applying SLLN. Under LMS sampling design, Assumption 2.1.4 can be shown to hold *a.s.* $[\mathbb{P}]$ in the same way as the proof of the result $\sigma_1^2 = \sigma_2^2$ in the proof of Lemma 2.7.6. Next, we have $\lim_{\nu\to\infty}\Sigma_1 = E_{\mathbb{P}}\big[\{h_i + \chi^{-1}(E_{\mathbb{P}}(X_i))^{-1}X_i\big(\lambda E_{\mathbb{P}}(h_iX_i) - E_{\mathbb{P}}(h_i)E_{\mathbb{P}}(X_i)\big)\}^T\{h_i + \chi^{-1}(E_{\mathbb{P}}(X_i))^{-1}X_i\big(\lambda E_{\mathbb{P}}(h_iX_i) - E_{\mathbb{P}}(h_i)E_{\mathbb{P}}(X_i)\big)\}\{E_{\mathbb{P}}(X_i)/X_i -$

$\lambda\}\big]$ *a.s.* $[\mathbb{P}]$ for $\mathbf{V}_i$=h_i, $h_i - \overline{h}X_i/\overline{X}$ and $h_i + \overline{h}X_i/\overline{X}$ under any πPS sampling design (i.e., a sampling design with π_i=$nX_i/\sum_{i=1}^{N} X_i$) by SLLN because Assumptions 2.1.1 and 2.2.1 hold, and $E_{\mathbb{P}}||h_i||^2 < \infty$. Here, χ=$E_{\mathbb{P}}(X_i) - \lambda(E_{\mathbb{P}}(X_i)^2/E_{\mathbb{P}}(X_i))$. Moreover, under any πPS sampling design, we have $\lim_{\nu\to\infty}\Sigma_1$=$E_{\mathbb{P}}\big[\big\{h_i - E_{\mathbb{P}}(h_i) + \lambda\chi^{-1}(E_{\mathbb{P}}(X_i))^{-1}X_iC_{xh}\big\}^T\big\{h_i - E_{\mathbb{P}}(h_i) + \lambda\chi^{-1}(E_{\mathbb{P}}(X_i))^{-1}X_iC_{xh}\big\}\times\ \big\{E_{\mathbb{P}}(X_i)/X_i - \lambda\big\}\big]$ *a.s.* $[\mathbb{P}]$ for $\mathbf{V}_i$=$h_i - \overline{h}$ and $\lim_{\nu\to\infty}\Sigma_1$= $E_{\mathbb{P}}\big[\big\{h_i - E_{\mathbb{P}}(h_i) - C_{xh}(X_i - E_{\mathbb{P}}(X_i))\big\}^T\big\{h_i - E_{\mathbb{P}}(h_i) - C_{xh}(X_i - E_{\mathbb{P}}(X_i))\big\}\big\{E_{\mathbb{P}}(X_i)/X_i - \lambda\big\}\big]$ *a.s.* $[\mathbb{P}]$ for $\mathbf{V}_i$=$h_i - \overline{h} - S_{xh}(X_i - \overline{X})/S_x^2$. Note that the above limits are p.d. matrices because Assumption 2.2.1 holds and Assumption 2.1.5 holds with $0 \leq \lambda < E_{\mathbb{P}}(X_i)/b$. Therefore, Assumption 2.1.4 holds *a.s.* $[\mathbb{P}]$ under any πPS sampling design. This completes the proof of *(ii)* in Lemma 2.7.8. □

Chapter 3

Estimators of the mean of infinite dimensional data in finite populations

In the recent past, [12], [13], [16], etc. considered the HT estimator (see [44]) of the finite population mean, when population observations are from some functional space. [14] and [15] also constructed a model assisted estimator for finite population mean function based on some homoscedastic linear regression models. This model assisted estimator can be related to the GREG estimator considered earlier in [22] for finite dimensional data. All these authors investigated different asymptotic properties of the HT and the model assisted estimators in $\mathcal{C}[0, T]$, the space of continuous functions defined on $[0, T]$, under sampling designs, which satisfy some regularity conditions. These sampling designs include SRSWOR, stratified sampling design with SRSWOR and rejective sampling designs. However, none of these authors compared the performance of the aforementioned estimators under different sampling designs.

In this chapter, we consider the extensions of the HT and the RHC estimators (see Table 2.1 in Chapter 2) for the population mean of a study variable that lies in an infinite dimensional separable Hilbert space $\mathcal{H}$ because these estimators are widely used design unbiased estimators of the population mean for finite dimensional data. We also consider the extension of the GREG estimator (see Table 2.1 in Chapter 2) for the population mean of the same study variable, which is not a design unbiased estimator but known to be asymptotically often more efficient than other estimators for finite dimensional data (see Sections 2.1 and 2.2 in Chapter 2). We compare the HT, the RHC and the GREG estimators using their asymptotic distributions under SRSWOR, LMS, HEπPS and RHC sampling designs (see the introduction), and some superpopulations satisfying linear regression models. The main results obtained from this comparison are the following.

- The GREG estimator is asymptotically at least as good as the HT estimator under each of SRSWOR, LMS and any HEπPS sampling designs. Also, the GREG estimator turns out to be asymptotically at least as good as the RHC estimator under RHC sampling design.
- If the degree of heteroscedasticity present in linear regression models is not very large, then the use of the well-known sampling designs like RHC and any HEπPS sampling designs instead of SRSWOR may have an adverse effect on the performance of the GREG estimator. In other words, the use of the auxiliary information in the design stage of sampling may have an adverse effect on the performance of the GREG estimator. On the other hand, if the degree of heteroscedasticity present in linear regression models is sufficiently large, then the sampling designs like RHC and any HEπPS sampling designs lead to an improvement in the performance of the GREG estimator.

In section 3.1, we discuss infinite dimensional extensions of the HT, the RHC and the GREG estimators of the population mean. In section 3.2, we compare these estimators using their asymptotic distributions under the sampling designs mentioned above and some superpopulations satisfying linear regression models. In this section, we also discuss the estimation of asymptotic covariance operators of several estimators and show that these estimators of asymptotic covariance operators are consistent. Some numerical results based on both synthetic and real data are presented in Section 3.3. Several methods of determining the degree of heteroscedasticity present in linear regression models are provided in Section 3.4. Proofs of various results are given in Sections 3.5 and 3.6.

3.1. Estimators based on infinite dimensional data

Suppose that $\mathcal{H}$ is an infinite dimensional separable Hilbert space with associated inner product $\langle \cdot, \cdot \rangle$, and y is a $\mathcal{H}$-valued study variable. Some examples of such a study variable are electricity consumption curve of household in the summer/winter (e.g., see [12], [13], [16], [14], etc.), rainfall curve in state/district over a particular time period (e.g., see the website of India Meteorological Department (*https://mausam.imd.gov.in/imd_latest/contents/rainfall_statistics_3.php*)), growth curve of height of male/female over a certain period of time (see [83]), micro-array expression levels of genes in cell/tissue (e.g., see the Colon dataset in the statistical software *R*), etc. Recall from the introduction that $Y_1, \ldots, Y_N$ are the population values of y. The HT estimator of the finite population mean of y, $\overline{Y}$=$\sum_{i=1}^{N} Y_i/N$, is defined as

$$\hat{\overline{Y}}_{HT} = \sum_{i \in s} (N\pi_i)^{-1} Y_i, \tag{3.1.1}$$

where π_i=$\sum_{s \ni i} P(s)$ is the inclusion probability of the i^{th} population unit for i=$1, \ldots, N$.

Before we write the expression of the RHC estimator, recall from the introduction that in the RHC sampling design, the population $\mathcal{P}$ is first divided randomly into n disjoint groups of sizes $\tilde{N}_1, \ldots, \tilde{N}_n$ such that $\sum_{r=1}^{n} \tilde{N}_r$=$N$, and then one unit is selected from each group independently. Also, recall from the beginning of Section 2.1 in Chapter 2 that G_i denotes the total of the x values of that randomly formed group from which the i^{th} unit is selected in the sample s. Then, the RHC estimator of $\overline{Y}$ can be expressed as

$$\hat{\overline{Y}}_{RHC} = \sum_{i \in s} (NX_i)^{-1} G_i Y_i, \tag{3.1.2}$$

where $X_1, \ldots, X_N$ are known population values on the size variable x in $(0, \infty)$.

[66] considered the RHC estimator for a real-valued study variable. The RHC estimator is more easily computable than other unbiased estimators under other unequal probability sampling designs without replacement (e.g., the HT or the Des Raj estimator (see [58]) under probability proportional to size sampling without replacement). Moreover, the RHC estimator has smaller variance than the usual unbiased estimator under the probability proportional to size sampling with replacement. Also, its variance can be estimated by a non negative unbiased estimator. These results continue to hold, when we consider the RHC estimator for a $\mathcal{H}$-valued study variable.

[68], [72], [28], [22], etc.considered the GREG estimator for finite dimensional data. Suppose that z=$(z_1, \ldots, z_d)$ is a $\mathbb{R}^d$-valued ($d \geq 1$) covariate with population values $Z_1, \ldots, Z_N$ and known population total $\sum_{i=1}^{N} Z_i$. It will be appropriate to note that the size variable x may be one of the real-valued components of z in some cases. As mentioned in Chapter 2, all vectors in Euclidean spaces will be taken as row vectors and superscript T will be used to denote their transpose. Further, suppose that $\mathcal{G}$ is any separable Hilbert space with inner product $\langle \cdot, \cdot \rangle$, and $\mathcal{B}(\mathcal{G}, \mathcal{H})$ is the class of all bounded linear operators from $\mathcal{G}$ to $\mathcal{H}$. It is to be noted that $\mathcal{B}(\mathcal{G}, \mathcal{H})$ is an infinite dimensional Hilbert space associated with the Hilbert-Schmidt (HS) inner product

(see [45]). For any $a \in \mathcal{G}$ and $b \in \mathcal{H}$, let us consider the tensor product $a \otimes b \in \mathcal{B}(\mathcal{G}, \mathcal{H})$, which is defined as $(a \otimes b)e=\langle a, e\rangle b$, $e \in \mathcal{G}$. Suppose that $\hat{\overline{Z}}=\sum_{i\in s}\pi_i^{-1}Z_i/\sum_{i\in s}\pi_i^{-1}$. Let us also suppose that the inverse of $\hat{S}_{zz}=\sum_{i\in s}\pi_i^{-1}(Z_i-\hat{\overline{Z}})^T(Z_i-\hat{\overline{Z}})/\sum_{i\in s}\pi_i^{-1}$ exists. Then, an infinite dimensional version of the GREG estimator for the population mean is defined as

$$\hat{\overline{Y}}_{GREG} = \hat{\overline{Y}} + \hat{S}_{zy}((\overline{Z} - \hat{\overline{Z}})\hat{S}_{zz}^{-1}), \tag{3.1.3}$$

where $\overline{Z}=\sum_{i=1}^{N} Z_i/N$, $\hat{\overline{Y}}=\sum_{i\in s}\pi_i^{-1}Y_i/\sum_{i\in s}\pi_i^{-1}$ and $\hat{S}_{zy}=\sum_{i\in s}\pi_i^{-1}(Z_i-\hat{\overline{Z}})\otimes(Y_i-\hat{\overline{Y}})/\sum_{i\in s}\pi_i^{-1}$. Under RHC sampling design, we consider the GREG estimator $\hat{\overline{Y}}_{GREG}$ after replacing π_i^{-1} by $G_iX_i^{-1}$ (cf. [22]).

3.2. Comparison of estimators under superpopulation models

In this section, we compare among the HT and the GREG estimators under SRSWOR, LMS and HEπPS sampling designs, and the RHC and the GREG estimators under RHC sampling design. For this, as mentioned in the introduction, we assume that the observations $\{(Y_i, Z_i, X_i) : 1 \leq i \leq N\}$ are i.i.d. $\mathcal{H} \times \mathbb{R}^{d+1}$-valued random variables on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Also, as in Section 2.2, we consider the function $P(s, \omega)$ that is defined on $\mathcal{S} \times \Omega$. Recall from Section 2.2 that for each $s \in \mathcal{S}$, $P(s, \omega)$ is a random variable on Ω, and for each $\omega \in \Omega$, $P(s, \omega)$ is a probability distribution on $\mathcal{S}$. It is to be noted that $P(s, \omega)$ is a sampling design for each $\omega \in \Omega$. Next, recall from Section 2.1 in Chapter 2 that our asymptotic framework is as follows. Let $\{\mathcal{P}_\nu\}$ be a sequence of populations with N_ν, $n_\nu \to \infty$ as $\nu \to \infty$, where N_ν and n_ν are, respectively, the population size and the sample size corresponding to the ν^{th} population. We shall frequently drop the limiting index ν for the sake of notational simplicity.

We now slightly modify the notation to describe high entropy sampling designs given in the introduction. Suppose that a sampling design $P(s, \omega)$ is such that the Kullback–Leibler divergence $D(P||R)=\sum_{s\in\mathcal{S}} P(s, \omega)\log\big(P(s, \omega)/R(s, \omega)\big)$ converges to 0 as $\nu \to \infty$ *a.s.* [$\mathbf{P}$] for some rejective sampling design $R(s, \omega)$ (for the description of rejective sampling design, see the introduction). Such a sampling design is known as the high entropy sampling design (cf. [4], [16], [7], etc.). We call a sampling design $P(s, \omega)$ a HEπPS sampling design if it is a high entropy sampling design as well as a πPS sampling design (see the introduction).

Before we state our main results, let us consider some assumptions on distributions of $\{Y_i, Z_i, X_i\}_{i=1}^N$. Recall from Section 2.2 in Chapter 2 that $E_{\mathbf{P}}$ denotes that expectation with respect to the probability measure **P**. The expectations of $\mathcal{H}$-valued random variables are defined using Bochner integrals (see [45]). Also, recall from Section 2.1 in Chapter 2 that in any finite dimensional Euclidean space, we consider the Euclidean norm and denote it by $||\cdot||$. On the other hand, in $\mathcal{H}$, we consider the norm induced by the inner product associated with $\mathcal{H}$ and denote it by $||\cdot||_{\mathcal{H}}$.

Assumption 3.2.1. *$n/N \to \lambda$ as $\nu \to \infty$, where $0 \leq \lambda < 1$.*

Assumption 3.2.2. *$0 < X_i \leq b$ a.s. $[\boldsymbol{P}]$ for some $b > 0$, $E_{\boldsymbol{P}}(X_i)^{-2} < \infty$, and $\max_{1\leq i\leq N} X_i/ \min_{1\leq i\leq N} X_i = O(1)$ as $\nu \to \infty$ a.s. $[\boldsymbol{P}]$.*

Assumption 3.2.3. *$E_{\boldsymbol{P}}||Y_i||_{\mathcal{H}}^4 < \infty$, $E_{\boldsymbol{P}}||Z_i||^4 < \infty$, and $E_{\boldsymbol{P}}(Z_i - E_{\boldsymbol{P}}(Z_i))^T(Z_i - E_{\boldsymbol{P}}(Z_i))$ is positive definite (p.d.).*

Assumptions 3.2.1 and 3.2.2 are discussed in Chapter 2 (see the discussion related to Assumptions 2.1.1, 2.1.3 and 2.2.1 in Chapter 2). Assumption 3.2.3 implies that the fourth order raw moments of Y_i and Z_i exist. In this chapter, Assumptions 3.2.1–3.2.3 are used to prove some technical results (see Lemmas 3.6.1–3.6.4 in Section 3.6) under LMS, HEπPS and RHC sampling designs, which will be required to show weak convergence of $\sqrt{n}(\hat{\overline{Y}}_{HT} - \overline{Y})$, $\sqrt{n}(\hat{\overline{Y}}_{RHC} - \overline{Y})$ and $\sqrt{n}(\hat{\overline{Y}}_{GREG} - \overline{Y})$ via uniform approximation (see [54]). Now, we state the following proposition.

Proposition 3.2.1. *Suppose that Assumptions 3.2.1–3.2.3 hold. Then, a.s. $[\boldsymbol{P}]$, under SRSWOR and LMS sampling design, $\sqrt{n}(\hat{\overline{Y}}_{HT} - \overline{Y}) \xrightarrow{\mathcal{L}} \mathcal{N}$ as $\nu \to \infty$, where $\mathcal{N}$ is a Gaussian distribution in $\mathcal{H}$ with mean 0 and some covariance operator. Moreover, if Assumption 3.2.1 holds with $0 \leq \lambda < E_{\boldsymbol{P}}(X_i)/b$, and Assumptions 3.2.2 and 3.2.3 hold, then the same result holds under any HEπPS sampling design.*

Next, as in Chapter 2, here also we consider the following assumption.

Assumption 3.2.4. *For the RHC sampling design, $\{\tilde{N}_r\}_{r=1}^n$ are such that*

$$\tilde{N}_r = \begin{cases} N/n, \text{ for } r = 1, \cdots, n, \text{ when } N/n \text{ is an integer}, \\ \lfloor N/n \rfloor, \text{ for } r = 1, \cdots, k, \text{ and} \\ \lfloor N/n \rfloor + 1, \text{ for } r = k+1, \cdots, n, \text{ when } N/n \text{ is not an integer}, \end{cases} \tag{3.2.1}$$

where k is such that $\sum_{r=1}^{n} \tilde{N}_r = N$. Here, $\lfloor N/n \rfloor$ is the integer part of N/n.

Now, we state the following propositions.

Proposition 3.2.2. *Suppose that Assumptions 3.2.1–3.2.4 hold. Then, a.s. $[\boldsymbol{P}]$, under RHC sampling design, $\sqrt{n}(\hat{\overline{Y}}_{RHC} - \overline{Y}) \xrightarrow{\mathcal{L}} \mathcal{N}$ as $\nu \to \infty$, where $\mathcal{N}$ is a Gaussian distribution in $\mathcal{H}$ with mean 0 and some covariance operator.*

Proposition 3.2.3. *Suppose that Assumptions 3.2.1–3.2.3 hold. Then, a.s. $[\boldsymbol{P}]$, under SRSWOR and LMS sampling design, $\sqrt{n}(\hat{\overline{Y}}_{GREG} - \overline{Y}) \xrightarrow{\mathcal{L}} \mathcal{N}$ as $\nu \to \infty$, where $\mathcal{N}$ is a Gaussian distribution in $\mathcal{H}$ with mean 0 and some covariance operator. Further, if Assumption 3.2.1 holds with $0 \leq \lambda < E_{\boldsymbol{P}}(X_i)/b$, and Assumptions 3.2.2 and 3.2.3 hold, then the same result holds under any HEπPS sampling design. Moreover, if Assumptions 3.2.1–3.2.4 hold, then the above result holds under RHC sampling design.*

The weak convergence of the HT and the GREG estimators is shown under SRSWOR, LMS and HEπPS sampling designs (see Propositions 3.2.1 and 3.2.3) using the weak convergence of the HT and the GREG estimators under high entropy sampling designs and the fact that the aforementioned sampling designs can be approximated by rejective sampling designs in Kullback-Liebler divergence. The technique used to prove Propositions 3.2.1–3.2.3 is based on the idea of convergence in distribution via uniform approximation considered in [54]. This idea was used in [54] to extend central limit theorem for independent random variables from finite dimensional Euclidean space to an infinite dimensional separable Hilbert space (see Proposition 2.1 in [54]). Any infinite dimensional separable Hilbert space (e.g., the space of square integrable functions equipped with L^2-inner product) is isometrically isomorphic to the space of square summable sequences l^2 because a separable Hilbert space always has a complete orthonormal basis. Further, the l^2 space can be conveniently viewed as an infinite dimensional extension of a finite dimensional Euclidean space. Thus it is relatively easy to extend the results from multivariate data setup to the functional data setup using the sequence structure of the l^2 space.

[12] and [15] showed the weak convergence of the HT and the model assisted estimators, respectively, in $\mathcal{C}[0, T]$ under some conditions on sampling designs (see pp. 110-111 in [12] and pp. 569-573 in [15]). These conditions hold under usual sampling designs like SRSWOR, stratified sampling design with SRSWOR, rejective sampling design, etc. We are able to dispense with these conditions, and show the weak convergence of the HT and the GREG estimators in a separable Hilbert space under SRSWOR, LMS and any HEπPS sampling designs. Many of

these sampling designs are not covered in the earlier literature. We are also able to show the weak convergence of the RHC and the GREG estimators in a separable Hilbert space under RHC sampling design. These results are not available in the earlier literature.

We develop our results in a separable Hilbert space framework rather than in a space of continuous functions equipped with supremum norm because we are able to prove Propositions 3.2.1–3.2.3 in the case of a separable Hilbert space framework. The space of continuous functions is a subset of the space of square integrable functions, which is a separable Hilbert space equipped with L^2 inner product. Random functions from the space of continuous functions can be expressed as linear combinations of orthonormal basis functions in the space of square integrable functions through the Karhunen-Loève expansion.

Next, we carry out the comparison of the estimators mentioned earlier based on the above results. We say that an estimator $\hat{\overline{Y}}_1$ with asymptotic covariance operator Γ is asymptotically at least as efficient as another estimator $\hat{\overline{Y}}_2$ with asymptotic covariance operator Δ if $\Delta - \Gamma$ is non negative definite (n.n.d.), i.e., if $\langle(\Delta - \Gamma)a, a\rangle \geq 0$ for any $a \in \mathcal{H}$. We also say that $\hat{\overline{Y}}_1$ is asymptotically more efficient than $\hat{\overline{Y}}_2$ if $\Delta - \Gamma$ is p.d, i.e., if $\langle(\Delta - \Gamma)a, a\rangle > 0$ for any $a \in \mathcal{H}$ and $a \neq 0$. We now state the following theorems.

Theorem 3.2.1. *Suppose that Assumptions 3.2.1–3.2.3 hold. Then, a.s. [**P**], the GREG estimator is asymptotically at least as efficient as the HT estimator under SRSWOR as well as LMS sampling design. Moreover, a.s. [**P**], both the GREG estimator has the same asymptotic distribution under SRSWOR and LMS sampling design.*

Before we state the next theorem, let us consider superpopulations satisfying the linear regression model

$$Y_i = \beta_0 + \sum_{j=1}^{d} Z_{ji}\beta_j + \epsilon_i X_i^{\eta}, \tag{3.2.2}$$

where i=$1, \ldots, N$, $\{\epsilon_i\}_{i=1}^N$ are i.i.d. $\mathcal{H}$-valued random variables independent of $\{Z_i, X_i\}_{i=1}^N$ with mean 0. Here, Z_i=$(Z_{1i}, \ldots, Z_{di})$, $\beta_j \in \mathcal{H}$ for j=$0, \ldots, d$, and $\eta \geq 0$ is the degree of heteroscedasticity present in the linear model given above. For any given $\eta > 0$, the conditional total variance of Y_i given (Z_i, X_i), the trace of the conditional covariance operator of Y_i given (Z_i, X_i), increases as the value of X_i increases (cf. [72]). In essence, the parameter η determines the rate at which this conditional total variance increases with X_i. Similar types of linear model as in (3.2.2) were used for constructing several estimators by earlier authors, when the observations on y are from some finite dimensional Euclidean space (see [17], [71], [72] and references therein). A homoscedastic (i.e., when η=0) version of the above linear regression model was considered earlier in [14] and [15] for constructing the model assisted estimator of $\overline{Y}$, when the observations on y are from some functional space. Now, we state the following theorem.

Theorem 3.2.2. *Suppose that* (3.2.2) *and Assumptions 3.2.1–3.2.4 hold. Then, a.s.* [$\boldsymbol{P}$]*, the GREG estimator is asymptotically at least as efficient as the RHC estimator under RHC sampling design. Further, if* (3.2.2) *holds, Assumption 3.2.1 holds with* $0 \leq \lambda < E_{\boldsymbol{P}}(X_i)/b$*, and Assumptions 3.2.2 and 3.2.3 hold, then a.s.* [$\boldsymbol{P}$]*, the GREG estimator is asymptotically at least as efficient as the HT estimator under any HEπPS sampling design.*

It follows from the preceding results that the GREG estimator is asymptotically at least as efficient as the HT and RHC estimators under each of the sampling designs considered in this chapter. Also, both the HT and the GREG estimators have the same asymptotic distribution under SRSWOR and LMS sampling design. Now, we compare the performance of the GREG estimator under SRSWOR, RHC sampling design and HEπPS sampling designs based on the degree of heteroscedasticity η.

Theorem 3.2.3. *Suppose that* (3.2.2) *holds, and* ϵ_i *has a p.d. covariance operator. Further, suppose that Assumption 3.2.1 holds with* $0 \leq \lambda < E_{\boldsymbol{P}}(X_i)/b$*, and Assumptions 3.2.2–3.2.4 hold. Then, the sampling designs among SRSWOR, HEπPS and RHC sampling designs under which the GREG estimator becomes the most efficient estimator a.s.* [$\boldsymbol{P}$] *are as mentioned in Table 3.1 below. Further, if Assumption 3.2.1 holds with* λ=0*, and Assumptions 3.2.2–3.2.4 hold, then the GREG estimator has the same asymptotic distribution under RHC and any HEπPS sampling designs.*

Proofs of Theorems 3.2.1–3.2.3 involve some results related to operator theory, which are available in [45]. It follows from (3.5.18) in the proof of Theorem 3.2.3 that $cov_{\mathbf{P}}(X_i^{2\eta-1}, X_i)$, the covariance between $X_i^{2\eta-1}$ and X_i, determines the sampling design among SRSWOR, HEπPS and RHC sampling designs under which the GREG estimator becomes the most efficient estimator. The GREG estimator performs more efficiently under SRSWOR than under RHC and any HEπPS sampling designs, whenever $cov_{\mathbf{P}}(X_i^{2\eta-1}, X_i) < 0$. On the other hand, the GREG estimator under RHC as well as any HEπPS sampling design becomes more efficient than the GREG estimator under SRSWOR in the case of λ=0, whenever $cov_{\mathbf{P}}(X_i^{2\eta-1}, X_i) > 0$, and the GREG estimator under any HEπPS sampling design becomes more efficient than

TABLE 3.1: Sampling designs for which the GREG estimator becomes the most efficient estimator.

	λ=0	$\lambda > 0$ & λ^{-1} is an integer	$\lambda > 0$ & λ^{-1} is a non-integer
$\eta < 0.5$	SRSWOR	SRSWOR	SRSWOR
$\eta = 0.5$	[1]SRSWOR, HEπPS & RHC	[1]SRSWOR, HEπPS & RHC	[2] SRSWOR & HEπPS
$\eta > 0.5$	[3] HEπPS & RHC	HEπPS	HEπPS

[1] GREG estimator has the same asymptotic distribution under SRSWOR, RHC sampling design and HEπPS sampling designs for η=0.5, $\lambda > 0$ and λ^{-1} an integer.

[2] GREG estimator has the same asymptotic distribution under SRSWOR and HEπPS sampling designs, when η=0.5, $\lambda > 0$ and λ^{-1} is a non-integer.

[3] GREG estimator has the same asymptotic distribution under HEπPS and RHC sampling designs for $\eta > 0.5$ and λ=0.

the GREG estimator under both SRSWOR and RHC sampling design in the case of $\lambda > 0$, whenever $cov_{\mathbf{P}}(X_i^{2\eta-1}, X_i) > 0$. Now, $x^{2\eta-1}$ is a decreasing function of x for $\eta < 0.5$ and an increasing function of x for $\eta > 0.5$. Therefore, $cov_{\mathbf{P}}(X_i^{2\eta-1}, X_i) < 0$ for $\eta < 0.5$ and $cov_{\mathbf{P}}(X_i^{2\eta-1}, X_i) > 0$ for $\eta > 0.5$. Thus the use of the auxiliary information in HEπPS and RHC sampling designs has an adverse effect on the performance of the GREG estimator, when $\eta < 0.5$. On the other hand, for the case of $\eta > 0.5$, the use of HEπPS and RHC sampling designs improves the performance of the GREG estimator.

Note that if we consider a generalized version of the linear regression model in (3.2.2) as Y_i=$\beta_0 + \sum_{j=1}^{d} Z_{ji}\beta_j + \epsilon_i g(X_i)$ for i=$1,\ldots,N$ and some non-negative real-valued function g, then it can be shown in the same way as in the proof of Theorem 3.2.2 that the conclusion of Theorem 3.2.2 holds under the above linear model. It can also be shown in the same way as in the proof of Theorem 3.2.3 that the results in 2^{nd}, 3^{rd} and 4^{th} rows in Table 3.1 related to Theorem 3.2.3 hold, whenever $cov_{\mathbf{P}}(g^2(X_i)X_i^{-1}, X_i) < 0$, $cov_{\mathbf{P}}(g^2(X_i)X_i^{-1}, X_i)$=0 and $cov_{\mathbf{P}}(g^2(X_i)X_i^{-1}, X_i) > 0$, respectively. In particular, the results in 2^{nd}, 3^{rd} and 4^{th} rows in Table 3.1 hold if $g^2(x)x^{-1}$ is decreasing, constant and increasing function of x, respectively.

Let us denote the asymptotic covariance operator of $\sqrt{n}(\hat{\overline{Y}} - \overline{Y})$ by Γ, where $\hat{\overline{Y}}$ denotes one of $\hat{\overline{Y}}_{HT}$, $\hat{\overline{Y}}_{RHC}$ and $\hat{\overline{Y}}_{GREG}$. Next, suppose that $\hat{\overline{Y}}$ is either $\hat{\overline{Y}}_{HT}$ or $\hat{\overline{Y}}_{GREG}$ under one of SRSWOR, LMS and any HEπPS sampling designs. Then, it follows from the proofs of

Propositions 3.2.1 and 3.2.3 that Γ=$\lim_{\nu\to\infty} nN^{-2}\sum_{i=1}^{N}(V_i - T_V\pi_i)\otimes(V_i - T_V\pi_i)(\pi_i^{-1}-1)$ *a.s.* [**P**], where T_V=$\sum_{i=1}^{N} V_i(1-\pi_i)/\sum_{i=1}^{N}\pi_i(1-\pi_i)$ and $\{\pi_i\}_{i=1}^{N}$ are inclusion probabilities. Further, V_i is Y_i for $\hat{\overline{Y}}$ being $\hat{\overline{Y}}_{HT}$. Also, V_i is $Y_i - \overline{Y} - S_{zy}((Z_i - \overline{Z})S_{zz}^{-1})$ for $\hat{\overline{Y}}$ being $\hat{\overline{Y}}_{GREG}$. Here, S_{zy}=$\sum_{i=1}^{N}(Z_i-\overline{Z})\otimes(Y_i-\overline{Y})/N$ and S_{zz}=$\sum_{i=1}^{N}(Z_i-\overline{Z})^T(Z_i-\overline{Z})/N$. We estimate Γ by

$$\hat{\Gamma} = (nN^{-2})\sum_{i\in s}(\hat{V}_i - \hat{T}_V\pi_i)\otimes(\hat{V}_i - \hat{T}_V\pi_i)(\pi_i^{-1}-1)\pi_i^{-1}, \tag{3.2.3}$$

where $\hat{V}_i$ is Y_i or $Y_i - \hat{\overline{Y}}_{HT} - \hat{S}_{zy}((Z_i - \hat{\overline{Z}}_{HT})\hat{S}_{zz}^{-1})$ for $\hat{\overline{Y}}$ being $\hat{\overline{Y}}_{HT}$ or $\hat{\overline{Y}}_{GREG}$, respectively. Also, $\hat{T}_V$=$\sum_{i\in s}\hat{V}_i(\pi_i^{-1}-1)/\sum_{i\in s}(1-\pi_i)$, $\hat{S}_{zz}$=$\sum_{i\in s}\pi_i^{-1}(Z_i-\hat{\overline{Z}})^T(Z_i-\hat{\overline{Z}})/\sum_{i\in s}\pi_i^{-1}$, and $\hat{S}_{zy}$=$\sum_{i\in s}\pi_i^{-1}(Z_i-\hat{\overline{Z}})\otimes(Y_i-\hat{\overline{Y}})/\sum_{i\in s}\pi_i^{-1}$.

Next, suppose that $\hat{\overline{Y}}$ is either $\hat{\overline{Y}}_{RHC}$ or $\hat{\overline{Y}}_{GREG}$ under RHC sampling design. Then, it can be shown from the proofs of Propositions 3.2.2 and 3.2.3 that Γ=$\lim_{\nu\to\infty} n\gamma\overline{X}N^{-1}\sum_{i=1}^{N}\big(V_i - X_i\overline{V}/\overline{X}\big)\otimes\big(V_i - X_i\overline{V}/\overline{X}\big)X_i^{-1}$ *a.s.* [**P**], where γ=$\sum_{r=1}^{n}\tilde{N}_r(\tilde{N}_r-1)/N(N-1)$ with $\tilde{N}_r$ being the size of the r^{th} group formed randomly in the first step of the RHC sampling design (see the introduction), r=$1,\ldots,n$. Further, V_i is Y_i for $\hat{\overline{Y}}$ being $\hat{\overline{Y}}_{RHC}$. Also, V_i is $Y_i - \overline{Y} - S_{zy}((Z_i - \overline{Z})S_{zz}^{-1})$ for $\hat{\overline{Y}}$ being $\hat{\overline{Y}}_{GREG}$. In this case, we estimate Γ by

$$\hat{\Gamma} = n\gamma(\overline{X}N^{-1})\sum_{i\in s}\Big(\hat{V}_i - X_i\hat{\overline{V}}_{RHC}/\overline{X}\Big)\otimes\Big(\hat{V}_i - X_i\hat{\overline{V}}_{RHC}/\overline{X}\Big)(G_iX_i^{-2}), \tag{3.2.4}$$

where $\hat{V}_i$ is Y_i or $Y_i - \hat{\overline{Y}}_{RHC} - \hat{S}_{zy}((Z_i - \hat{\overline{Z}}_{RHC})\hat{S}_{zz}^{-1})$ for $\hat{\overline{Y}}$ being $\hat{\overline{Y}}_{RHC}$ or $\hat{\overline{Y}}_{GREG}$, respectively. Further, $\hat{\overline{V}}_{RHC}$=$\sum_{i\in s}(NX_i)^{-1}G_i\hat{V}_i$, $\hat{\overline{Z}}_{RHC}$=$\sum_{i\in s}(NX_i)^{-1}G_iZ_i$ and $\hat{S}_{zy}$ and $\hat{S}_{zz}$ are the same as above with π_i^{-1} replaced by $G_iX_i^{-1}$. Also, recall b from Assumption 3.2.2. Now, we state the following theorem concerning the consistency of $\hat{\Gamma}$ as an estimator of Γ with respect to the HS norm (see [45]).

Theorem 3.2.4. *Let us consider Γ, the asymptotic covariance operator of $\sqrt{n}(\hat{\overline{Y}} - \overline{Y})$, and its estimator $\hat{\Gamma}$ from the preceding discussion. Suppose that Assumptions 3.2.1–3.2.3 hold. Then, a.s. [**P**], under SRSWOR and LMS sampling design, $\hat{\Gamma} \xrightarrow{p} \Gamma$ as $\nu\to\infty$. Here, the convergence in probability holds with respect to the HS norm. Further, if Assumption 3.2.1 holds with $0 \le \lambda < E_{\boldsymbol{P}}(X_i)/b$, and Assumptions 3.2.2 and 3.2.3 hold, then the same result holds under any HEπPS sampling design. Moreover, if Assumptions 3.2.1–3.2.4 hold, then the above result holds under RHC sampling design.*

3.3. Data analysis

3.3.1 Analysis based on synthetic data

In this section, we consider a finite population of size N=1000 generated as follows. We first generate the observations $X_1, \ldots, X_N$ on the size variable x from a gamma distribution with mean 500 and standard deviation (s.d.) 100. Here, we assume that the covariate z and the size variable x are same. Then, we generate the population observations on y from $L^2[0,1]$ using linear regression models $Y_i(t)$=$1000 + \beta(t)X_i + \epsilon_i(t)X_i^{\eta}$, where $\beta(t)$=1, t and $1-(t-0.5)^2$, η=$k/10$ for k=$0, 1, \ldots, 10$, and $\{\epsilon_i(t)\}_{t\in[0,1]}$'s are i.i.d. copies of standard Brownian motion with mean 0 and covariance kernel $\sigma(s,t)$=$s \wedge t$. The population observations on y are generated at $t_1, \ldots, t_r$, where r=100 and t_j=jr^{-1} for j=$1, \ldots, r$. We now consider the estimation of the mean of y. We compare the HT and the GREG estimators under SRSWOR and RS sampling design, and the RHC and the GREG estimators under RHC sampling design in terms of relative efficiencies as defined in the following paragraph. The RS sampling design is chosen as a HEπPS sampling design since it is easier to implement than any other HEπPS sampling design. We shall not report the results under LMS sampling design because these results are very close to the results under SRSWOR as expected from our theoretical results.

Suppose that each curve in a population of N curves from $L^2[0,1]$ is observed at $t_1, \ldots, t_r \in [0,1]$ for some $r > 1$. Let us consider I samples each of size n from this population. Then, the MSE of an estimator of $\overline{Y}$, say $\hat{\overline{Y}}$, under sampling design $P(s)$ is computed as $MSE(\hat{\overline{Y}}, P)$=$(rI)^{-1}\sum_{l=1}^{I}\sum_{j=1}^{r}(\hat{\overline{Y}}_l(t_j) - \overline{Y}(t_j))^2$ (see [12], [14], etc.), where $\hat{\overline{Y}}_l$ is an estimate of $\overline{Y}$ based on the l^{th} sample, l=$1, \ldots, I$. Further, we define the relative efficiency of an estimator $\hat{\overline{Y}}_1$ under sampling design $P_1(s)$ compared to another estimator $\hat{\overline{Y}}_2$ under sampling design $P_2(s)$ by

$$RE(\hat{\overline{Y}}_1, P_1|\hat{\overline{Y}}_2, P_2) = MSE(\hat{\overline{Y}}_2, P_2)/MSE(\hat{\overline{Y}}_1, P_1).$$

We say that $\hat{\overline{Y}}_1$ under $P_1(s)$ is more efficient than $\hat{\overline{Y}}_2$ under $P_2(s)$ if RE($\hat{\overline{Y}}_1$, P_1 | $\hat{\overline{Y}}_2$, P_2)> 1. We compute relative efficiencies of the estimators mentioned in the preceding paragraph based on I=1000 samples each of size n=100. We plot the relative efficiency of the HT estimator compared to the GREG estimator under each of SRSWOR and RS sampling design as well as the relative efficiency of the RHC estimator compared to the GREG estimator under RHC sampling design for different η. We also plot the relative efficiency of the GREG estimator under SRSWOR compared to the GREG estimator under each of RS and RHC sampling designs. We use the R

software for drawing samples as well as computing estimators. For RS sampling design, we use the 'pps' package in R. The results obtained from this analysis are summarized as follows.

(i) It follows from Figures 3.1, 3.2 and 3.3 that the relative efficiency curve of the HT estimator compared to the GREG estimator under each of SRSWOR and RS sampling design and that of the RHC estimator compared to the GREG estimator under RHC sampling design always lie below the $y = 1$ line (dashed line), when $\beta(t)$=$1, t$ or $1 - (t - 0.5)^2$. This implies that the GREG estimator is more efficient than the HT estimator under SRSWOR and RS sampling design, and the GREG estimator is more efficient than the RHC estimator under RHC sampling design for different η. The above results are in conformity with Theorems 3.2.1 and 3.2.2.

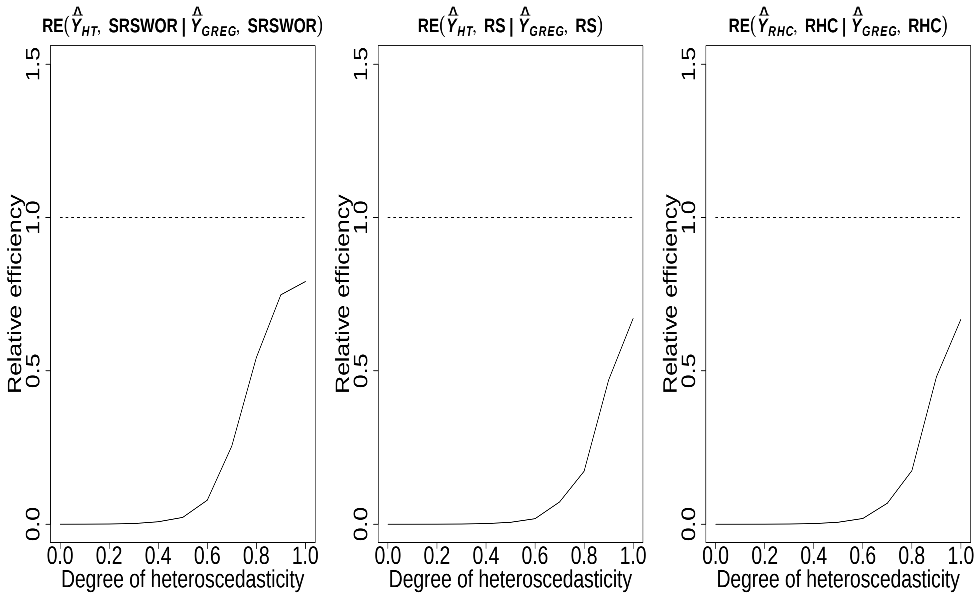

FIGURE 3.1: Comparison of HT, GREG and RHC estimators under different sampling designs for $\beta(t)$=1.

(ii) We see from Figures 3.4, 3.5 and 3.6 that the relative efficiency curve of the GREG estimator under SRSWOR compared to that under each of RS and RHC sampling designs lies above $y = 1$ line, when $\eta < 0.5$ and $\beta(t)$=$1, t$ or $1 - (t - 0.5)^2$. However, these lines lie below $y = 1$ line, when $\eta > 0.5$. This means that the use of the sampling designs like RS and RHC have an adverse effect on the performance of the GREG estimator, when $\eta < 0.5$. However, the use of the above sampling designs improves the performance of the GREG estimator, when $\eta > 0.5$. Thus the above empirical results corroborate the theoretical results stated in Theorem 3.2.3.

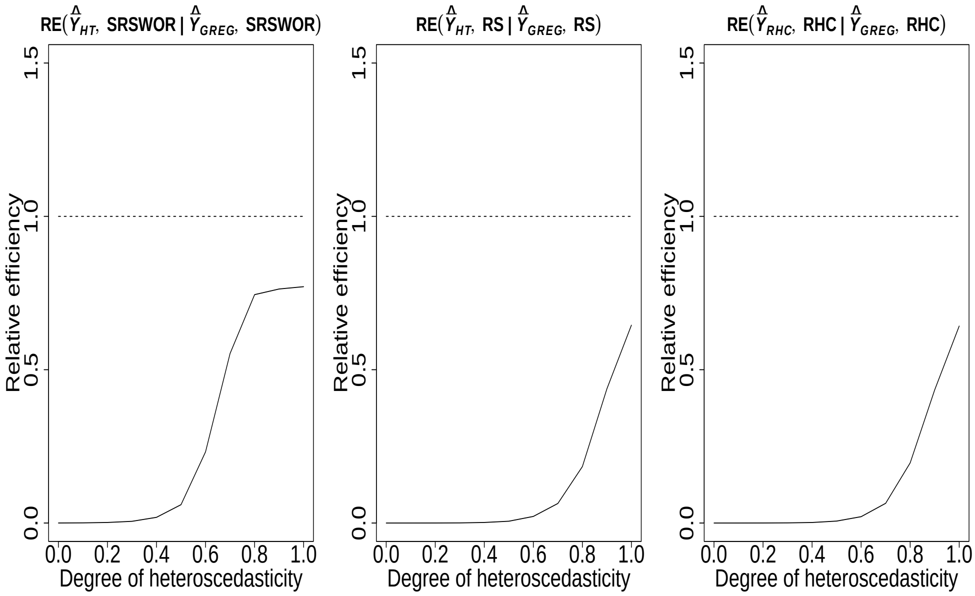

FIGURE 3.2: Comparison of HT, GREG and RHC estimators under different sampling designs for $\beta(t)$=t.

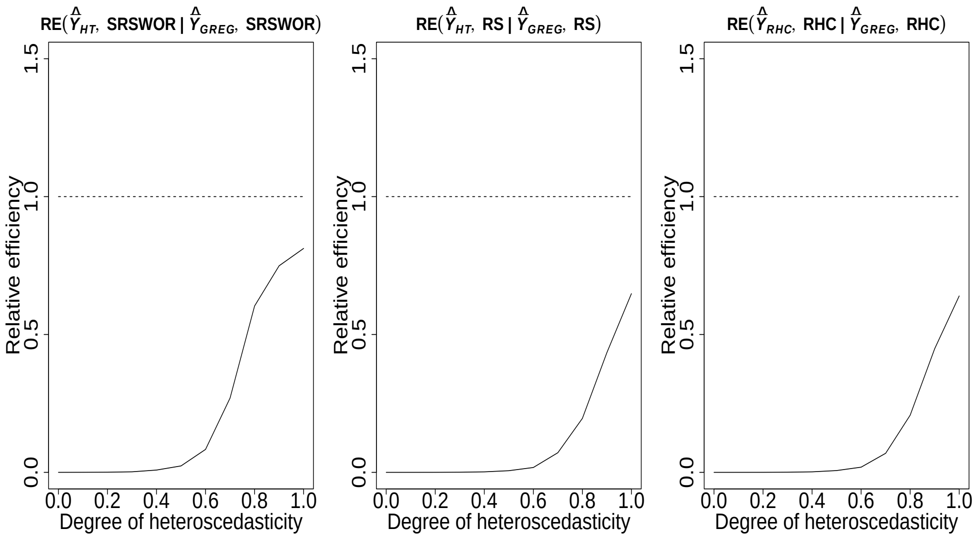

FIGURE 3.3: Comparison of HT, GREG and RHC estimators under different sampling designs for $\beta(t)$=$1-(t-0.5)^2$.

3.3.2 Analysis based on real data

In this section, we consider Electricity Customer Behaviour Trial data available in Irish Social Science Data Archive (ISSDA, *https://www.ucd.ie/issda/*). In this data set, we have electricity consumption of Irish households measured (in kWh) at the end of every half an hour during the period, 14^{th} July in 2009 to 31^{st} December in 2010. We are interested in the estimation of the

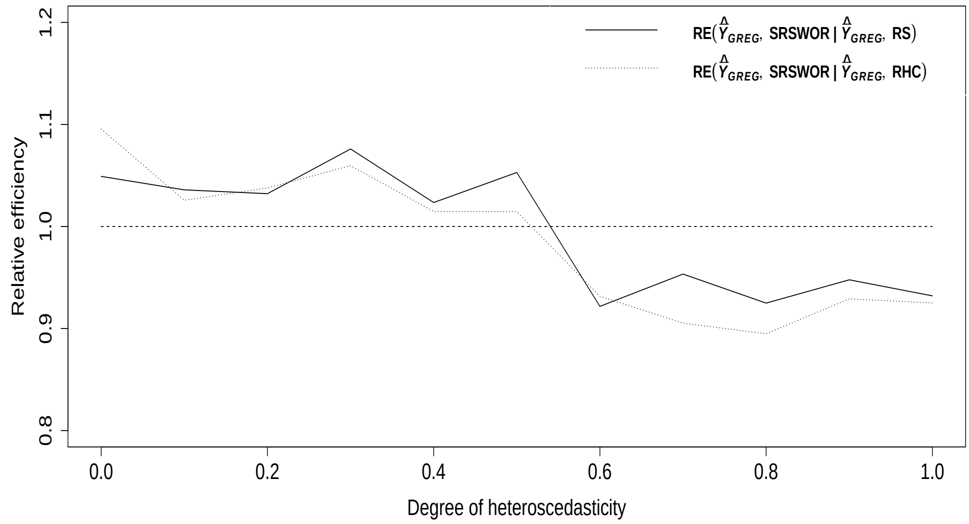

FIGURE 3.4: Comparison of GREG estimators under different sampling designs for $\beta(t)$=1.

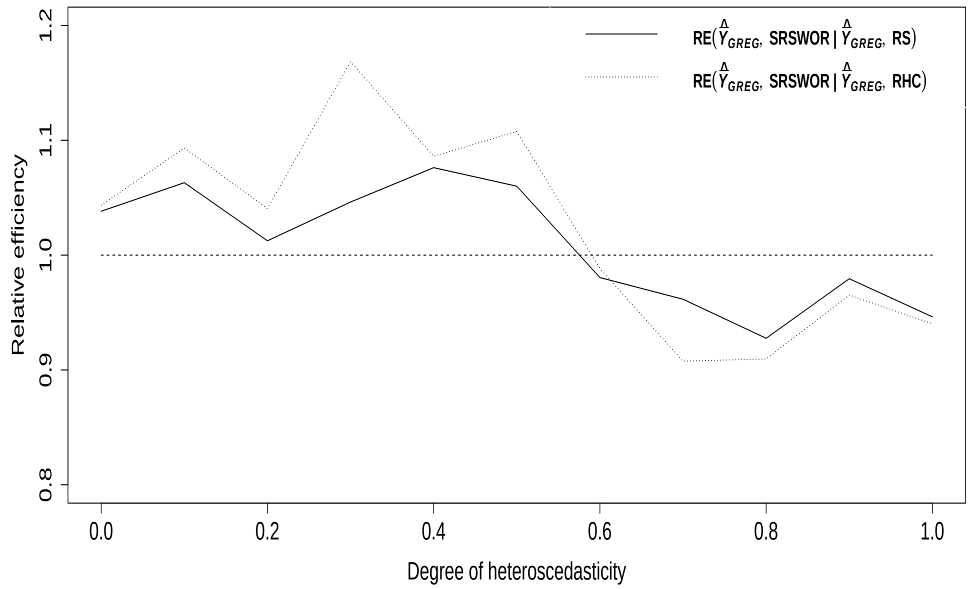

FIGURE 3.5: Comparison of GREG estimators under different sampling designs for $\beta(t)$=t.

mean electricity consumption curve in the summer months, viz. June, July and August in 2010 and in the winter month of December in 2010. It is to be noted that we consider the estimation of the mean electricity consumption curve only in the winter month of December in 2010 because the data for the other two months in the winter of 2010, viz. January and February in 2011 are unavailable. In this data set, we have N=5372 households for which electricity consumption

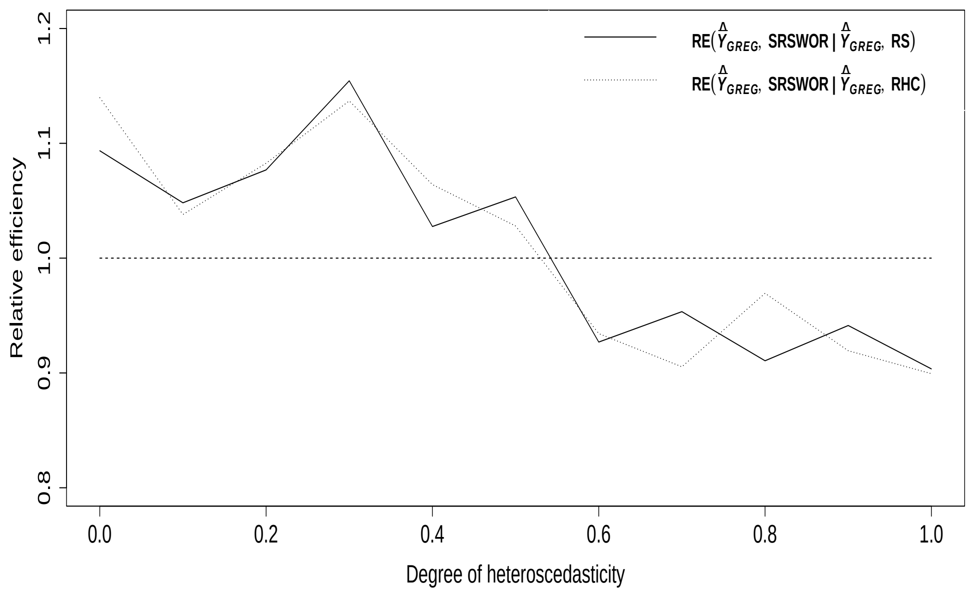

FIGURE 3.6: Comparison of GREG estimators under different sampling designs for $\beta(t)$=1 $-$ $(t-0.5)^2$.

data are available during July and August of 2009 and all the summer months of 2010. We also have N=5092 households for which electricity consumption data are available during December of both 2009 and 2010. Further, for each unit, there are 4416 and 1488 measurement points in summer months and December of 2010, respectively. Electricity consumption in summer months and December of 2010 can be viewed as electricity consumption curves in $L^2[0, T_1]$ and $L^2[0, T_2]$, respectively, where T_1=30 $\times$ 4416=132480 and T_2=30 $\times$ 1488=44640. For estimating the mean electricity consumption curve in the summer months of 2010, we choose the mean electricity consumption in July and August of 2009 as the size variable x, the mean electricity consumption in July of 2009 as the first covariate z_1 and the mean electricity consumption in August of 2009 as the second covariate z_2. On the other hand, for estimating the mean electricity consumption curve in December of 2010, we choose the mean electricity consumption in December of 2009 as both the size variable x and the covariate z. In case of the above estimation problems, we compare the estimators considered in the preceding section in terms of relative efficiencies (see Section 3.3.1). We compute relative efficiencies of these estimators based on I=1000 samples each of size n=100, where these samples are selected from the two data sets consisting of 5372 and 5092 observations, respectively. The results obtained from this analysis are summarized as follows.

TABLE 3.2: Relative efficiencies of the HT, the GREG and the RHC estimators under various sampling designs.

Relative efficiency	Jun, July and August in 2010	December in 2010
RE($\hat{\overline{Y}}_{GREG}$, SRSWOR | $\hat{\overline{Y}}_{HT}$, SRSWOR)	1.529	1.805
RE($\hat{\overline{Y}}_{GREG}$, RS | $\hat{\overline{Y}}_{HT}$, RS)	1.427	1.263
RE($\hat{\overline{Y}}_{GREG}$, RHC | $\hat{\overline{Y}}_{RHC}$, RHC)	1.531	1.251

TABLE 3.3: Relative efficiencies of the GREG estimator under various sampling designs.

Relative efficiency	Jun, July and August in 2010	December in 2010
RE($\hat{\overline{Y}}_{GREG}$, RS | $\hat{\overline{Y}}_{GREG}$, SRSWOR)	2.32	1.76
RE($\hat{\overline{Y}}_{GREG}$, RS | $\hat{\overline{Y}}_{GREG}$, RHC)	1.018	1.012

(i) We see from Table 3.2 that the GREG estimator is more efficient than the HT estimator under SRSWOR and RS sampling design in both the data sets. Also, the GREG estimator is more efficient than the RHC estimator under RHC sampling design in both the data sets. Therefore, these results support the results stated in Theorems 3.2.1 and 3.2.2.

(ii) In the cases of both the data sets, we observe the presence of substantial heteroscedasticity in electricity consumption data, when we plot each of the first three principal components (PC) of electricity consumption data against the size variable (see Figures 3.7 and 3.8). Further, it follows from Table 3.3 that the GREG estimator under RS sampling design is more efficient than any other estimator under any other sampling design for both the data sets. Thus the empirical results stated here are in conformity with the theoretical results stated in Theorem 3.2.3.

3.4. Determining the degree of heteroscedasticity η

In this section, we provide two methods for checking whether the degree of heteroscedasticity η in the linear regression model in (3.2.2) in Section 3.2 is bigger than 0.5 or smaller than 0.5 based on a pilot survey using SRSWOR. In the first method, we estimate η based on some non-parametric estimation methods. In the second method, we choose η based on statistical tests of heteroscedasticity.

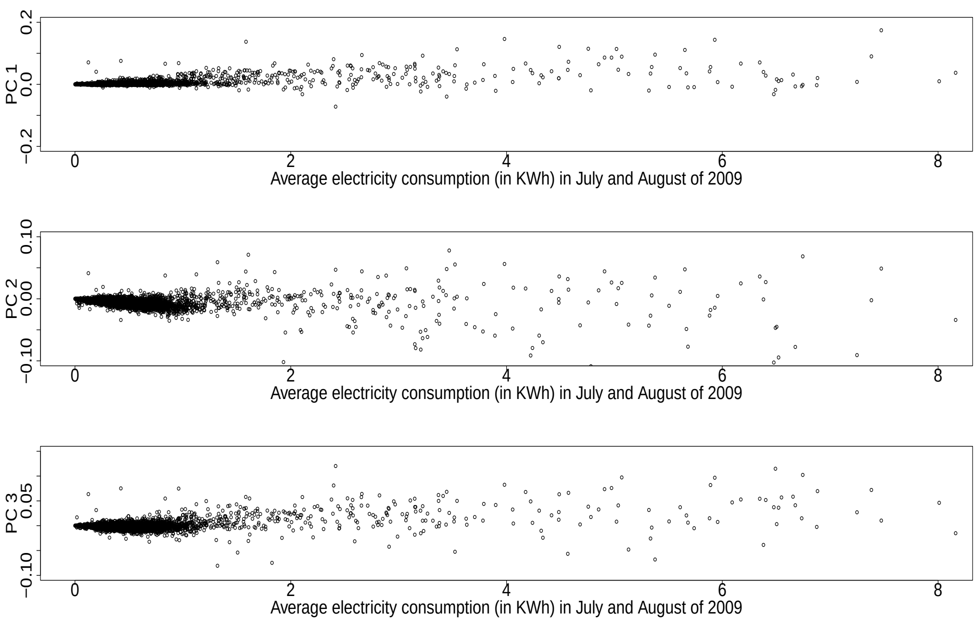

FIGURE 3.7: Scatter plots of the first three principal components of electricity consumption data versus the size variable.

3.4.1 Estimation of η

Under the linear regression model in (3.2.2), we have the conditional total variance $tr(cov_{\mathbf{P}}(Y_i|Z_i, X_i))$= $tr(cov_{\mathbf{P}}(\epsilon_i))X_i^{2\eta}$, where tr denotes the trace of an operator, and $cov_{\mathbf{P}}(Y_i|Z_i, X_i)$ is the conditional covariance operator of Y_i given (Z_i, X_i). Thus according to the linear model (3.2.2), $\log\big(tr(cov_{\mathbf{P}}(Y_i|Z_i, X_i))\big)$ and $log(X_i)$ are linearly related with the slope 2η. Now, in the case of $\mathcal{H}$=$L^2[0,T]$, we have $tr(cov_{\mathbf{P}}(Y_i|Z_i, X_i))$=$\int_{[0,T]} var_{\mathbf{P}}(Y_i(t)|Z_i, X_i)dt$, where $var_{\mathbf{P}}(Y_i(t)|Z_i, X_i)$ is the conditional variance of $Y_i(t)$ given (Z_i, X_i). Suppose that the observations $\{(Y_i, Z_i, X_i) : 1 \leq i \leq N\}$ in the population are generated from the linear model in (3.2.2) and the observations on the study variable y are obtained at $t_1, \ldots, t_r$ in $[0, T]$. Further, suppose that s is a sample of size n drawn based on a pilot survey using SRSWOR. Then, we estimate $tr(cov_{\mathbf{P}}(Y_i|Z_i, X_i))$ based on $\big\{\big(Y_i(t_l), Z_i, X_i\big) : i \in s, l = 1, \ldots, r\big\}$ as follows. For any $i \in s$ and l=$1, \ldots, r$, we first construct the local average estimator of $E_{\mathbf{P}}(Y_i(t_l)|Z_i, X_i)$, the conditional mean of $Y_i(t_l)$ given (Z_i, X_i), as

$$\hat{E}_{\mathbf{P}}(Y_i(t_l)|Z_i, X_i) = \sum_{k\in s}\prod_{j=1}^{d} \mathbb{1}_{[|Z_{ji}-Z_{jk}|\leq h_{1l}]}\mathbb{1}_{[|X_i-X_k|\leq h_{1l}]}Y_k(t_l) \Bigg/ \sum_{k\in s}\prod_{j=1}^{d} \mathbb{1}_{[|Z_{ji}-Z_{jk}|\leq h_{1l}]}\mathbb{1}_{[|X_i-X_k|\leq h_{1l}]}. \quad (3.4.1)$$

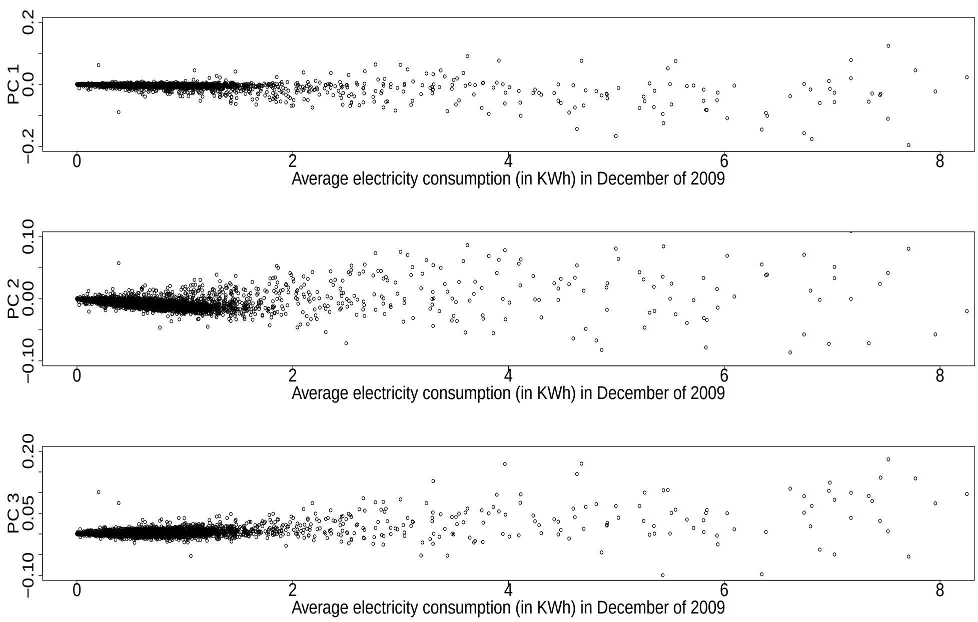

FIGURE 3.8: Scatter plots of the first three principal components of electricity consumption data versus the size variable.

Here, Z_{ji} is j^{th} component of Z_i. For any given l=$1,\ldots,r$, we compute the bandwidth h_{1l} using leave one out cross validation based on $\{(Y_i(t_l), Z_i, X_i) \, : \, i \in s\}$. Now, using $\{\hat{E}_{\mathbf{P}}(Y_i(t_l)|Z_i, X_i) : i \in s\}$, we estimate $var_{\mathbf{P}}(Y_i(t_l)|Z_i, X_i)$ by local sample variance

$$\widehat{var}_{\mathbf{P}}(Y_i(t_l)|Z_i, X_i) = \sum_{k\in s}\prod_{j=1}^{d} \mathbb{1}_{[|Z_{ji}-Z_{jk}|\le h_{2l}]}\mathbb{1}_{[|X_i-X_k|\le h_{2l}]}\times \big(Y_k(t_l) - \hat{E}_{\mathbf{P}}(Y_k(t_l)|Z_k, X_k)\big)^2 \Big/ \sum_{k\in s}\prod_{j=1}^{d} \mathbb{1}_{[|Z_{ji}-Z_{jk}|\le h_{2l}]}\mathbb{1}_{[|X_i-X_k|\le h_{2l}]} \tag{3.4.2}$$

for any $i \in s$ and l=$1,\ldots,r$. We compute the bandwidth h_{2l} based on $\big\{\big(\big(Y_i(t_l) - \hat{E}_{\mathbf{P}}(Y_i(t_l)|Z_i, X_i)\big)^2, Z_i, X_i\big) : i \in s\big\}$ using leave one out cross validation in the same way as we compute the bandwidth h_{1l}. Now, given $\big\{\widehat{var}_{\mathbf{P}}(Y_i(t_l)|Z_i, X_i) \, : \, i \in s, l = 1,\ldots,r\big\}$, we estimate $tr(cov_{\mathbf{P}}(Y_i|Z_i, X_i))$ by $Tr^{-1}\sum_{l=1}^{r}\widehat{var}_{\mathbf{P}}(Y_i(t_l)|Z_i, X_i)$ for any $i \in s$. Then, we fit a least square regression line to the data $\big\{\big(log\big(Tr^{-1}\sum_{l=1}^{r}\widehat{var}_{\mathbf{P}}(Y_i(t_l)|Z_i, X_i)\big), log(X_i)\big) : i \in s\big\}$, and compute the slope of this line. The slope, say $\hat{\theta}$, is expected to be close to 2η if the linear model in (3.2.2) holds. Thus $\hat{\eta}$=$0.5\hat{\theta}$ can be considered as an estimator of η. We demonstrate this method based on real and synthetic data as follows.

(i) Let us first consider the data sets from Section 3.3.2. Recall from Section 3.3.2 that in the case of the estimation of the mean electricity consumption curve in June, July and

August of 2010, we have r=4416. On the other hand, in the case of the estimation of the mean electricity consumption curve in the December of 2010, we have r=1488. Also, recall that T=30r in the cases of the estimation problems for both the data sets. We draw I=100 samples each of size n=500 from these populations using SRSWOR and estimate η as above based on these samples. Then, we compute the proportion of cases, when $\hat{\eta} > 0.5$. It follows that this proportion is 0.72 in the case of the estimation of the mean electricity consumption curve in June, July and August of 2010 and 0.76 in the case of the estimation of the mean electricity consumption curve in the December of 2010. Recall from Section 3.3.2 that in the cases of the estimation problems for both the data sets, the GREG estimator under RS sampling design is more efficient than any other estimator under any other sampling design when compared in terms of relative efficiencies. These corroborate the results stated in Theorem 3.2.3.

(ii) Next, suppose that finite populations each of size N=5000 are generated from linear models in the same way as in Section 3.3.1. Recall from Section 3.3.1 that r=100, T=1 and η=0.1k for k=0, . . . , 10 in this case. We draw I=100 samples each of size n=500 from these populations using SRSWOR. Based on each sample s, we estimate η. Now, suppose that $\hat{\eta}_{lk}$ is the estimate of $0.1k$ based on the l^{th} sample for k=0, . . . , 10 and l=1, . . . , I. Then, we compute the proportion $l^{-1}\#\{l : \hat{\eta}_{lk} \leq 0.5\}$ for different η's and $\beta(t)$'s (see Section 3.3.1) in Table 3.4. It follows from Table 3.4 that for the values of η smaller than 0.5, the proportions are close to 1. On the other hand, these proportions gradually decrease and become 0, when η becomes larger than 0.5. Once again, these corroborate the results stated in Theorem 3.2.3.

3.4.2 Tests for η

Under the linear regression model in (3.2.2), in the case of $\mathcal{H}$=$L^2[0, T]$, we have $X_i^{-\eta} \int_{[0,T]} Y_i(t)dt$= $X_i^{-\eta} \int_{[0,T]} \beta_0(t)dt + \sum_{j=1}^{d} \big(\int_{[0,T]} \beta_j(t)dt\big) Z_{ji} X_i^{-\eta} + \int_{[0,T]} \epsilon_i(t)dt$ for i=1, . . . , N. As in the preceding section, suppose that observations on the study variable y are obtained at $t_1, \ldots, t_r$ in $[0, T]$, and s is a sample of size n drawn based on a pilot survey using SRSWOR. Then we can say that $\{(\tilde{Y}_i X_i^{-\eta}, (1, Z_i) X_i^{-\eta}) : i \in s\}$ are generated from a homoscedastic linear model. Here, $\tilde{Y}_i$=$\int_{[0,T]} Y_i(t)dt$ for $i \in s$. We approximate $\tilde{Y}_i$ by $\hat{Y}_i$=$Tr^{-1} \sum_{l=1}^{r} Y_i(t_l)$. Next, for every η in $\{0.1k : k = 0, \ldots, 10\}$, we test the null hypothesis $H_{0,\eta}$: the data $\{(\hat{Y}_i X_i^{-\eta}, (1, Z_i) X_i^{-\eta}) : i \in s\}$ are generated from a homoscedastic linear model against the alternative hypothesis $H_{1,\eta}$:

TABLE 3.4: Proportion of cases when $\hat{\eta} \leq 0.5$ for different η's and $\beta(t)$'s in the case of synthetic data.

η	$\beta(t)$=1	$\beta(t)$=t	$\beta(t)$=$1-(t-0.5)^2$
0	1	1	1
0.1	1	1	1
0.2	1	1	1
0.3	1	0.99	0.98
0.4	0.99	0.95	0.96
0.5	0.9	0.92	0.94
0.6	0.52	0.56	0.59
0.7	0.2	0.24	0.2
0.8	0.01	0.02	0
0.9	0	0	0
1	0	0	0

heteroscedasticity is present in the data $\{(\hat{Y}_i X_i^{-\eta}, (1, Z_i)X_i^{-\eta}) : i \in s\}$. For this purpose, we use the Breusch-Pagan (BP, see [11]), the White (see [86]) and the Glejser (see [38]) tests because these are some well-known tests for heteroscedasticity. In these tests, the residuals obtained from the ordinary least square regression between the response and the explanatory variables are expressed in terms of explanatory variables by means of different parametric models, and it is checked whether the explanatory variables have any influence on these residuals. Large P-values of the BP, the White and the Glejser tests are indicative of substantial evidence in favour of $H_{0,\eta}$. Thus, we select the η from $\{0.1k : k = 0, \ldots, 10\}$ for which we have the highest P-value. We denote this η by $\hat{\eta}$. Now, we demonstrate this method based on real and synthetic data as follows.

(i) As in the preceding section, let us first consider the data sets used in Section 3.3.2. We draw I=100 samples each of size n=500 from these data sets using SRSWOR and compute $\hat{\eta}$ as above based on each of these samples. Then, for each of the three tests and each of the data sets, we compute the proportion of cases, when $\hat{\eta} > 0.5$ (see Table 3.5). As mentioned in the preceding Section, in the cases of both the estimation problems, the GREG estimator under RS sampling design becomes the most efficient estimator when compared in terms of relative efficiencies. These corroborate the results stated in Theorem 3.2.3.

(ii) Next, we determine η as above based on the synthetic data considered in Section 3.4.1. We draw I=100 samples each of size n=500 from these data sets using SRSWOR and compute $\hat{\eta}$ based on each of these samples. Then, for each of the three tests, every η in

TABLE 3.5: Proportion of cases when $\hat{\eta} > 0.5$ for different tests and data sets in the case of electricity consumption data.

Test	Jun, July and August in 2010	December in 2010
BP	0.79	0.83
White	0.76	0.78
Glejser	0.84	0.8

$\{0.1k : k = 1, \ldots, 10\}$ and each $\beta(t)$ (see Section 3.3.1), we compute the proportion of cases, $\hat{\eta} \leq 0.5$ (see Table 3.6). As in the previous section, it follows from Table 3.6 that for the values of η smaller than 0.5, these proportions are close to 1. On the other hand, these proportions gradually decrease and become 0, when η becomes larger than 0.5. Once again, these corroborate the results stated in Theorem 3.2.3.

TABLE 3.6: Proportion of cases when $\hat{\eta} \leq 0.5$ for different η's and $\beta(t)$'s in the case of synthetic data.

η	$\beta(t)$=1			$\beta(t)$=t			$\beta(t)$=$1-(t-0.5)^2$		
	BP	White	Glejser	BP	White	Glejser	BP	White	Glejser
0	1	1	1	1	1	1	1	1	1
0.1	1	1	1	1	1	1	1	1	1
0.2	0.96	0.99	0.99	0.99	1	1	1	0.97	0.93
0.3	0.95	0.77	0.98	0.9	0.91	0.93	0.98	0.9	0.88
0.4	0.85	0.75	0.84	0.83	0.72	0.88	0.87	0.9	0.79
0.5	0.6	0.65	0.68	0.67	0.58	0.69	0.58	0.69	0.72
0.6	0.47	0.29	0.36	0.43	0.46	0.47	0.39	0.45	0.36
0.7	0.17	0.22	0.15	0.16	0.21	0.14	0.13	0.25	0.17
0.8	0.09	0.07	0.06	0.07	0.06	0.05	0.03	0.04	0.09
0.9	0.01	0.02	0	0.01	0.01	0.01	0.01	0.01	0
1	0	0	0	0	0	0	0	0	0

3.5. Proofs of the main results

In this section, we give the proofs of different Propositions and Theorems. For technical details, which are related to operator theory and used in the proofs of Propositions and Theorems, the reader is referred to [45]. Let us first introduce some notations. Let $\{e_j\}_{j=1}^{\infty}$ be an orthonormal basis of the separable Hilbert space $\mathcal{H}$. Suppose that V_i is either Y_i or

$Y_i - \overline{Y} - S_{zy}((Z_i - \overline{Z})S_{zz}^{-1})$, where S_{zy}=$\sum_{i=1}^{N}(Z_i - \overline{Z}) \otimes (Y_i - \overline{Y})/N$ and S_{zz}=$\sum_{i=1}^{N}(Z_i - \overline{Z})^T(Z_i - \overline{Z})/N$. Further, suppose that $\hat{\overline{V}}_1$=$\sum_{i\in s}(N\pi_i)^{-1}V_i$ and Σ_1=$nN^{-2}\sum_{i=1}^{N}(V_i - T_V\pi_i) \otimes (V_i - T_V\pi_i)(\pi_i^{-1} - 1)$, where T_V=$\sum_{i=1}^{N}V_i(1-\pi_i)/\sum_{i=1}^{N}\pi_i(1-\pi_i)$, and π_i is the inclusion probability of the i^{th} population unit. Moreover, in the case of RHC sampling design, suppose that $\hat{\overline{V}}_2$=$\sum_{i\in s}(NX_i)^{-1}G_iV_i$ and Σ_2=$n\gamma\overline{X}N^{-1}\sum_{i=1}^{N}\left(V_i - X_i\overline{V}/\overline{X}\right) \otimes \left(V_i - X_i\overline{V}/\overline{X}\right)X_i^{-1}$, where $\overline{V}$=$\sum_{i=1}^{N}V_i/N$, $\overline{X}$=$\sum_{i=1}^{N}X_i/N$, G_i is the total of the x values of that randomly formed group from which the i^{th} population unit is selected in the sample by RHC sampling design (see the introduction), and γ=$\sum_{r=1}^{n}\tilde{N}_r(\tilde{N}_r - 1)/N(N-1)$ with $\tilde{N}_r$ being the size of the r^{th} group formed randomly in the first step of the RHC sampling design for r=$1,\ldots,n$. Let us also assume that S_k=$\sqrt{n}(\hat{\overline{V}}_k - \overline{V})$ for k=$1,2$.

Proof of Proposition 3.2.1. Recall the expression of $\hat{\overline{Y}}_{HT}$ from (3.1.1) in Section 3.1 and note that S_1= $\sqrt{n}(\hat{\overline{Y}}_{HT} - \overline{Y})$ if we substitute V_i=Y_i in S_1. It follows from Lemma 3.6.3 in Section 3.6 that $(\langle S_1, e_1\rangle, \ldots, \langle S_1, e_r\rangle) \xrightarrow{\mathcal{L}} N_r(0, \Gamma_{1,r})$ as $\nu \to \infty$ for any $r \geq 1$ under SRSWOR, LMS and any HEπPS sampling designs *a.s.* [**P**]. Here, $\Gamma_{1,r}$ is a $r \times r$ matrix such that $((\Gamma_{1,r}))_{jl}$=$\langle\Gamma_1 e_j, e_l\rangle$, and Γ_1=$\lim_{\nu\to\infty}\Sigma_1$ *a.s.* [**P**]. Further, it follows from the 1^{st} paragraph in the proof of Lemma 3.6.2 in Section 3.6 that Γ_1=Δ_1 for SRSWOR and LMS sampling design, and Γ_1=Δ_2 for any HEπPS sampling design. Here,

$$\begin{aligned}\Delta_1 &= (1-\lambda)E_{\mathbf{P}}(Y_i - E_{\mathbf{P}}(Y_i)) \otimes (Y_i - E_{\mathbf{P}}(Y_i)) \text{ and}\\ \Delta_2 &= E_{\mathbf{P}}\Bigg[\Bigg\{Y_i - \chi^{-1}X_i\Big(E_{\mathbf{P}}(Y_i) - \lambda E_{\mathbf{P}}(X_iY_i)/E_{\mathbf{P}}(X_i)\Big)\Bigg\}\otimes\\ &\Bigg\{Y_i - \chi^{-1}X_i\Big(E_{\mathbf{P}}(Y_i) - \lambda E_{\mathbf{P}}(X_iY_i)/E_{\mathbf{P}}(X_i)\Big)\Bigg\}\Bigg\{X_i^{-1}E_{\mathbf{P}}(X_i) - \lambda\Bigg\}\Bigg]\end{aligned} \tag{3.5.1}$$

with χ=$E_{\mathbf{P}}(X_i) - \lambda E_{\mathbf{P}}(X_i)^2/E_{\mathbf{P}}(X_i)$. Now, suppose that Π_r denotes the orthogonal projection onto the linear span of $\{e_1, \ldots, e_r\}$, i.e., $\Pi_r(a)$=$\sum_{j=1}^{r}\langle a, e_j\rangle e_j$ for any $r \geq 1$ and $a \in \mathcal{H}$. Then, by continuous mapping theorem, $\Pi_r(S_1)$= $\sum_{j=1}^{r}\langle S_1, e_j\rangle e_j \xrightarrow{\mathcal{L}} \mathcal{N}_1 \circ \Pi_r^{-1}$ as $\nu \to \infty$ under the above sampling designs for any $r \geq 1$ *a.s.* [**P**], where $\mathcal{N}_1$ is the Gaussian distribution in $\mathcal{H}$ with mean 0 and covariance operator Γ_1. Moreover, in view of Lemma 3.6.4 in Section 3.6, we have $\lim_{r\to\infty}\overline{\lim}_{\nu\to\infty}\sum_{s\in B_{1,r}}P(s,\omega)$=$0$ *a.s.* [**P**], where $P(s,\omega)$ denotes one of the above sampling designs. Then, by Proposition 2.1 in [54], $\sqrt{n}(\hat{\overline{Y}}_{HT} - \overline{Y}) \xrightarrow{\mathcal{L}} \mathcal{N}_1$ as $\nu \to \infty$ under the above sampling designs *a.s.* [**P**]. □

Proof of Proposition 3.2.2. Recall the expression of $\hat{\overline{Y}}_{RHC}$ from (3.1.2) in Section 3.1 and note that S_2= $\sqrt{n}(\hat{\overline{Y}}_{RHC} - \overline{Y})$ if we substitute V_i=Y_i in S_2. It follows in view of Lemma

3.6.3 in Section 3.6 that under RHC sampling design, $(\langle S_2, e_1\rangle, \ldots, \langle S_2, e_r\rangle) \xrightarrow{\mathcal{L}} N_r(0, \Gamma_{2,r})$ as $\nu \to \infty$ for any $r \geq 1$ *a.s.* [**P**]. Here, $\Gamma_{2,r}$ is a $r \times r$ matrix such that $((\Gamma_{2,r}))_{jl}$=$\langle \Gamma_2 e_j, e_l\rangle$, and Γ_2=$\lim_{\nu\to\infty} \Sigma_2$ *a.s.* [**P**]. Further, it follows from the 2^{nd} paragraph in the proof of Lemma 3.6.2 in Section 3.6 that Γ_2=Δ_3. Here,

$$\Delta_3 = c\big\{E_{\mathbf{P}}(X_i)E_{\mathbf{P}}\big((Y_i \otimes Y_i)X_i^{-1}\big) - E_{\mathbf{P}}(Y_i) \otimes E_{\mathbf{P}}(Y_i)\big\} \tag{3.5.2}$$

with c=$\lim_{\nu\to\infty} n\gamma > 0$. It is to be noted that $n\gamma \to c$ as $\nu \to \infty$ for some $c \geq 1-\lambda > 0$ by Lemma 2.7.5 in Section 2.7 of Chapter 2. Therefore, by continuous mapping theorem, $\Pi_r(S_2)$=$\sum_{j=1}^{r}\langle S_2, e_j\rangle e_j \xrightarrow{\mathcal{L}} \mathcal{N}_2 \circ \Pi_r^{-1}$ as $\nu \to \infty$ under RHC sampling design for any $r \geq 1$ *a.s.* [**P**], where $\mathcal{N}_2$ is the Gaussian distribution in $\mathcal{H}$ with mean 0 and covariance operator Γ_2. Next, it follows from Lemma 3.6.4 in Section 3.6 that $\lim_{r\to\infty} \overline{\lim}_{\nu\to\infty} \sum_{s \in B_{2,r}} P(s,\omega)$=0 *a.s.* [**P**], where $P(s,\omega)$ denotes RHC sampling design. Then, by Proposition 2.1 in [54], $\sqrt{n}(\hat{\overline{Y}}_{RHC} - \overline{Y}) \xrightarrow{\mathcal{L}} \mathcal{N}_2$ as $\nu \to \infty$ under RHC sampling design *a.s.* [**P**]. □

Proof of Proposition 3.2.3. Recall from (3.1.3) in Section 3.1 that $\hat{\overline{Y}}_{GREG}$=$\hat{\overline{Y}} + \hat{S}_{zy}((\overline{Z} - \hat{\overline{Z}})\hat{S}_{zz}^{-1})$, where $\hat{\overline{Y}}$=$\sum_{i\in s}\pi_i^{-1}Y_i/\sum_{i\in s}\pi_i^{-1}$, $\hat{\overline{Z}}$=$\sum_{i\in s}\pi_i^{-1}Z_i/\sum_{i\in s}\pi_i^{-1}$, $\hat{S}_{zz}$=$\sum_{i\in s}\pi_i^{-1}(Z_i - \hat{\overline{Z}})^T(Z_i - \hat{\overline{Z}})/\sum_{i\in s}\pi_i^{-1}$, and $\hat{S}_{zy}$=$\sum_{i\in s}\pi_i^{-1}(Z_i - \hat{\overline{Z}}) \otimes (Y_i - \hat{\overline{Y}})/\sum_{i\in s}\pi_i^{-1}$. Note that

$$\hat{\overline{Y}}_{GREG} - \overline{Y} = \Theta(\hat{\overline{V}}_1 - \overline{V}) + B, \tag{3.5.3}$$

where $\hat{\overline{V}}_1$=$\sum_{i\in s}(N\pi_i)^{-1}V_i$, V_i=$Y_i - \overline{Y} - S_{zy}((Z_i - \overline{Z})S_{zz}^{-1})$, Θ=$(\sum_{i\in s}\pi_i^{-1})^{-1}$, B=$S_{zy}((\hat{\overline{Z}} - \overline{Z})S_{zz}^{-1}) - \hat{S}_{zy}((\hat{\overline{Z}} - \overline{Z})\hat{S}_{zz}^{-1})$, S_{zy}=$\sum_{i=1}^{N}(Z_i - \overline{Z}) \otimes (Y_i - \overline{Y})/N$, and S_{zz}=$\sum_{i=1}^{N}(Z_i - \overline{Z})^T(Z_i - \overline{Z})/N$. Using Lemmas 3.6.3 and 3.6.4 in Section 3.6, it can be shown in the same way as in the proof of Proposition 3.2.1 that as $\nu \to \infty$, $\sqrt{n}(\hat{\overline{V}}_1 - \overline{V}) \xrightarrow{\mathcal{L}} \mathcal{N}_3$ under SRSWOR, LMS and any HEπPS sampling designs *a.s.* [**P**], where $\mathcal{N}_3$ is the Gaussian distribution in $\mathcal{H}$ with mean 0 and covariance operator Γ_1. Here, Γ_1=$\lim_{\nu\to\infty}\Sigma_1$ *a.s.* [**P**]. It follows from the last paragraph in the proof of Lemma 3.6.2 in Section 3.6 that Γ_1=Δ_4 under SRSWOR and LMS sampling design, and Γ_1=Δ_5 under any HEπPS sampling design. Here,

$$\Delta_4 = (1-\lambda)E_{\mathbf{P}}\Bigg\{\bigg(Y_i - E_{\mathbf{P}}(Y_i) - C_{zy}\big((Z_i - E_{\mathbf{P}}(Z_i))C_{zz}^{-1}\big)\bigg)\otimes \\ \bigg(Y_i - E_{\mathbf{P}}(Y_i) - C_{zy}\big((Z_i - E_{\mathbf{P}}(Z_i))C_{zz}^{-1}\big)\bigg)\Bigg\} \text{ and} \tag{3.5.4}$$

$$\Delta_5 = E_{\mathbf{P}}\bigg[\Big\{Y_i - E_{\mathbf{P}}(Y_i) - C_{zy}\big((Z_i - E_{\mathbf{P}}(Z_i))C_{zz}^{-1}\big)+ \\ \chi^{-1}X_i\lambda\Big(E_{\mathbf{P}}(X_iY_i) - E_{\mathbf{P}}(X_i)E_{\mathbf{P}}(Y_i) - C_{zy}\big((E_{\mathbf{P}}(X_iZ_i) - E_{\mathbf{P}}(X_i)E_{\mathbf{P}}(Z_i))\times \\ C_{zz}^{-1}\big)\Big)(E_{\mathbf{P}}(X_i))^{-1}\Big\} \otimes \Big\{Y_i - E_{\mathbf{P}}(Y_i) - C_{zy}\big((Z_i - E_{\mathbf{P}}(Z_i))C_{zz}^{-1}\big)+ \\ \chi^{-1}X_i\lambda\Big(E_{\mathbf{P}}(X_iY_i) - E_{\mathbf{P}}(X_i)E_{\mathbf{P}}(Y_i) - C_{zy}\big((E_{\mathbf{P}}(X_iZ_i) - E_{\mathbf{P}}(X_i)E_{\mathbf{P}}(Z_i))\times \\ C_{zz}^{-1}\big)\Big)(E_{\mathbf{P}}(X_i))^{-1}\Big\}\Big\{X_i^{-1}E_{\mathbf{P}}(X_i) - \lambda\Big\}\bigg] \tag{3.5.5}$$

with χ=$E_{\mathbf{P}}(X_i) - \lambda E_{\mathbf{P}}(X_i)^2/E_{\mathbf{P}}(X_i)$. Now, to establish the weak convergence of $\sqrt{n}(\hat{\overline{Y}}_{GREG} - \overline{Y})$ under the above sampling designs *a.s.* [**P**], it is enough to show that $\Theta \xrightarrow{p} 1$ and $\sqrt{n}B \xrightarrow{p} 0$ under these sampling designs as $\nu \to \infty$ *a.s.* [**P**].

Suppose that $||\cdot||_{op}$ denotes the operator norm. Note that except the operator norm, we use only the HS norm for the operators considered in this chapter and denote it by $||\cdot||_{HS}$. Also, note that

$$||B||_{\mathcal{H}} \leq (||S_{zz}^{-1}||_{op}||S_{zy} - \hat{S}_{zy}||_{op} + ||\hat{S}_{zy}||_{op}||S_{zz}^{-1} - \hat{S}_{zz}^{-1}||_{op})||\hat{\overline{Z}} - \overline{Z}||. \tag{3.5.6}$$

It follows in view of Lemma 3.6.5 in Section 3.6 that as $\nu \to \infty$,

$$\bigg|\bigg|\sum_{i\in s}(N\pi_i)^{-1}(Y_i \otimes Z_i) - \sum_{i=1}^{N}(Y_i \otimes Z_i)/N\bigg|\bigg|_{HS} = o_p(1), \bigg|\bigg|\sum_{i\in s}(N\pi_i)^{-1}Z_i^TZ_i - \\ \sum_{i=1}^{N}Z_i^TZ_i/N\bigg|\bigg| = o_p(1), \sqrt{n}\bigg|\bigg|\hat{\overline{Z}}_1 - \overline{Z}\bigg|\bigg| = O_p(1), \text{ and } \sum_{i\in s}(N\pi_i)^{-1} - 1 = o_p(1) \tag{3.5.7}$$

under the sampling designs considered in the previous paragraph *a.s.* [**P**], where $\hat{\overline{Z}}_1$=$\sum_{i\in s}(N\pi_i)^{-1}$ $\times Z_i$. Consequently, in view of Assumption 3.2.3,

$$\sqrt{n}\bigg|\bigg|\hat{\overline{Z}} - \overline{Z}\bigg|\bigg| = O_p(1), \bigg|\bigg|\hat{S}_{zz} - S_{zz}\bigg|\bigg|_{op} \leq \bigg|\bigg|\hat{S}_{zz} - S_{zz}\bigg|\bigg| = o_p(1) \text{ and} \\ \bigg|\bigg|\hat{S}_{zy} - S_{zy}\bigg|\bigg|_{op} \leq \bigg|\bigg|\hat{S}_{zy}^* - S_{zy}^*\bigg|\bigg|_{HS} = o_p(1) \tag{3.5.8}$$

as $\nu \to \infty$ under these sampling designs *a.s.* [**P**]. Here, $\hat{S}_{zy}^*$=$\sum_{i\in s}\pi_i^{-1}(Y_i - \hat{\overline{Y}}) \otimes (Z_i - \hat{\overline{Z}})/\sum_{i\in s}\pi_i^{-1}$ and S_{zy}^*=$\sum_{i=1}^{N}(Y_i - \overline{Y}) \otimes (Z_i - \overline{Z})/N$ are adjoints of $\hat{S}_{zy}$ and S_{zy}, respectively. Now, recall C_{zz} and C_{zy} from the 2^{nd} paragraph in the proof of Lemma 3.6.2 in Section 3.6. Note that $||S_{zz} - C_{zz}||$=$o(1)$ and $||S_{zy} - C_{zy}||_{HS}$=$o(1)$ as $\nu \to \infty$ *a.s.* [**P**] in view of Assumption

3.2.3. Also, note that C_{zz}^{-1} exists by Assumption 3.2.3. Consequently, $||S_{zz}^{-1}||_{op}$=$O(1)$, $||\hat{S}_{zz}^{-1} - S_{zz}^{-1}||_{op}$=$o_p(1)$ and $||\hat{S}_{zy}||_{op}$=$O_p(1)$ as $\nu \to \infty$ *a.s.* [**P**]. Thus $\sqrt{n}||B||_{\mathcal{H}}$=$o_p(1)$ and $\Theta - 1$=$o_p(1)$ as $\nu \to \infty$ under the above-mentioned sampling designs *a.s.* [**P**]. Hence, the weak convergence of $\sqrt{n}(\hat{\overline{Y}}_{GREG} - \overline{Y})$ follows under these sampling designs by using Proposition 2.1 in [54].

Let us next consider the RHC sampling design. Recall from Section 3.1 that we consider $\hat{\overline{Y}}_{GREG}$ under RHC sampling design with π_i^{-1} replacing $G_i X_i^{-1}$. Then, under this sampling design,

$$\hat{\overline{Y}}_{GREG} - \overline{Y} = \Theta(\hat{\overline{V}}_2 - \overline{V}) + B, \tag{3.5.9}$$

where $\hat{\overline{V}}_2$=$\sum_{i\in s}(NX_i)^{-1}G_iV_i$ for V_i=$Y_i - \overline{Y} - S_{zy}((Z_i - \overline{Z})S_{zz}^{-1})$, and Θ and B are the same as defined in the 1^{st} paragraph of this proof with π_i^{-1} replaced by $G_iX_i^{-1}$. Using Lemmas 3.6.3 and 3.6.4 in Section 3.6, it can be shown in a similar way as in the proof of Proposition 3.2.2 that $\sqrt{n}(\hat{\overline{V}}_2 - \overline{V}) \xrightarrow{\mathcal{L}} \mathcal{N}_4$ as $\nu \to \infty$ under RHC sampling design *a.s.* [**P**], where $\mathcal{N}_4$ is the Gaussian distribution in $\mathcal{H}$ with mean 0 and covariance operator Γ_2. It follows from the last paragraph in the proof of Lemma 3.6.2 in Section 3.6 that Γ_2=Δ_6=$\lim_{\nu\to\infty}\Sigma_2$ *a.s.* [**P**]. Here,

$$\begin{aligned}\Delta_6 = cE_{\mathbf{P}}(X_i)E_{\mathbf{P}}\bigg\{\Big(Y_i - E_{\mathbf{P}}(Y_i) - C_{zy}\big((Z_i - E_{\mathbf{P}}(Z_i))C_{zz}^{-1}\big)\Big)\otimes\\ \Big(Y_i - E_{\mathbf{P}}(Y_i) - C_{zy}\big((Z_i - E_{\mathbf{P}}(Z_i))C_{zz}^{-1}\big)\Big)X_i^{-1}\bigg\}\end{aligned} \tag{3.5.10}$$

with c=$\lim_{\nu\to\infty} n\gamma > 0$. It is to be noted that $n\gamma \to c$ as $\nu \to \infty$ for some $c \geq 1 - \lambda > 0$ by Lemma 2.7.5 in Section 2.7 of Chapter 2. Moreover, using Lemma 3.6.5 in Section 3.6, it can be shown in the same way as in the preceding paragraph of this proof that $\Theta \xrightarrow{p} 1$ and $\sqrt{n}B \xrightarrow{p} 0$ as $\nu \to \infty$ under RHC sampling design *a.s.* [**P**]. Threfore, the weak convergence of $\sqrt{n}(\hat{\overline{Y}}_{GREG} - \overline{Y})$ follows under this sampling design by using Proposition 2.1 in [54]. □

Proof of Theorem 3.2.1. Let us recall the expressions of Δ_1 and Δ_4 from the proofs of Propositions 3.2.1 and 3.2.2, respectively. It follows from the proof of Proposition 3.2.3 that *a.s.* [**P**], $\sqrt{n}(\hat{\overline{Y}}_{GREG} - \overline{Y})$ has the same asymptotic covariance operator Δ_4 under SRSWOR and LMS sampling design. It further follows from the proof of Proposition 3.2.1 that *a.s.* [**P**], the asymptotic covariance operator of $\sqrt{n}(\hat{\overline{Y}}_{HT} - \overline{Y})$ is Δ_1 under SRSWOR as well as LMS sampling design. Let A_i=$\langle Y_i, a\rangle$ for $a \in \mathcal{H}$ and i=$1,\ldots,N$. Then, we have

$$\begin{aligned}\langle(\Delta_1 - \Delta_4)a, a\rangle = (1-\lambda)\big(E_{\mathbf{P}}(A_i - E_{\mathbf{P}}(A_i))^2 - E_{\mathbf{P}}(A_i - E_{\mathbf{P}}(A_i) -\\ C_{za}C_{zz}^{-1}(Z_i - E_{\mathbf{P}}(Z_i))^T)^2\big) = (1-\lambda)C_{za}C_{zz}^{-1}C_{za}^T\end{aligned} \tag{3.5.11}$$

for C_{za}=$E_{\mathbf{P}}(A_i - E_{\mathbf{P}}(A_i))(Z_i - E_{\mathbf{P}}(Z_i))$ and C_{zz}=$E_{\mathbf{P}}(Z_i - E_{\mathbf{P}}(Z_i))^T(Z_i - E_{\mathbf{P}}(Z_i))$. Note that $C_{za}C_{zz}^{-1}C_{za}^T \geq 0$ for any $a \in \mathcal{H}$ by Assumption 3.2.3. In fact, there exists $a \in \mathcal{H}$ such that $a \neq 0$ and C_{za}=0. Therefore, $\Delta_1 - \Delta_4$ is p.s.d. Hence, *a.s.* [**P**], the GREG estimator is asymptotically at least as efficient as the HT estimator under SRSWOR and LMS sampling design. Moreover, *a.s.* [**P**], both the GREG estimator has the same asymptotic distribution under SRSWOR and LMS sampling design. □

Proof of Theorem 3.2.2. Let us recall the expressions of Δ_2, Δ_3, Δ_5 and Δ_6 from the proofs of Propositions 3.2.1–3.2.3. It can be shown from the proofs of Propositions 3.2.2 and 3.2.3 that *a.s.* [**P**], asymptotic covariance operators of $\sqrt{n}(\hat{\overline{Y}}_{RHC} - \overline{Y})$ and $\sqrt{n}(\hat{\overline{Y}}_{GREG} - \overline{Y})$ under RHC sampling design are Δ_3 and Δ_6, respectively. Now, it follows from the linear regression model in (3.2.2) in Section 3.2 that

$$\langle \Delta_3 a, a\rangle = c\Bigg[\mu_x E_{\mathbf{P}}(\tilde{\epsilon}_i)^2 E_{\mathbf{P}}\big(X_i^{2\eta-1}\big) + \mu_x E_{\mathbf{P}}\Bigg(\tilde{\beta}_0 + \sum_{j=1}^{d}\tilde{\beta}_j Z_{ji}\Bigg)^2 X_i^{-1} - \Bigg(\sum_{j=0}^{d}\tilde{\beta}_j\mu_j\Bigg)^2\Bigg] \text{and } \langle \Delta_6 a, a\rangle = c\mu_x E_{\mathbf{P}}(\tilde{\epsilon}_i)^2 E_{\mathbf{P}}\big(X_i^{2\eta-1}\big), \tag{3.5.12}$$

where c=$\lim_{\nu\to\infty} n\gamma > 0$, $a \in \mathcal{H}$, $\tilde{\epsilon}_i$=$\langle \epsilon_i, a\rangle$, μ_x=$E_{\mathbf{P}}(X_i)$, $\tilde{\beta}_j$=$\langle \beta_j, a\rangle$ for j=0, …, d, μ_0=1, and μ_j=$E_{\mathbf{P}}(Z_{ji})$ for j=1, …, d. Therefore,

$$\langle (\Delta_3 - \Delta_6)a, a\rangle = c\mu_x E_{\mathbf{P}}\Bigg(\tilde{\beta}_0 + \sum_{j=1}^{d}\tilde{\beta}_j Z_{ji} - X_i\sum_{j=0}^{d}\tilde{\beta}_j\mu_j\mu_x^{-1}\Bigg)^2 X_i^{-1} \geq 0 \tag{3.5.13}$$

for any $a \in \mathcal{H}$. Thus $\Delta_3 - \Delta_6$ is n.n.d. Hence, *a.s.* [**P**], the GREG estimator is asymptotically at least as efficient as the RHC estimator under RHC sampling design. Next, it follows from the proofs of Propositions 3.2.1 and 3.2.3 that *a.s.* [**P**], asymptotic covariance operators of $\sqrt{n}(\hat{\overline{Y}}_{HT} - \overline{Y})$ and $\sqrt{n}(\hat{\overline{Y}}_{GREG} - \overline{Y})$ under any HEπPS sampling design are Δ_2 and Δ_5, respectively. Further, it follows from the linear regression model in (3.2.2) in Section 3.2 that

$$\langle \Delta_2 a, a\rangle = \Bigg[E_{\mathbf{P}}(\tilde{\epsilon}_i)^2\Big\{\mu_x E_{\mathbf{P}}\big(X_i^{2\eta-1}\big) - \lambda E_{\mathbf{P}}\big(X_i^{2\eta}\big)\Big\} + E_{\mathbf{P}}\Bigg\{\Bigg(\tilde{\beta}_0 + \sum_{j=1}^{d}\tilde{\beta}_j Z_{ji}\Bigg)^2 \times \big(X_i^{-1}\mu_x - \lambda\big)\Bigg\} - \chi^{-1}\mu_x^{-1}\Bigg\{(1-\lambda)\tilde{\beta}_0\mu_x + \Bigg(\sum_{j=1}^{d}\tilde{\beta}_j\big(\mu_j\mu_x - \lambda\mu_{jx}\big)\Bigg\}^2\Bigg] \tag{3.5.14}$$
$$\text{and } \langle \Delta_5 a, a\rangle = E_{\mathbf{P}}(\tilde{\epsilon}_i)^2\Big(\mu_x E_{\mathbf{P}}\big(X_i^{2\eta-1}\big) - \lambda E_{\mathbf{P}}\big(X_i^{2\eta}\big)\Big),$$

where μ_{jx}=$E_{\mathbf{P}}(Z_{ji}X_i)$ for j=$1,\ldots,d$ and χ=$\mu_x - \lambda E_{\mathbf{P}}(X_i)^2(\mu_x)^{-1}$. Now, since Assumption 3.2.2 holds and $0 \leq \lambda \leq \mu_x b^{-1}$, we have

$$\begin{aligned}&\langle(\Delta_2 - \Delta_5)a, a\rangle = E_{\mathbf{P}}\Bigg[\Bigg\{\Bigg(\tilde{\beta}_0 + \sum_{j=1}^{d}\tilde{\beta}_j Z_{ji}\Bigg) - \chi^{-1}X_i\Bigg(\sum_{j=0}^{d}\tilde{\beta}_j\mu_j - \lambda\tilde{\beta}_0 - \\ &\sum_{j=1}^{d}\lambda\tilde{\beta}_j\mu_{jx}\mu_x^{-1}\Bigg)\Bigg\}^2\big(X_i^{-1}\mu_x - \lambda\big)\Bigg] \geq 0\end{aligned} \tag{3.5.15}$$

Thus using similar arguments as above, we can say that *a.s.* [**P**], the GREG estimator is asymptotically at least as efficient as the HT estimator under any HEπPS sampling design. □

Proof of Theorem 3.2.3. Recall from the proofs of Theorems 3.2.1 and 3.2.2 that *a.s.* [**P**], the asymptotic covariance operators of the GREG estimator under SRSWOR, any HEπPS sampling design and RHC sampling design are Δ_4, Δ_5 and Δ_6, respectively. Also, recall from (3.5.12) and (3.5.14) in the proof of Theorem 3.2.2 that

$$\begin{aligned}&\langle\Delta_5 a, a\rangle = E_{\mathbf{P}}(\tilde{\epsilon}_i)^2\Big(\mu_x E_{\mathbf{P}}\big(X_i^{2\eta-1}\big) - \lambda E_{\mathbf{P}}\big(X_i^{2\eta}\big)\Big) \text{ and } \langle\Delta_6 a, a\rangle \\ &= c\mu_x E_{\mathbf{P}}(\tilde{\epsilon}_i)^2 E_{\mathbf{P}}\big(X_i^{2\eta-1}\big)\end{aligned} \tag{3.5.16}$$

for any $a \in \mathcal{H}$ under the linear regression model in (3.2.2) in Section 3.2. It can be further shown using (3.2.2) in Section 3.2 and (3.6.10) in the proof of Lemma 3.6.2 in Section 3.6 that

$$\langle\Delta_4 a, a\rangle = (1-\lambda)E_{\mathbf{P}}(\tilde{\epsilon}_i)^2 E_{\mathbf{P}}(X_i^{2\eta}) \tag{3.5.17}$$

for any $a \in \mathcal{H}$. Therefore, we have

$$\begin{aligned}&\langle(\Delta_4 - \Delta_5)a, a\rangle = E_{\mathbf{P}}(\tilde{\epsilon}_i)^2 cov_{\mathbf{P}}\Big(X_i^{2\eta-1}, X_i\Big) \\ &\langle(\Delta_6 - \Delta_5)a, a\rangle = E_{\mathbf{P}}(\tilde{\epsilon}_i)^2\Big(\lambda E_{\mathbf{P}}\big(X_i^{2\eta}\big) - (1-c)E_{\mathbf{P}}\big(X_i^{2\eta-1}\big)\mu_x\Big) \text{ and} \\ &\langle(\Delta_4 - \Delta_6)a, a\rangle = E_{\mathbf{P}}(\tilde{\epsilon}_i)^2\Big((1-\lambda)E_{\mathbf{P}}\big(X_i^{2\eta}\big) - cE_{\mathbf{P}}\big(X_i^{2\eta-1}\big)\mu_x\Big)\end{aligned} \tag{3.5.18}$$

for any $a \in \mathcal{H}$. Note that $E_{\mathbf{P}}(\tilde{\epsilon}_i)^2$=$\langle E_{\mathbf{P}}(\epsilon_i \otimes \epsilon_i)a, a\rangle > 0$ for any $a \in \mathcal{H}$ since $E_{\mathbf{P}}(\epsilon_i \otimes \epsilon_i)$ is p.d. Also, note that $cov_{\mathbf{P}}\big(X_i^{2\eta-1}, X_i\big) > 0$ for $\eta > 0.5$, $cov_{\mathbf{P}}\big(X_i^{2\eta-1}, X_i\big) = 0$ for $\eta = 0.5$ and $cov_{\mathbf{P}}\big(X_i^{2\eta-1}, X_i\big) < 0$ for $\eta < 0.5$. Further, it follows from Lemma 2.7.5 in Section 2.7 of Chapter 2 that c=1 for λ=0, c=$1-\lambda$ for $\lambda > 0$ and λ^{-1} an integer, and $c > 1-\lambda$ when $\lambda > 0$

and λ^{-1} is a non-integer. Therefore, the results in Table 3.7 below hold, and hence the results stated in Table 3.1 hold.

TABLE 3.7: Relations among Δ_4, Δ_5 and Δ_6.

	λ=0	$\lambda > 0$ & λ^{-1} is an integer	$\lambda > 0$ & λ^{-1} is a non-integer
$\eta < 0.5$	$\Delta_5 - \Delta_4$ and $\Delta_6 - \Delta_4$ are p.d.	$\Delta_5 - \Delta_4$ and $\Delta_6 - \Delta_4$ are p.d.	$\Delta_5 - \Delta_4$ and $\Delta_6 - \Delta_4$ are p.d.
$\eta = 0.5$	Δ_4=Δ_5=Δ_6	Δ_4=Δ_5=Δ_6	Δ_4=Δ_5 and $\Delta_6 - \Delta_4$ is p.d.
$\eta > 0.5$	Δ_5=Δ_6 and $\Delta_4 - \Delta_5$ is p.d.	$\Delta_4 - \Delta_5$ and $\Delta_6 - \Delta_5$ are p.d.	$\Delta_4 - \Delta_5$ and $\Delta_6 - \Delta_5$ are p.d.

Next, if we put λ=0 and c=1, respectively, in the expressions of Δ_5 and Δ_6 in the proof of Lemma 3.6.2 in Section 3.6, we have Δ_5=Δ_6. Thus *a.s.* [**P**], the GREG estimator has the same asymptotic covariance operator under RHC and any HEπPS sampling designs. Hence, *a.s.* [**P**], the GREG estimator has the same asymptotic distribution under RHC and any HEπPS sampling designs. This completes the proof of the theorem. $\square$

Proof of Theorem 3.2.4. Recall the expression of $\hat{\Gamma}$ from (3.2.3) in Section 3.2 and note that

$$\hat{\Gamma} = (nN^{-2})\bigg(\sum_{i\in s}(\hat{V}_i \otimes \hat{V}_i)(\pi_i^{-1} - 1)\pi_i^{-1} - \sum_{i\in s}(1 - \pi_i)\hat{T}_V \otimes \hat{T}_V\bigg) \tag{3.5.19}$$

with $\hat{T}_V$=$\sum_{i\in s}\hat{V}_i(\pi_i^{-1} - 1)/\sum_{i\in s}(1 - \pi_i)$. Let us first consider the case, when Γ denotes the asymptotic covariance operator of $\sqrt{n}(\hat{\overline{Y}}_{HT} - \overline{Y})$ and $\hat{\Gamma}$ is its estimator. Then, we have $\hat{V}_i$=Y_i in $\hat{\Gamma}$. Now, recall the expression of Σ_1 from the beginning of this section and note that

$$\Sigma_1 = (nN^{-2})\bigg(\sum_{i=1}^{N}(V_i \otimes V_i)(\pi_i^{-1} - 1) - \sum_{i=1}^{N}\pi_i(1 - \pi_i)T_V \otimes T_V\bigg) \tag{3.5.20}$$

with T_V=$\sum_{i=1}^{N} V_i(1 - \pi_i)/\sum_{i=1}^{N}\pi_i(1 - \pi_i)$. Let us substitute V_i=Y_i in Σ_1. We shall first show that under SRSWOR, LMS and any HEπPS sampling designs, $\hat{\Gamma} - \Sigma_1 \xrightarrow{p} 0$ with respect to the HS norm as $\nu \to \infty$ *a.s.* [**P**]. It follows by Assumption 3.2.3 that $\sum_{i=1}^{N}||Y_i||_{\mathcal{H}}^2/N$=$O(1)$ as $\nu \to \infty$ *a.s.* [**P**]. It also follows by (3.6.1) in the statement of Lemma 3.6.1 in Section 3.6 that as $\nu \to \infty$, $\sum_{i=1}^{N}\big(N\pi_i(1 - \pi_i)/n\big)^2/N$=$O(1)$ under the above sampling designs *a.s.* [**P**]. Then, using the same line of arguments as in the proof of Lemma 3.6.5 in Section 3.6, it can be shown that

$$\Big(\sum_{i\in s}(1-\pi_i)-\sum_{i=1}^{N}\pi_i(1-\pi_i)\Big)\Big/n=o_p(1) \text{ and}$$
$$\Big|\Big|\sum_{i\in s}\hat{V}_i(\pi_i^{-1}-1)-\sum_{i=1}^{N}V_i(1-\pi_i)\Big|\Big|_{\mathcal{H}}\Big/N=o_p(1) \tag{3.5.21}$$

as $\nu\to\infty$ *a.s.* [**P**]. Moreover, $\sum_{i=1}^{N}\pi_i(1-\pi_i)/n$ is bounded away from 0 as $\nu\to\infty$ *a.s.* [**P**] because (3.6.1) and Assumption 3.2.1 hold. Consequently, under all of the above-mentioned sampling designs, $(nN^{-2})\big(\sum_{i\in s}(1-\pi_i)(\hat{T}_V\otimes\hat{T}_V)-\sum_{i=1}^{N}\pi_i(1-\pi_i)(T_V\otimes T_V)\big)\xrightarrow{p}0$ with respect to the HS norm as $\nu\to\infty$ *a.s.* [**P**]. Similarly, $(nN^{-2})\big(\sum_{i\in s}(\hat{V}_i\otimes\hat{V}_i)(\pi_i^{-1}-1)\pi_i^{-1}-\sum_{i=1}^{N}(V_i\otimes V_i)(\pi_i^{-1}-1)\big)\xrightarrow{p}0$ with respect to the HS norm as $\nu\to\infty$ *a.s.* [**P**]. Thus under the above sampling designs, $\hat{\Gamma}-\Sigma_1\xrightarrow{p}0$ with respect to the HS norm as $\nu\to\infty$ *a.s.* [**P**]. Recall from Section 3.2 that Γ=$\lim_{\nu\to\infty}\Sigma_1$ *a.s.* [**P**]. Therefore, under the aforesaid sampling designs, $\hat{\Gamma}\xrightarrow{p}\Gamma$ with respect to the HS norm as $\nu\to\infty$ *a.s.* [**P**].

Let us next consider the case, when Γ denotes the asymptotic covariance operator of $\sqrt{n}(\hat{\overline{Y}}_{GREG}-\overline{Y})$ and $\hat{\Gamma}$ denotes its estimator. Then, $\hat{\Gamma}$ is the same as described in the preceding paragraph with $\hat{V}_i$=$Y_i-\hat{\overline{Y}}_{HT}-\hat{S}_{zy}((Z_i-\hat{\overline{Z}}_{HT})\hat{S}_{zz}^{-1})$. Let us also consider Σ_1 with V_i=$Y_i-\overline{Y}-S_{zy}((Z_i-\overline{Z})S_{zz}^{-1})$. Note that

$$\Big(\sum_{i\in s}\hat{V}_i(\pi_i^{-1}-1)-\sum_{i=1}^{N}V_i(1-\pi_i)\Big)\Big/N=\sum_{i\in s}(\hat{V}_i-V_i)(\pi_i^{-1}-1)/N+$$
$$\Big(\sum_{i\in s}V_i(\pi_i^{-1}-1)-\sum_{i=1}^{N}V_i(1-\pi_i)\Big)\Big/N. \tag{3.5.22}$$

It can be shown in the same way as in the proof of Lemma 3.6.5 in Section 3.6 that $||(\sum_{i\in s}V_i\times(\pi_i^{-1}-1)-\sum_{i=1}^{N}V_i(1-\pi_i))/N||_{\mathcal{H}}$=$o_p(1)$ under the sampling designs considered in the previous paragraph as $\nu\to\infty$ *a.s.* [**P**]. Further, it can be shown that $||\sum_{i\in s}(\hat{V}_i-V_i)(\pi_i^{-1}-1)/N||_{\mathcal{H}}$=$o_p(1)$ as $\nu\to\infty$ *a.s.* [**P**] since $||\hat{\overline{Y}}_{HT}-\overline{Y}||_{\mathcal{H}}$=$o_p(1)$, $||\hat{S}_{zy}-S_{zy}||_{op}$=$o_p(1)$, $||\hat{S}_{zz}^{-1}-S_{zz}^{-1}||_{op}$=$o_p(1)$, $||\hat{S}_{zy}||_{op}$=$O_p(1)$ and $||S_{zz}^{-1}||_{op}$=$O(1)$ as $\nu\to\infty$ *a.s.* [**P**] (see the proof of Proposition 3.2.3). Then, $(nN^{-2})(\sum_{i\in s}(1-\pi_i)(\hat{T}_V\otimes\hat{T}_V)-\sum_{i=1}^{N}\pi_i(1-\pi_i)(T_V\otimes T_V))\xrightarrow{p}0$ with respect to the HS norm as $\nu\to\infty$ *a.s.* [**P**]. Similarly, $(nN^{-2})(\sum_{i\in s}(\hat{V}_i\otimes\hat{V}_i)(\pi_i^{-1}-1)\pi_i^{-1}-\sum_{i=1}^{N}(V_i\otimes V_i)(\pi_i^{-1}-1))\xrightarrow{p}0$ with respect to the HS norm as $\nu\to\infty$ *a.s.* [**P**]. Hence, under the above sampling designs, $\hat{\Gamma}-\Sigma_1\xrightarrow{p}0$, and hence $\hat{\Gamma}\xrightarrow{p}\Gamma$ with respect to the HS norm as $\nu\to\infty$ *a.s.* [**P**].

Next, consider the case, when Γ denotes the asymptotic covariance operator of $\sqrt{n}(\hat{\overline{Y}}_{RHC} - \overline{Y})$ or $\sqrt{n}(\hat{\overline{Y}}_{GREG} - \overline{Y})$ under RHC sampling design, and $\hat{\Gamma}$ denotes its estimator. Recall from (3.2.4) in Section 3.2 that in this case,

$$\hat{\Gamma} = n\gamma(\overline{X}N^{-1})\sum_{i\in s}\Big(\hat{V}_i - X_i\hat{\overline{V}}_{RHC}/\overline{X}\Big)\otimes\Big(\hat{V}_i - X_i\hat{\overline{V}}_{RHC}/\overline{X}\Big)(G_iX_i^{-2}) = n\gamma\bigg((\overline{X}N^{-1})\sum_{i\in s}(\hat{V}_i\otimes\hat{V}_i)G_iX_i^{-2} - \hat{\overline{V}}_{RHC}\otimes\hat{\overline{V}}_{RHC}\bigg). \tag{3.5.23}$$

Also, recall the expression of Σ_2 from the beginning of this section and note that

$$\Sigma_2 = n\gamma\bigg((\overline{X}N^{-1})\sum_{i=1}^{N}(V_i\otimes V_i)X_i^{-1} - \overline{V}\otimes\overline{V}\bigg). \tag{3.5.24}$$

Then, it can be shown in a similar way as in the earlier cases that under RHC sampling design, $\hat{\Gamma} - \Sigma_2 \xrightarrow{p} 0$ with respect to the HS norm as $\nu\to\infty$ *a.s.* [**P**]. Therefore, under RHC sampling design, $\hat{\Gamma} \xrightarrow{p} \Gamma$ with respect to the HS norm as $\nu\to\infty$ *a.s.* [**P**] because Γ=$\lim_{\nu\to\infty}\Sigma_2$ *a.s.* [**P**] (see Section 3.2). □

3.6. Proofs of additional results required to prove the main results

In this section, we state and prove some technical lemmas, which will be required to prove our main results.

Lemma 3.6.1. *Suppose that Assumption 3.2.2 holds. Then, LMS sampling design is a high entropy sampling design. Moreover, under each of SRSWOR, LMS and any HEπPS sampling designs, we have, for all sufficiently large ν,*

$$L \le N\pi_i/n \le L' \text{ for some constants } L, L' > 0 \text{ and all } 1\le i\le N \text{ a.s. } [\boldsymbol{P}]. \tag{3.6.1}$$

Lemma 3.6.1 is similar to Lemma 2.7.1 in Chapter 2.

Proof. The proof of the above Lemma follows exactly the same way as the proof of Lemma 2.7.1. □

Before we state the next lemma, let us recall $\{e_j\}_{j=1}^{\infty}$, $\{V_i\}_{i=1}^{N}$ Σ_1 and Σ_2 from the paragraph preceding the proof of Proposition 3.2.1 in Section 3.5. Let us also recall b from Assumption 3.2.2. We now state the following lemma.

Lemma 3.6.2. *Suppose that Assumptions 3.2.1–3.2.3 hold. Then, under SRSWOR and LMS sampling design, $\Sigma_1 \to \Gamma_1$ with respect to the HS norm as $\nu \to \infty$ a.s. [**P**] for some n.n.d. HS operator Γ_1. Also, $\sum_{j=1}^{\infty}\langle \Gamma_1 e_j, e_j\rangle < \infty$, and $\sum_{j=1}^{\infty}\langle \Sigma_1 e_j, e_j\rangle \to \sum_{j=1}^{\infty}\langle \Gamma_1 e_j, e_j\rangle$ under the above sampling designs as $\nu \to \infty$ a.s. [**P**]. Further, if Assumption 3.2.1 holds with $0 \leq \lambda < E_{\mathbf{P}}(X_i)/b$, and Assumptions 3.2.2 and 3.2.3 hold, then, the above results hold under any HEπPS sampling design. Moreover, if Assumptions 3.2.1–3.2.4 hold, then in the case of RHC sampling design, $\Sigma_2 \to \Gamma_2$ with respect to the HS norm as $\nu \to \infty$ a.s. [**P**] for some n.n.d. HS operator Γ_2. Also, $\sum_{j=1}^{\infty}\langle \Gamma_2 e_j, e_j\rangle < \infty$, and $\sum_{j=1}^{\infty}\langle \Sigma_2 e_j, e_j\rangle \to \sum_{j=1}^{\infty}\langle \Gamma_2 e_j, e_j\rangle$ as $\nu \to \infty$ a.s. [**P**].*

Proof. Let us first consider the case V_i=Y_i for i=$1,\ldots,N$. Then, we have

$$\Sigma_1 = nN^{-2}\sum_{i=1}^{N}(V_i - T_V\pi_i)\otimes(V_i - T_V\pi_i)(\pi_i^{-1}-1) = nN^{-2}\Big\{\sum_{i=1}^{N}(Y_i\otimes Y_i)\times (\pi_i^{-1}-1) - \Big(\sum_{i=1}^{N}Y_i(1-\pi_i)\otimes\sum_{i=1}^{N}Y_i(1-\pi_i)\Big)\Big/\sum_{i=1}^{N}\pi_i(1-\pi_i)\Big\}. \tag{3.6.2}$$

Now, substituting π_i=n/N for SRSWOR, we obtain Σ_1=$(1-n/N)\sum_{i=1}^{N}(Y_i-\overline{Y})\otimes(Y_i-\overline{Y})/N$. Note that $E_{\mathbf{P}}||Y_i||_{\mathcal{H}}^2 < \infty$ in view of Assumption 3.2.3. Then, under SRSWOR,

$$\Sigma_1 \to \Delta_1 = (1-\lambda)E_{\mathbf{P}}(Y_i - E_{\mathbf{P}}(Y_i))\otimes(Y_i - E_{\mathbf{P}}(Y_i)) \tag{3.6.3}$$

with respect to the HS norm as $\nu \to \infty$ *a.s.* [**P**] by SLLN and Assumption 3.2.1. Now, suppose that $\Sigma_1^{(1)}$ and $\Sigma_1^{(2)}$ denote Σ_1 under SRSWOR and LMS sampling design, respectively. Further, suppose that $\{\pi_i\}_{i=1}^{N}$ are the inclusion probabilities of LMS sampling design. Then, we have

$$\Sigma_1^{(2)} - \Sigma_1^{(1)} = nN^{-2}\Big\{\sum_{i=1}^{N}(\pi_i^{-1} - n^{-1}N)(Y_i\otimes Y_i)\Big\} - nN^{-2}\Big\{\Big(\sum_{i=1}^{N}Y_i(1-\pi_i)\otimes\sum_{i=1}^{N}Y_i(1-\pi_i)\Big)\Big/\sum_{i=1}^{N}\pi_i(1-\pi_i) - \Big(\sum_{i=1}^{N}Y_i(1-n/N)\otimes\sum_{i=1}^{N}Y_i(1-n/N)\Big)\Big/n(1-n/N)\Big\} \tag{3.6.4}$$

by (3.6.2). Further, it follows from the proof of Lemma 2.7.1 in Section 2.7 of Chapter 2 that as $\nu \to \infty$, $\max_{1\leq i\leq N} |n^{-1}N\pi_i - 1| \to 0$ *a.s.* [**P**]. It also follows from Assumption 3.2.3 that $N^{-1}\sum_{i=1}^{N} ||Y_i||_{\mathcal{H}}^2 = O(1)$ as $\nu \to \infty$ *a.s.* [**P**]. Therefore, it can be shown that as $\nu \to \infty$,

$$nN^{-2}\Big\{\sum_{i=1}^{N}(\pi_i^{-1} - n^{-1}N)(Y_i \otimes Y_i)\Big\} \to 0 \text{ and} \tag{3.6.5}$$

$$\begin{aligned} &nN^{-2}\Big\{\Big(\sum_{i=1}^{N} Y_i(1-\pi_i) \otimes \sum_{i=1}^{N} Y_i(1-\pi_i)\Big)\Big/\sum_{i=1}^{N}\pi_i(1-\pi_i) - \\ &\Big(\sum_{i=1}^{N} Y_i(1-n/N) \otimes \sum_{i=1}^{N} Y_i(1-n/N)\Big)\Big/n\big(1-n/N\big)\Big\} \to 0, \text{ and hence} \end{aligned} \tag{3.6.6}$$

$\Sigma_1^{(2)} - \Sigma_1^{(1)} \to 0$ with respect to the HS norm *a.s.* [**P**]. Thus $\Sigma_1 \to \Gamma_1$ as $\nu \to \infty$ under SRSWOR as well as under LMS sampling design *a.s.* [**P**] with Γ_1=Δ_1. Next, under any HEπPS sampling design (i.e., a sampling design with π_i=$nX_i/\sum_{i=1}^{N} X_i$),

$$\begin{aligned} &\Sigma_1 \to \Delta_2 = E_{\mathbf{P}}\Bigg[\Big\{Y_i - \chi^{-1}X_i\Big(E_{\mathbf{P}}(Y_i) - \lambda E_{\mathbf{P}}(X_iY_i)/E_{\mathbf{P}}(X_i)\Big)\Big\}\otimes \\ &\Big\{Y_i - \chi^{-1}X_i\Big(E_{\mathbf{P}}(Y_i) - \lambda E_{\mathbf{P}}(X_iY_i)/E_{\mathbf{P}}(X_i)\Big)\Big\}\Big\{X_i^{-1}E_{\mathbf{P}}(X_i) - \lambda\Big\}\Bigg] \end{aligned} \tag{3.6.7}$$

with respect to the HS norm as $\nu \to \infty$ *a.s.* [**P**] by SLLN because $E_{\mathbf{P}}||Y_i||_{\mathcal{H}}^2 < \infty$, Assumptions 3.2.1 and 3.2.2 hold. Here, χ=$E_{\mathbf{P}}(X_i) - \lambda E_{\mathbf{P}}(X_i)^2/E_{\mathbf{P}}(X_i)$. Note that Δ_2 is a n.n.d. HS operator since Assumption 3.2.1 holds with $0 \leq \lambda < E_{\mathbf{P}}(X_i)/b$. Thus as $\nu \to \infty$, $\Sigma_1 \to \Gamma_1$ under any HEπPS sampling design *a.s.* [**P**] with Γ_1=Δ_2. Next, note that $\sum_{j=1}^{\infty}\langle\Delta_1 e_j, e_j\rangle$=$E_{\mathbf{P}}||Y_i - E_{\mathbf{P}}(Y_i)||_{\mathcal{H}}^2 < \infty$ and $\sum_{j=1}^{\infty}\langle\Delta_2 e_j, e_j\rangle$=$E_{\mathbf{P}}\big[||Y_i - \chi^{-1}X_i\{E_{\mathbf{P}}(Y_i) - \lambda E_{\mathbf{P}}(X_iY_i)/E_{\mathbf{P}}(X_i)\}||_{\mathcal{H}}^2\{X_i^{-1}E_{\mathbf{P}}(X_i) - \lambda\}\big] < \infty$ since Assumption 3.2.2 holds, and $E_{\mathbf{P}}||Y_i||_{\mathcal{H}}^2 < \infty$. Then, it can be shown in the same way as argued above that as $\nu \to \infty$, $\sum_{j=1}^{\infty}\langle\Sigma_1 e_j, e_j\rangle$= $nN^{-2}\big\{\sum_{i=1}^{N}(\pi_i^{-1} - 1)||Y_i||_{\mathcal{H}}^2 - \sum_{i=1}^{N} ||Y_i(1-\pi_i)||_{\mathcal{H}}^2/\sum_{i=1}^{N}\pi_i(1-\pi_i)\big\} \to \sum_{j=1}^{\infty}\langle\Delta_1 e_j, e_j\rangle$ under SRSWOR and LMS sampling design, and $\sum_{j=1}^{\infty}\langle\Sigma_1 e_j, e_j\rangle \to \sum_{j=1}^{\infty}\langle\Delta_2 e_j, e_j\rangle$ under any HEπPS sampling design *a.s.* [**P**].

Next, consider the case of RHC sampling design and Σ_2 with V_i=Y_i. Then, we have

$$\Sigma_2 = n\gamma\overline{X}N^{-1}\sum_{i=1}^{N}\Big(V_i - X_i\overline{V}/\overline{X}\Big) \otimes \Big(V_i - X_i\overline{V}/\overline{X}\Big)X_i^{-1} \tag{3.6.8}$$

$$= n\gamma\Big\{\overline{X}N^{-1}\sum_{i=1}^{N}(Y_i\otimes Y_i)X_i^{-1} - \overline{Y}\otimes\overline{Y}\Big\},$$

where γ=$\sum_{r=1}^{n}\tilde{N}_r(\tilde{N}_r-1)/N(N-1)$ with $\tilde{N}_r$ being the size of the r^{th} group formed randomly in the first step of the RHC sampling design (see the introduction) for r=$1,\ldots,n$. Note that $n\gamma\to c$ as $\nu\to\infty$ for some $c\geq 1-\lambda>0$ by Lemma 2.7.5 in Section 2.7 of Chapter 2. Then, by SLLN,

$$\Sigma_2\to\Delta_3 = c\big\{E_{\mathbf{P}}(X_i)E_{\mathbf{P}}\big((Y_i\otimes Y_i)X_i^{-1}\big) - E_{\mathbf{P}}(Y_i)\otimes E_{\mathbf{P}}(Y_i)\big\} \tag{3.6.9}$$

with respect to the HS norm as $\nu\to\infty$ *a.s.* [**P**]. Thus Γ_2=Δ_3 in this case. It follows that $\sum_{j=1}^{\infty}\langle\Delta_3 e_j,e_j\rangle$ =$c\big\{E_{\mathbf{P}}(X_i)E_{\mathbf{P}}(||Y_i||_{\mathcal{H}}^2X_i^{-1}) - ||E_{\mathbf{P}}(Y_i)||_{\mathcal{H}}^2\big\}<\infty$ since Assumption 3.2.2 holds, and $E_{\mathbf{P}}||Y_i||_{\mathcal{H}}^2<\infty$. Further, it can be shown using SLLN that $\sum_{j=1}^{\infty}\langle\Sigma_2 e_j,e_j\rangle$=$n\gamma\times$ $\big\{\overline{X}N^{-1}\sum_{i=1}^{N}||Y_i||_{\mathcal{H}}^2X_i^{-1} - ||\overline{Y}||_{\mathcal{H}}^2\big\}\to\sum_{j=1}^{\infty}\langle\Delta_3 e_j,e_j\rangle$ as $\nu\to\infty$ *a.s.* [**P**].

Let us next consider the case V_i=$Y_i-\overline{Y}-S_{zy}((Z_i-\overline{Z})S_{zz}^{-1})$ for i=$1,\ldots,N$. It follows from SLLN that $\sum_{i=1}^{N}||V_i||_{\mathcal{H}}^2/N$=$O(1)$ as $\nu\to\infty$ *a.s.* [**P**] because Assumption 3.2.3 holds. Then, it can be shown using similar arguments as in the 1^{st} paragraph of this proof that as $\nu\to\infty$,

$$\begin{aligned}&\Sigma_1\to\Delta_4 = (1-\lambda)E_{\mathbf{P}}\Bigg\{\bigg(Y_i - E_{\mathbf{P}}(Y_i) - C_{zy}\big((Z_i-E_{\mathbf{P}}(Z_i))C_{zz}^{-1}\big)\bigg)\otimes\\&\bigg(Y_i - E_{\mathbf{P}}(Y_i) - C_{zy}\big((Z_i-E_{\mathbf{P}}(Z_i))C_{zz}^{-1}\big)\bigg)\Bigg\}\end{aligned} \tag{3.6.10}$$

under SRSWOR and LMS sampling design, and

$$\begin{aligned}&\Sigma_1\to\Delta_5 = E_{\mathbf{P}}\Bigg[\bigg\{Y_i - E_{\mathbf{P}}(Y_i) - C_{zy}\big((Z_i-E_{\mathbf{P}}(Z_i))C_{zz}^{-1}\big)+\\&\chi^{-1}X_i\lambda\bigg(E_{\mathbf{P}}(X_iY_i) - E_{\mathbf{P}}(X_i)E_{\mathbf{P}}(Y_i) - C_{zy}\big((E_{\mathbf{P}}(X_iZ_i) - E_{\mathbf{P}}(X_i)E_{\mathbf{P}}(Z_i))\times\\&C_{zz}^{-1}\big)\bigg)(E_{\mathbf{P}}(X_i))^{-1}\bigg\}\otimes\bigg\{Y_i - E_{\mathbf{P}}(Y_i) - C_{zy}\big((Z_i-E_{\mathbf{P}}(Z_i))C_{zz}^{-1}\big)+\\&\chi^{-1}X_i\lambda\bigg(E_{\mathbf{P}}(X_iY_i) - E_{\mathbf{P}}(X_i)E_{\mathbf{P}}(Y_i) - C_{zy}\big((E_{\mathbf{P}}(X_iZ_i) - E_{\mathbf{P}}(X_i)E_{\mathbf{P}}(Z_i))\times\\&C_{zz}^{-1}\big)\bigg)(E_{\mathbf{P}}(X_i))^{-1}\bigg\}\bigg\{X_i^{-1}E_{\mathbf{P}}(X_i)-\lambda\bigg\}\Bigg]\end{aligned} \tag{3.6.11}$$

under any HEπPS sampling design with respect to the HS norm *a.s.* [**P**]. Here, C_{zy}=$E_{\mathbf{P}}(Z_i - E_{\mathbf{P}}(Z_i))\otimes(Y_i-E_{\mathbf{P}}(Y_i))$ and C_{zz}=$E_{\mathbf{P}}(Z_i-E_{\mathbf{P}}(Z_i))^T(Z_i-E_{\mathbf{P}}(Z_i))$. Thus as $\nu\to\infty$, $\Sigma_1\to\Gamma_1$ with Γ_1=Δ_4 under SRSWOR and LMS sampling design, and $\Sigma_1\to\Gamma_1$ with Γ_1=Δ_5 under any

HEπPS sampling design *a.s.* [**P**]. Note that $\sum_{j=1}^{\infty}\langle\Delta_4 e_j, e_j\rangle$=$(1-\lambda)E_{\mathbf{P}}\big|\big|Y_i-E_{\mathbf{P}}(Y_i)-C_{zy}((Z_i-E_{\mathbf{P}}(Z_i))C_{zz}^{-1})\big|\big|_{\mathcal{H}}^2<\infty$, and $\sum_{j=1}^{\infty}\langle\Delta_5 e_j, e_j\rangle$=$E_{\mathbf{P}}\big[\big|\big|Y_i-E_{\mathbf{P}}(Y_i)-C_{zy}\big((Z_i-E_{\mathbf{P}}(Z_i))C_{zz}^{-1}\big)+\chi^{-1}X_i\lambda\big\{E_{\mathbf{P}}(X_iY_i)-E_{\mathbf{P}}(X_i)E_{\mathbf{P}}(Y_i)-C_{zy}\big((E_{\mathbf{P}}(X_iZ_i)-E_{\mathbf{P}}(X_i)E_{\mathbf{P}}(Z_i))C_{zz}^{-1}\big)\big\}(E_{\mathbf{P}}(X_i))^{-1}\big|\big|_{\mathcal{H}}^2\times\big\{X_i^{-1}E_{\mathbf{P}}(X_i)-\lambda\big\}\big]<\infty$ since Assumptions 3.2.2 and 3.2.3 hold. Then, it can be shown in a similar way as in the 1^{st} paragraph of this proof that $\sum_{j=1}^{\infty}\langle\Sigma_1 e_j, e_j\rangle\rightarrow\sum_{j=1}^{\infty}\langle\Delta_4 e_j, e_j\rangle$ under SRSWOR and LMS sampling design, and $\sum_{j=1}^{\infty}\langle\Sigma_1 e_j, e_j\rangle\rightarrow\sum_{j=1}^{\infty}\langle\Delta_5 e_j, e_j\rangle$ under any HEπPS sampling design as $\nu\rightarrow\infty$ *a.s.* [**P**]. Further, it can be shown using the same line of argument as in the 2^{nd} paragraph of this proof that for RHC sampling design,

$$\Sigma_2\rightarrow\Delta_6=cE_{\mathbf{P}}(X_i)E_{\mathbf{P}}\Bigg\{\bigg(Y_i-E_{\mathbf{P}}(Y_i)-C_{zy}\big((Z_i-E_{\mathbf{P}}(Z_i))C_{zz}^{-1}\big)\bigg)\otimes \\ \bigg(Y_i-E_{\mathbf{P}}(Y_i)-C_{zy}\big((Z_i-E_{\mathbf{P}}(Z_i))C_{zz}^{-1}\big)\bigg)X_i^{-1}\Bigg\} \tag{3.6.12}$$

with respect to the HS norm, and $\sum_{j=1}^{\infty}\langle\Sigma_2 e_j, e_j\rangle\rightarrow\sum_{j=1}^{\infty}\langle\Delta_6 e_j, e_j\rangle$ as $\nu\rightarrow\infty$ *a.s.* [**P**]. Thus Γ_2=Δ_6 in this case. □

Recall $\{e_j\}_{j=1}^{\infty}$, $\{V_i\}_{i=1}^{N}$, S_1, S_2, $\Sigma_{1,r}$ and $\Sigma_{2,r}$ from the paragraph preceding the proof of Proposition 3.2.1 in Section 3.5 and define $\mathbf{W}_i$= $(\langle V_i, e_1\rangle,\ldots,\langle V_i, e_r\rangle)$ for i=$1,\ldots,N$ and $r\geq 1$. Suppose that $\hat{\overline{\mathbf{W}}}_1$= $\sum_{i\in s}(N\pi_i)^{-1}\mathbf{W}_i$ and $\overline{\mathbf{W}}$=$N^{-1}\sum_{i=1}^{N}\mathbf{W}_i$. Moreover, suppose that $\hat{\overline{\mathbf{W}}}_2$=$\sum_{i\in s}(NX_i)^{-1}G_i\mathbf{W}_i$, where G_i is the total of the x values of that randomly formed group from which the i^{th} population unit is selected in the sample by RHC sampling design (see the introduction). Let us also assume that $\Sigma_{k,r}$ is a $r\times r$ matrix such that $((\Sigma_{k,r}))_{jl}$=$\langle\Sigma_k e_j, e_l\rangle$ for $j,l=1,\ldots,r$, k=$1,2$ and $r\geq 1$. We now state the following lemma.

Lemma 3.6.3. *Fix $r\geq 1$. Suppose that Assumptions 3.2.1–3.2.3 hold. Then, under SRSWOR and LMS sampling design, $(\langle S_1, e_1\rangle,\ldots,\langle S_1, e_r\rangle)\xrightarrow{\mathcal{L}}N_r(0,\Gamma_{1,r})$ as $\nu\rightarrow\infty$ a.s. [**P**], where $\Gamma_{1,r}$ is a $r\times r$ matrix such that $((\Gamma_{1,r}))_{jl}$=$\langle\Gamma_1 e_j, e_l\rangle$ for $j,l=1,\ldots,r$, and Γ_1 is as in the statement of Lemma 3.6.2. Further, if Assumption 3.2.1 holds with $0\leq\lambda<E_{\mathbf{P}}(X_i)/b$, and Assumptions 3.2.2 and 3.2.3 hold, then, the above result holds under any HEπPS sampling design. Moreover, if Assumptions 3.2.1–3.2.3 hold, then $(\langle S_2, e_1\rangle,\ldots,\langle S_2, e_r\rangle)\xrightarrow{\mathcal{L}}N_r(0,\Gamma_{2,r})$ as $\nu\rightarrow\infty$ under RHC sampling design a.s. [**P**]. Here, $\Gamma_{2,r}$ is a $r\times r$ matrix such that $((\Gamma_{2,r}))_{jl}$=$\langle\Gamma_2 e_j, e_l\rangle$ for $j,l=1,\ldots,r$, and Γ_2 is as in the statement of Lemma 3.6.2.*

Proof. Note that $(\langle S_1, e_1\rangle,\ldots,\langle S_1, e_r\rangle)$=$\sqrt{n}(\hat{\overline{\mathbf{W}}}_1-\overline{\mathbf{W}})$. Let us first consider SRSWOR, LMS and any HEπPS sampling designs. Note that under the above-mentioned sampling designs, $\Sigma_{1,r}\rightarrow\Gamma_{1,r}$ as $\nu\rightarrow\infty$ *a.s.* [**P**] because $\Sigma_1\rightarrow\Gamma_1$ under these sampling designs as $\nu\rightarrow\infty$ *a.s.*

[**P**] in view of Lemma 3.6.2. Moreover, $\Gamma_{1,r}$ is a n.n.d. matrix since Σ_1 is a n.n.d. operator. Now, consider the case, when $\Gamma_{1,r}$ is p.d. Then, under the above sampling designs, $\mathbf{m}\Gamma_{1,r}\mathbf{m}^T > 0$ for any $\mathbf{m} \in \mathbb{R}^r$ and $\mathbf{m} \neq 0$, and all sufficiently large ν *a.s.* [**P**]. It can be shown that $\sqrt{n}\mathbf{m}(\hat{\overline{\mathbf{W}}}_1 - \overline{\mathbf{W}})^T \xrightarrow{\mathcal{L}} N(0, \mathbf{m}\Gamma_{1,r}\mathbf{m}^T)$ as $\nu \to \infty$ for any $\mathbf{m} \neq 0$ under these sampling designs *a.s.* [**P**] in the same way as $\sqrt{n}\mathbf{m}_1(\hat{\overline{\mathbf{V}}}_1 - \overline{\mathbf{V}})^T \xrightarrow{\mathcal{L}} N(0, \mathbf{m}_1\Gamma_1\mathbf{m}_1^T)$ as $\nu \to \infty$ under each of the above sampling designs for any $\mathbf{m}_1 \in \mathbb{R}^p$, $\mathbf{m}_1 \neq 0$ and Γ_1=$\lim_{\nu\to\infty}\Sigma_1$ in the proof of Lemma 2.7.2 in Section 2.7 of Chapter 2. This implies that under these sampling designs, $\sqrt{n}(\hat{\overline{\mathbf{W}}}_1 - \overline{\mathbf{W}}) \xrightarrow{\mathcal{L}} N_r(0, \Gamma_{1,r})$ as $\nu \to \infty$ *a.s.* [**P**].

Next, consider the case, when $\Gamma_{1,r}$ is a positive semi definite (p.s.d.) matrix. Let A_1=$\{\mathbf{m} \neq 0 : \mathbf{m}\Gamma_{1,r}\mathbf{m}^T > 0\}$ and A_2=$\{\mathbf{m} \neq 0 : \mathbf{m}\Gamma_{1,r}\mathbf{m}^T = 0\}$. Then, under the sampling designs mentioned in the preceding paragraph, $\sqrt{n}\mathbf{m}(\hat{\overline{\mathbf{W}}}_1 - \overline{\mathbf{W}})^T \xrightarrow{\mathcal{L}} N(0, \mathbf{m}\Gamma_{1,r}\mathbf{m}^T)$ as $\nu \to \infty$ for any $\mathbf{m} \in A_1$ *a.s.* [**P**] in the same way as argued above. Next, suppose that $P(s,\omega)$ denotes one of these sampling designs, and $Q(s,\omega)$ is a rejective sampling design with inclusion probabilities equal to those of $P(s,\omega)$ (cf. [4]). Note that under $Q(s,\omega)$, $var(\sqrt{n}\mathbf{m}(\hat{\overline{\mathbf{W}}}_1 - \overline{\mathbf{W}})^T)$=$\mathbf{m}\Sigma_{1,r}\mathbf{m}^T(1+h)$ (see Theorem 6.1 in [40]) for any ω and $\mathbf{m}$, where $h \to 0$ as $\nu \to \infty$ if $\sum_{i=1}^N \pi_i(1-\pi_i) \to \infty$ as $\nu \to \infty$. Also, note that $\sum_{i=1}^N \pi_i(1-\pi_i) \to \infty$ as $\nu \to \infty$ under $P(s,\omega)$ *a.s.* [**P**] because (3.6.1) in Lemma 3.6.1 holds under $P(s,\omega)$. Therefore, $\sum_{i=1}^N \pi_i(1-\pi_i) \to \infty$ as $\nu \to \infty$ under $Q(s,\omega)$ *a.s.* [**P**]. Next, note that $\Sigma_{1,r}$ depends on the sampling design only through the inclusion probabilities, and $\Sigma_{1,r} \to \Gamma_{1,r}$ as $\nu \to \infty$ under $P(s,\omega)$ *a.s.* [**P**] as mentioned in the previous paragraph. Therefore, $\mathbf{m}\Sigma_{1,r}\mathbf{m}^T \to 0$ as $\nu \to \infty$ for any $\mathbf{m} \in A_2$ under $Q(s,\omega)$ *a.s.* [**P**]. Hence, $\sqrt{n}\mathbf{m}(\hat{\overline{\mathbf{W}}}_1 - \overline{\mathbf{W}})^T$=$o_p(1)$ as $\nu \to \infty$ for any $\mathbf{m} \in A_2$ under $Q(s,\omega)$ *a.s.* [**P**]. Now, it follows from Lemmas 2 and 3 in [4] that

$$\begin{aligned}\sum_{s\in A} P(s,\omega) \le \sum_{s\in A} Q(s,\omega) + \sum_{s\in\mathcal{S}} |P(s,\omega) - Q(s,\omega)| \le \sum_{s\in A} Q(s,\omega) + \\ (2D(P||Q))^{1/2} \le \sum_{s\in A} Q(s,\omega) + (2D(P||R))^{1/2},\end{aligned} \tag{3.6.13}$$

where A=$\{s \in \mathcal{S} : |\sqrt{n}\mathbf{m}(\hat{\overline{\mathbf{W}}}_1 - \overline{\mathbf{W}})^T| > \epsilon\}$ for $\epsilon > 0$, and $R(s,\omega)$ is any other rejective sampling design. Since $P(s,\omega)$ is a high entropy sampling design as discussed earlier in this proof, there exists a rejective sampling design $R(s,\omega)$ such that $D(P||R) \to 0$ as $\nu \to \infty$ *a.s.* [**P**]. Then, under $P(s,\omega)$, $\sqrt{n}\mathbf{m}(\hat{\overline{\mathbf{W}}}_1 - \overline{\mathbf{W}})^T$=$o_p(1)$ as $\nu \to \infty$ for any $\mathbf{m} \in A_2$ *a.s.* [**P**]. Therefore, under $P(s,\omega)$, as $\nu \to \infty$, $\sqrt{n}\mathbf{m}(\hat{\overline{\mathbf{W}}}_1 - \overline{\mathbf{W}})^T \xrightarrow{\mathcal{L}} N(0, \mathbf{m}\Gamma_{1,r}\mathbf{m}^T)$ for any $m \neq 0$, and hence $\sqrt{n}(\hat{\overline{\mathbf{W}}}_1 - \overline{\mathbf{W}}) \xrightarrow{\mathcal{L}} N_r(0, \Gamma_{1,r})$ *a.s.* [**P**].

Next, note that $(\langle S_2, e_1\rangle, \ldots, \langle S_2, e_r\rangle)$=$\sqrt{n}(\hat{\overline{\mathbf{W}}}_2 - \overline{\mathbf{W}})$. Also, note that $\Sigma_{2,r} \rightarrow \Gamma_{2,r}$ as $\nu \rightarrow \infty$ *a.s.* [**P**] since $\Sigma_2 \rightarrow \Gamma_2$ as $\nu \rightarrow \infty$ *a.s.* [**P**] in view of Lemma 3.6.2. Moreover, $\Gamma_{2,r}$ is a n.n.d. matrix since Σ_2 is a n.n.d. operator. Let us consider the case, when $\Gamma_{2,r}$ is p.d. Then, $m\Gamma_{2,r}m^T > 0$ for any $m \neq 0$ and all sufficiently large ν *a.s.* [**P**]. It can be shown that under RHC sampling design, $\sqrt{n}\mathbf{m}(\hat{\overline{\mathbf{W}}}_2 - \overline{\mathbf{W}})^T \xrightarrow{\mathcal{L}} N(0, \mathbf{m}\Gamma_{2,r}\mathbf{m}^T)$ as $\nu \rightarrow \infty$ for any $\mathbf{m} \neq 0$ *a.s.* [**P**] in the same way as $\sqrt{n}\mathbf{m}_1(\hat{\overline{\mathbf{V}}}_2 - \overline{\mathbf{V}})^T \xrightarrow{\mathcal{L}} N(0, \mathbf{m}_1\Gamma_2\mathbf{m}_1^T)$ as $\nu \rightarrow \infty$ under RHC sampling design for any $\mathbf{m}_1 \in \mathbb{R}^p$, $\mathbf{m}_1 \neq 0$ and Γ_2=$\lim_{\nu\rightarrow\infty} \Sigma_2$ in the proof of Lemma 2.7.2 in Section 2.7 of Chapter 2. Therefore, under RHC sampling design, $\sqrt{n}(\hat{\overline{\mathbf{W}}}_2 - \overline{\mathbf{W}}) \xrightarrow{\mathcal{L}} N_r(0, \Gamma_{2,r})$ as $\nu \rightarrow \infty$ *a.s.* [**P**].

Next, consider the case, when $\Gamma_{2,r}$ is p.s.d. Let A_1=$\{\mathbf{m} \neq 0 : \mathbf{m}\Gamma_{2,r}\mathbf{m}^T > 0\}$ and A_2=$\{\mathbf{m} \neq 0 : \mathbf{m}\Gamma_{2,r}\mathbf{m}^T = 0\}$. Then, under RHC sampling design, $\sqrt{n}\mathbf{m}(\hat{\overline{\mathbf{W}}}_2 - \overline{\mathbf{W}})^T \xrightarrow{\mathcal{L}} N(0, \mathbf{m}\Gamma_{2,r}\mathbf{m}^T)$ as $\nu \rightarrow \infty$ for any $\mathbf{m} \in A_1$ *a.s.* [**P**] in the same way as above. Under RHC sampling design, $var(\sqrt{n}\mathbf{m}(\hat{\overline{\mathbf{W}}}_2 - \overline{\mathbf{W}})^T)$=$\mathbf{m}\Sigma_{2,r}\mathbf{m}^T$ (see [61]) for any ω and $\mathbf{m}$. Note that $\mathbf{m}\Sigma_{2,r}\mathbf{m}^T \rightarrow 0$ as $\nu \rightarrow \infty$ for any $\mathbf{m} \in A_2$ *a.s.* [**P**]. Then, under RHC sampling design, $\sqrt{n}\mathbf{m}(\hat{\overline{\mathbf{W}}}_2 - \overline{\mathbf{W}})^T$=$o_p(1)$ as $\nu \rightarrow \infty$ for any $\mathbf{m} \in A_2$ *a.s.* [**P**]. Therefore, under RHC sampling design, as $\nu \rightarrow \infty$, $\sqrt{n}\mathbf{m}(\hat{\overline{\mathbf{W}}}_2 - \overline{\mathbf{W}})^T \xrightarrow{\mathcal{L}} N(0, \mathbf{m}\Gamma_{2,r}\mathbf{m}^T)$ for any $\mathbf{m} \neq 0$, and hence $\sqrt{n}(\hat{\overline{\mathbf{W}}}_2 - \overline{\mathbf{W}}) \xrightarrow{\mathcal{L}} N_r(0, \Gamma_{2,r})$ *a.s.* [**P**]. □

Recall from the proof of Proposition 3.2.1 in Section 3.5 that Π_r denotes the orthogonal projection onto the linear span of $\{e_1, \ldots, e_r\}$, i.e., $\Pi_r(a)$=$\sum_{j=1}^{r}\langle a, e_j\rangle e_j$ for any $r \geq 1$ and $a \in \mathcal{H}$. Further, suppose that $B_{1,r}$=$\{s \in \mathcal{S} : ||S_1 - \Pi_r(S_1)||_{\mathcal{H}} > \epsilon\}$ and $B_{2,r}$=$\{s \in \mathcal{S} : ||S_2 - \Pi_r(S_2)||_{\mathcal{H}} > \epsilon\}$ for any $\epsilon > 0$. Now, we state the following lemma.

Lemma 3.6.4. *Suppose that Assumptions 3.2.1–3.2.3 hold, and $P(s, \omega)$ denotes one of SRSWOR and LMS sampling design. Then, for any $\epsilon > 0$, $\lim_{r\rightarrow\infty} \overline{\lim}_{\nu\rightarrow\infty} \sum_{s\in B_{1,r}} P(s, \omega)$=0 a.s. [**P**]. Further, if Assumption 3.2.1 holds with $0 \leq \lambda < E_{\mathbf{P}}(X_i)/b$, and Assumptions 3.2.2 and 3.2.3 hold, then the above result holds under any HEπPS sampling design. Moreover, suppose that Assumptions 3.2.1–3.2.4 hold, and $P(s, \omega)$ denotes RHC sampling design. Then, for any $\epsilon > 0$, $\lim_{r\rightarrow\infty} \overline{\lim}_{\nu\rightarrow\infty} \sum_{s\in B_{2,r}} P(s, \omega)$=0 a.s. [**P**].*

Proof. Let us first consider the case, when $P(s, \omega)$ is one of SRSWOR, LMS and any HEπPS sampling designs. Suppose that $Q(s, \omega)$ is as described in the 2^{nd} paragraph of the proof of Lemma 3.6.3. Then, following similar arguments as in the proof of Theorem 6.1 in [40], we can show that

$$E\langle S_1, e_j\rangle^2 = (nN^{-2})\sum_{i=1}^{N}\langle V_i - T_V\pi_i, e_j\rangle^2(\pi_i^{-1}-1)(1+h) = \langle\Sigma_1 e_j, e_j\rangle(1+h) \quad (3.6.14)$$

under $Q(s,\omega)$ for any ω and $j \geq 1$. Here, h does not depend on $\{e_j\}_{j=1}^{\infty}$, and $h \to 0$ as $\nu \to \infty$ whenever $\sum_{i=1}^{N}\pi_i(1-\pi_i) \to \infty$ as $\nu \to \infty$. Recall from the 2^{nd} paragraph in the proof of Lemma 3.6.3 that $\sum_{i=1}^{N}\pi_i(1-\pi_i) \to \infty$ as $\nu \to \infty$ under $Q(s,\omega)$ *a.s.* [**P**]. It follows from Lemma 3.6.2 that under $P(s,\omega)$, $\Sigma_1 \to \Gamma_1$ with respect to the HS norm and $\sum_{j=1}^{\infty}\langle\Sigma_1 e_j, e_j\rangle \to \sum_{j=1}^{\infty}\langle\Gamma_1 e_j, e_j\rangle$ as $\nu \to \infty$ *a.s.* [**P**]. Therefore, $\Sigma_1 \to \Gamma_1$ and $\sum_{j=1}^{\infty}\langle\Sigma_1 e_j, e_j\rangle \to \sum_{j=1}^{\infty}\langle\Gamma_1 e_j, e_j\rangle$ as $\nu \to \infty$ under $Q(s,\omega)$ *a.s.* [**P**] because Σ_1 depends on the sampling design only through inclusion probabilities, and $P(s,\omega)$ and $Q(s,\omega)$ have the same inclusion probabilities. Thus as $\nu \to \infty$, $E\langle S_1, e_j\rangle^2 \to \langle\Gamma_1 e_j, e_j\rangle$ for any $j \geq 1$, and $\sum_{j=1}^{\infty} E\langle S_1, e_j\rangle^2 \to \sum_{j=1}^{\infty}\langle\Gamma_1 e_j, e_j\rangle$ under $Q(s,\omega)$ *a.s.* [**P**]. Then, following the same line of arguments as in the proof of Theorem 1.1 in [54], we can say that

$$\overline{\lim}_{\nu\to\infty}\sum_{s\in B_{1,r}} Q(s,\omega) \leq \sum_{j=r+1}^{\infty}\langle\Gamma_1 e_j, e_j\rangle\epsilon^{-2} \quad (3.6.15)$$

a.s. [**P**] for any $r \geq 1$. Therefore, $\lim_{r\to\infty}\overline{\lim}_{\nu\to\infty}\sum_{s\in B_{1,r}} Q(s,\omega)$=0 *a.s.* [**P**]. Further, it can be shown that $\lim_{r\to\infty}\overline{\lim}_{\nu\to\infty}\sum_{s\in B_{1,r}} P(s,\omega)$=0 *a.s.* [**P**] in the same way as the result $\sqrt{nm}(\hat{\overline{\mathbf{W}}}_1 - \overline{\mathbf{W}})^T$=$o_p(1)$ as $\nu \to \infty$ under $P(s,\omega)$ *a.s.* [**P**] is shown in the 2^{nd} paragraph of the proof of Lemma 3.6.3.

Let us next consider the case, when $P(s,\omega)$ is RHC sampling design. Note that

$$E\langle S_2, e_j\rangle^2 = (n\gamma)(\overline{X}N^{-1})\sum_{i=1}^{N}\Big(\langle V_i, e_j\rangle - \langle\overline{V}/\overline{X}, e_j\rangle X_i\Big)^2 X_i^{-1} = \langle\Sigma_2 e_j, e_j\rangle \quad (3.6.16)$$

under RHC sampling design for any $j \geq 1$ and ω (cf. [61]). Also, note that as $\nu \to \infty$, $\Sigma_2 \to \Gamma_2$ with respect to the HS norm and $\sum_{j=1}^{\infty}\langle\Sigma_2 e_j, e_j\rangle \to \sum_{j=1}^{\infty}\langle\Gamma_2 e_j, e_j\rangle$ *a.s.* [**P**] in view of Lemma 3.6.2. Then, under RHC sampling design, as $\nu \to \infty$, $E\langle S_2, e_j\rangle^2 \to \langle\Gamma_2 e_j, e_j\rangle$ for any $j \geq 1$, and $\sum_{j=1}^{\infty} E\langle S_2, e_j\rangle^2 \to \sum_{j=1}^{\infty}\langle\Gamma_2 e_j, e_j\rangle$ *a.s.* [**P**]. Therefore, $\lim_{r\to\infty}\overline{\lim}_{\nu\to\infty}\sum_{s\in B_{2,r}} P(s,\omega)$=0 *a.s.* [**P**] using similar arguments as in the proof of Theorem 1.1 in [54]. □

Before we state the next lemma, let $V_i^{\sharp}$ be one of $Y_i \otimes Z_i$, $Z_i^T Z_i$, Z_i and 1 for i=1, . . . , N. Also, let $\overline{V}^{\sharp}$=$N^{-1}\sum_{i=1}^{N}V_i^{\sharp}$, $S_1^{\sharp}$=$\sqrt{n}(\sum_{i\in s}(N\pi_i)^{-1}V_i^{\sharp}-\overline{V}^{\sharp})$ and $S_2^{\sharp}$=$\sqrt{n}(\sum_{i\in s}(NX_i)^{-1}G_iV_i^{\sharp}-\overline{V}^{\sharp})$. In this case, we denote the associated norm by $||\cdot||_{\mathcal{G}}$. Note that $||\cdot||_{\mathcal{G}}$=the Euclidean norm, when $V_i^{\sharp}$ is one of $Z_i^TZ_i$, Z_i and 1, and $||\cdot||_{\mathcal{G}}$=the HS norm, when $V_i^{\sharp}$=$Y_i\otimes Z_i$.

Lemma 3.6.5. *Suppose that Assumptions 3.2.1–3.2.4 hold. Then,* $||S_1^{\sharp}||_{\mathcal{G}}$=$O_p(1)$ *under SRSWOR, LMS and any HEπPS sampling designs, and* $||S_2^{\sharp}||_{\mathcal{G}}$=$O_p(1)$ *under RHC sampling design as* $\nu\rightarrow\infty$ *a.s.* [***P***].

Proof. Note that $\{V_i^{\sharp}\}_{i=1}^N$ are elements of either an infinite dimensional separable Hilbert space or a finite dimensional Euclidean space. Let $\{e_j^{\sharp}\}$ be an orthonormal basis of that space. Further, note that $N^{-1}\sum_{i=1}^{N}||V_i^{\sharp}||_{\mathcal{G}}^2$=$O(1)$ as $\nu\rightarrow\infty$ *a.s.* [**P**] by SLLN and Assumption 3.2.3. Now, suppose that $P(s,\omega)$ is one of SRSWOR, LMS and any HEπPS sampling designs, and $Q(s,\omega)$ is the corresponding rejective sampling design as described in the 2^{nd} paragraph of the proof of Lemma 3.6.3. Then, one can show that

$$E||S_1^{\sharp}||_{\mathcal{G}}^2=E\bigg(\sum_j\Big\langle S_1^{\sharp},e_j^{\sharp}\Big\rangle^2\bigg)=(nN^{-2})\sum_j\sum_{i=1}^{N}\Big\langle V_i^{\sharp}-T^{\sharp}\pi_i,e_j^{\sharp}\Big\rangle^2\times \\ (\pi_i^{-1}-1)(1+h) \tag{3.6.17}$$

for any ω under $Q(s,\omega)$ in the same way as the derivation of $E\langle S_1,e_j\rangle^2$=$\langle\Sigma_1e_j,e_j\rangle(1+h)$ in the proof of Lemma 3.6.4. Here, $T^{\sharp}$=$\sum_{i=1}^{N}V_i^{\sharp}(1-\pi_i)\big(\sum_{i=1}^{N}\pi_i(1-\pi_i)\big)^{-1}$, h does not depend on $\{e_j^{\sharp}\}$, and $h\rightarrow 0$ as $\nu\rightarrow\infty$ if $\sum_{i=1}^{N}\pi_i(1-\pi_i)\rightarrow\infty$ as $\nu\rightarrow\infty$. Note that (3.6.1) in Lemma 3.6.1 holds under $Q(s,\omega)$ because (3.6.1) holds under $P(s,\omega)$ by Lemma 3.6.1, and $P(s,\omega)$ and $Q(s,\omega)$ have the same inclusion probabilities. Then, $\sum_{i=1}^{N}\pi_i(1-\pi_i)\rightarrow\infty$ as $\nu\rightarrow\infty$ under $Q(s,\omega)$ *a.s.* [**P**]. Therefore, as $\nu\rightarrow\infty$,

$$(nN^{-2})\sum_j\sum_{i=1}^{N}\Big\langle V_i^{\sharp}-T^{\sharp}\pi_i,e_j^{\sharp}\Big\rangle^2(\pi_i^{-1}-1)(1+h)=(nN^{-2})\times \\ \sum_{i=1}^{N}||V_i^{\sharp}-T^{\sharp}\pi_i||_{\mathcal{G}}^2(\pi_i^{-1}-1)(1+h)=(nN^{-2})\bigg[\sum_{i=1}^{N}||V_i^{\sharp}||_{\mathcal{G}}^2(\pi_i^{-1}-1)- \\ ||T^{\sharp}||_{\mathcal{G}}^2\sum_{i=1}^{N}\pi_i(1-\pi_i)\bigg](1+h)\leq(nN^{-2})\sum_{i=1}^{N}\pi_i^{-1}||V_i^{\sharp}||_{\mathcal{G}}^2(1+h)=O(1) \tag{3.6.18}$$

under $Q(s,\omega)$ *a.s.* [**P**] since $N^{-1}\sum_{i=1}^{N}||V_i^{\sharp}||_{\mathcal{G}}^2$=$O(1)$ as $\nu\rightarrow\infty$ *a.s.* [**P**]. Hence, $E||S_1^{\sharp}||_{\mathcal{G}}^2$=$O(1)$ as $\nu\rightarrow\infty$ under $Q(s,\omega)$ *a.s.* [**P**]. Thus $||S_1^{\sharp}||_{\mathcal{G}}$=$O_p(1)$ as $\nu\rightarrow\infty$ under $Q(s,\omega)$ *a.s.* [**P**]. Now, it can be shown that $||S_1^{\sharp}||_{\mathcal{G}}=O_p(1)$ as $\nu\rightarrow\infty$ under $P(s,\omega)$ *a.s.* [**P**] in the same way as the result $\sqrt{n}m(\hat{\overline{\mathbf{W}}}_1 - \overline{\mathbf{W}})^T = o_p(1)$ as $\nu \to \infty$ under $P(s,\omega)$ *a.s.* [**P**] is shown in the 2^{nd} paragraph of the proof of Lemma 3.6.3.

Next, note that under RHC sampling design, as $\nu \to \infty$,

$$E||S_2^\sharp||_{\mathcal{G}}^2 = E\bigg(\sum_j \Big\langle S_2^\sharp, e_j^\sharp \Big\rangle^2\bigg) = (n\gamma)(\overline{X}N^{-1})\sum_j \sum_{i=1}^N \bigg(\langle V_i^\sharp, e_j^\sharp\rangle - \langle \overline{V}^\sharp/\overline{X}, e_j^\sharp\rangle X_i\bigg)^2 X_i^{-1} \leq (n\gamma)\sum_j N^{-1}\sum_{i=1}^N \langle V_i^\sharp, e_j^\sharp\rangle^2 \overline{X} X_i^{-1} \leq (n\gamma)N^{-1}\sum_{i=1}^N ||V_i^\sharp||_{\mathcal{G}}^2 \overline{X} X_i^{-1} = O(1) \tag{3.6.19}$$

a.s. [**P**] because $N^{-1}\sum_{i=1}^N ||V_i^\sharp||_{\mathcal{G}}^2 = O(1)$ as $\nu \to \infty$ *a.s.* [**P**], and Assumption 3.2.2 holds. Also, note that $n\gamma = O(1)$ as $\nu \to \infty$ since Assumption 3.2.4 holds. Therefore, $||S_2^\sharp||_{\mathcal{G}} = O_p(1)$ as $\nu \to \infty$ under RHC sampling design *a.s.* [**P**]. □

Chapter 4

Quantile processes and their applications in finite populations

The estimation of the finite population median instead of the population mean is meaningful, when the population observations are generated from skewed and heavy-tailed distributions. The estimation of the population trimmed means, which are constructed based on the population quantile function, can also be considered for a similar reason. [18], [35], [52], [53], [67], [85], etc. considered the estimation of the population median, whereas [77] considered the estimation of the population trimmed means. The estimation of some specific population quantiles (eg., population quartiles) are also of interest because estimators of population parameters like interquartile range, quantile based measure of skewness (Bowley's measure of skewness), etc. can be constructed based on the estimators of the population quantiles. [35] considered the estimation of the interquartile range, whereas [77] considered the estimation of the Bowley's measure of skewness and several other functions of the population quantiles. The median and the trimmed means in the population are more robust and resistant to outliers than the population mean. Several problems due to outliers in sample survey were discussed in detail in [3], [34], [47] and references therein.

Weak convergence of quantiles and qauntile processes were studied in classical set up, when sample observations are i.i.d. random variables from a probability distribution (see [76], [79], etc.). It becomes challenging, when we deal with samples drawn from a finite population using a without replacement sampling design. In this case, we face difficulty as sample observations may neither be independent nor identical. It becomes more challenging, when we consider the quantile processes constructed based on estimators other than the sample quantile, namely the ratio, the difference and the regression estimators of the population quantile. Furthermore, different quantile processes are considered under different sampling designs unlike in the case of i.i.d. sample observations.

The weak convergence of several empirical processes were shown in the earlier literature (see [7], [43] and references therein) under some conditions on sampling designs. These conditions seem to hold under only SRSWOR, Poisson sampling design and rejective sampling design. There is no result available in the literature related to the weak convergence of empirical processes under LMS, πPS, RHC and stratified multistage cluster sampling designs. These sampling designs, especially stratified multistage cluster sampling designs, are of practical importance in sample surveys. In this chapter, we show the weak convergence of an empirical process similar to the Hájek empirical process considered in [7] and [43] under high entropy sampling designs, which include SRSWOR, LMS and HEπPS sampling designs. We also show the weak convergence of the above empirical process under RHC and stratified multistage cluster sampling designs.

Asymptotic results related to the weak convergence of empirical processes were applied to study the asymptotic behaviour of poverty rate (see [7]) and to deal with different regression and classification problems (see [43]). However, neither [7] nor [43] considered quantiles and quantile processes in the context of sample survey. [78] proved strong and weak versions of Bahadur type representations for the sample quantile process under simple random sampling in the presence of superpopulation model. [26] constructed a quantile process based on the sample quantile, which is obtained by inverting the Hájek estimator of finite population distribution function under high entropy sampling designs. There is no available result related to the weak convergence of quantile processes based on well-known quantile estimators like the ratio (see [67]), the difference (see [67]), and the regression (see [27] and [70]) estimators, which are constructed using an auxiliary information. There is also no result available in the literature related to the weak convergence of a quantile process under RHC and stratified multistage cluster sampling designs. In this chapter, we establish the weak convergence of the quantile processes, which are constructed based on the sample quantile as well as the ratio, the difference and the regression estimators of the finite population quantile, under the aforementioned sampling designs using the weak convergence of empirical process, Hadamard differentiability of the quantile map and the functional delta method. The weak convergence of the empirical and the quantile processes are shown under a probability distribution, which is generated by a sampling design and a superpopulation model jointly.

In this chapter, we apply asymptotic results for quantile processes to derive asymptotic distributions of the smooth L-estimators (see [77]) and the estimators of smooth functions of population quantiles. We estimate asymptotic variances of these estimators consistently. Confidence intervals for finite population parameters like the median, the α-trimmed means, the interquartile range and the quantile based measure of skewness are constructed based on asymptotic distributions of these estimators.

We also compare several estimators based on their asymptotic distributions. It is shown that the use of the auxiliary information in the estimation stage may have an adverse effect on the performances of the smooth L-estimators and the estimators of smooth functions of population quantiles based on the ratio, the difference and the regression estimators under each of SRSWOR, LMS, HEπPS and RHC sampling designs. Moreover, each of the aforementioned estimators may have worse performance under HEπPS and RHC sampling designs, which use the auxiliary information, than under SRSWOR. In practice, SRSWOR is easier to implement than the sampling designs that use the auxiliary information. Thus the above result is significant in view of selecting the appropriate sampling design.

In this chapter, it is further shown that the sample median is more efficient than the sample mean under SRSWOR, whenever the finite population observations are generated from some symmetric and heavy-tailed superpopulation distributions with the same superpopulation mean and median. A similar result is known to hold in the classical set up with i.i.d. sample observations. However, for the cases of symmetric superpopulation distributions with the same superpopulation mean and median, it is shown that the GREG estimator of the finite population mean is more efficient than the sample median under SRSWOR, whenever there is substantial correlation present between the study and the auxiliary variables. This stands in contrast to what happens in the case of i.i.d. observations.

In Section 4.1, we give the expressions of the sample quantile and the ratio, the difference and the regression estimators of the population quantile. In this section, we also construct several quantile processes based on these estimators. We present asymptotic results related to the weak convergence of empirical and quantile processes in Sections 4.2 and 4.3 for single stage and stratified multistage cluster sampling designs. Asymptotic results related to the smooth L-estimators and the estimators of smooth functions of population quantiles are presented in Section 4.4. In Section 4.5, we compare different estimators. Some numerical results based on real data are presented in Section 4.6. Proofs of several results are given in Sections 4.7 and 4.8.

4.1. Quantile processes based on different estimators

We recall from the introduction that (Y_i, X_i) denotes the value of (y, x) for the i^{th} population unit, i=$1, \ldots, N$, where y is a finite/infinite dimensional study variable, and x is a positive real-valued size variable. In this chapter, we assume that y is a real-valued study variable. As in Chapter 2, here also we assume that the covariate z and the size variable x are the same. Recall from the introduction that the population values $\{X_i\}_{i=1}^{N}$ on x are assumed to be known and utilized to implement sampling designs as well as to construct estimators. Let $F_{y,N}(t)$=$\sum_{i=1}^{N} \mathbb{1}_{[Y_i \leq t]}/N$ be the finite population distribution function of y, where $t \in \mathbb{R}$. Then, the finite population p^{th} quantile of y is defined as $Q_{y,N}(p)$=$\inf\{t \in \mathbb{R} : F_{y,N}(t) \geq p\}$, where $0 < p < 1$. The HT estimator $\sum_{i\in s}(N\pi_i)^{-1}\mathbb{1}_{[Y_i \leq t]}$ (cf. 2^{nd} row in Table 2.1 in Chapter 2) and the RHC estimator $\sum_{i\in s}(NX_i)^{-1}G_i\mathbb{1}_{[Y_i \leq t]}$ (cf. 3^{rd} row in Table 2.1 in Chapter 2) are well-known design unbiased estimators of $F_{y,N}(t)$. Here, π_i is the inclusion probability of the i^{th} population unit under any sampling design $P(s)$, and G_i is the x total of that group of population units formed in the first step of the RHC sampling design from which the i^{th} population unit is selected in the sample (see the beginning of Section 2.1 in Chapter 2). A unified way of writing these estimators is $\sum_{i\in s} d(i, s)\mathbb{1}_{[Y_i \leq t]}$. An estimator of $Q_{y,N}(p)$ can be constructed as $\inf\{t \in \mathbb{R} : \sum_{i\in s} d(i, s)\mathbb{1}_{[Y_i \leq t]} \geq p\}$ (see [52]). However, $\inf\{t \in \mathbb{R} : \sum_{i\in s} d(i, s)\mathbb{1}_{[Y_i \leq t]} \geq p\}$ is not well defined, when $\max_{t\in\mathbb{R}} \sum_{i\in s} d(i, s)\mathbb{1}_{[Y_i \leq t]}$=$\sum_{i\in s} d(i, s) < p$ for some $0 < p < 1$ and $s \in \mathcal{S}$. On the other hand, $\sum_{i\in s} d(i, s)\mathbb{1}_{[Y_i \leq t]}$ violates the properties of the distribution functions, when $\max_{t\in\mathbb{R}} \sum_{i\in s} d(i, s)\mathbb{1}_{[Y_i \leq t]} > 1$ for some $s \in \mathcal{S}$. To eliminate these problems, we consider $\hat{F}_y(t)$=$\sum_{i\in s} d(i, s)\mathbb{1}_{[Y_i \leq t]}/\sum_{i\in s} d(i, s)$ (see [26], [52] and [85]) as an estimator of $F_{y,N}(t)$. $\hat{F}_y(t)$ becomes the Hájek estimator of $F_{y,N}(t)$ for $d(i, s)$=$(N\pi_i)^{-1}$ (see [41]). Based on $\hat{F}_y(t)$, the sample p^{th} quantile of y is defined as

$$\hat{Q}_y(p) = \inf\{t \in \mathbb{R} : \hat{F}_y(t) \geq p\}. \tag{4.1.1}$$

Note that $\hat{F}_y(t)$ satisfies all the properties of a distribution function, and $\max_{t\in\mathbb{R}} \hat{F}_y(t)$=$1 > p$ for any $0 < p < 1$ and $s \in \mathcal{S}$. Thus $\hat{Q}_y(p)$ is a well defined estimator of $Q_{y,N}(p)$. The estimator $\hat{Q}_y(p)$ was considered for $d(i, s)$=$(N\pi_i)^{-1}$ in [26], [85], etc. We also consider $\hat{Q}_y(p)$ for $d(i, s)$=$(NX_i)^{-1}G_i$ under RHC sampling design. Further, we consider some estimators of $Q_{y,N}(p)$, which are constructed using the auxiliary variable x in the estimation stage. Suppose that $Q_{x,N}(p)$ and $\hat{Q}_x(p)$ are the population and the sample p^{th} quantiles of x, respectively. Then, the ratio, the difference and the regression estimators of $Q_{y,N}(p)$ are defined as

$$\hat{Q}_{y,RA}(p) = \big(\hat{Q}_y(p)/\hat{Q}_x(p)\big)Q_{x,N}(p),$$
$$\hat{Q}_{y,DI}(p) = \hat{Q}_y(p) + \left(\sum_{i\in s} d(i,s)Y_i \Big/ \sum_{i\in s} d(i,s)X_i\right)(Q_{x,N}(p) - \hat{Q}_x(p)) \text{ and} \tag{4.1.2}$$
$$\hat{Q}_{y,REG}(p) = \hat{Q}_y(p) + \hat{\beta}(Q_{x,N}(p) - \hat{Q}_x(p)),$$

respectively, where $\hat{\beta}$=$\sum_{i\in s} d(i,s)X_iY_i/\sum_{i\in s} d(i,s)X_i^2$ is the estimator of finite population regression coefficient of y on x through the origin. The estimators $\hat{Q}_{y,RA}(p)$ and $\hat{Q}_{y,DI}(p)$ were considered in [67] for $d(i,s)$=$(N\pi_i)^{-1}$, whereas $\hat{Q}_{y,REG}(p)$ was considered in [27] and [70]. for $d(i,s)$=$(N\pi_i)^{-1}$. We also consider these estimators for $d(i,s)$=$(NX_i)^{-1}G_i$ under RHC sampling design.

Now, suppose that for any $0<\alpha<\beta<1$, $D[\alpha,\beta]$ is the space of all left continuous functions on $[\alpha,\beta]$ having right hand limits at each point, and $\mathcal{D}$ is the σ-field on $D[\alpha,\beta]$ generated by the open balls (ball σ-field) with respect to the sup norm metric. Note that $\mathcal{D}$ coincides with the Borel σ-field on $D[\alpha,\beta]$ with respect to the Skorohod metric (cf. [6] and [79]). Thus the quantile processes $\{\sqrt{n}(G(p)-Q_{y,N}(p)) : p\in[\alpha,\beta]\}$ for $G(p)$=$\hat{Q}_y(p)$, $\hat{Q}_{y,DI}(p)$, $\hat{Q}_{y,RA}(p)$ and $\hat{Q}_{y,REG}(p)$ are random functions in $(D[\alpha,\beta],\mathcal{D})$. Following the notion of weak convergence in [6] and [79], we shall show that the above quantile processes converge weakly in $(D[\alpha,\beta],\mathcal{D})$ with respect to the sup norm metric (see Sections 4.2 and 4.3). The weak convergence in $(D[\alpha,\beta],\mathcal{D})$ with respect to the sup norm metric implies and is implied by the weak convergence in $(D[\alpha,\beta],\mathcal{D})$ with respect to the Skorohod metric given that the limiting process has almost sure continuous paths.

4.2. Weak convergence of quantile processes under single stage sampling designs

As in the earlier chapters, we first consider a superpopulation model such that $\{(Y_i,X_i) : 1\leq i\leq N\}$ are i.i.d. random vectors on some probability space $(\Omega,\mathcal{F},\mathbf{P})$. Also, as in Section 2.2 of Chapter 2 and Section 3.1 of Chapter 3, we consider the function $P(s,\omega)$ that is defined on $\mathcal{S}\times\Omega$. Recall from these sections that for each $s\in\mathcal{S}$, $P(s,\omega)$ is a random variable on Ω, and for each $\omega\in\Omega$, $P(s,\omega)$ is a probability distribution on $\mathcal{S}$. It is to be noted that $P(s,\omega)$ is a sampling design for each $\omega\in\Omega$. Suppose that $\mathcal{A}$ is the power set of S. Then, we consider the probability measure $\mathbf{P}^*(B \times E)$=$\int_E \sum_{s\in B} P(s,\omega)d\mathbf{P}(\omega)$ (see [7] and [43]) defined on the product space ($\mathcal{S} \times \Omega$, $\mathcal{A} \times \mathcal{F}$), where $B \in \mathcal{A}$, $E \in \mathcal{F}$ and $B \times E$ is a cylinder subset of $\mathcal{S} \times \Omega$. Recall from Section 2.2 of Chapter 2 and Section 3.1 of Chapter 3 that we denote expectations of random quantities with respect to $\mathbf{P}$ by $E_{\mathbf{P}}$. We also denote expectations of random quantities with respect to $P(s,\omega)$ and $\mathbf{P}^*$ by E and $E_{\mathbf{P}^*}$, respectively. Also, recall from these sections that we define our asymptotic framework as follows. Let $\{\mathcal{P}_\nu\}$ be a sequence of populations with $N_\nu, n_\nu \to \infty$ as $\nu \to \infty$, where N_ν and n_ν are, respectively, the population and sample sizes corresponding to the ν^{th} population. We suppress the limiting index ν for the sake of notational simplicity.

We shall first show the weak convergence of the quantile processes introduced in Section 4.1 under high entropy sampling designs. Recall from Section 3.2 in Chapter 3 (see also the introduction) that a sampling design $P(s,\omega)$ is called the high entropy sampling design if

$$D(P||R) = \sum_{s\in\mathcal{S}} P(s,\omega)\log(P(s,\omega)/R(s,\omega)) \to 0 \text{ as } \nu \to \infty \ a.s. \ [\mathbf{P}] \tag{4.2.1}$$

for some rejective sampling design $R(s,\omega)$ (for the description of the rejective sampling design, see the introduction). Some examples of high entropy sampling designs are SRSWOR, RS sampling design (see [4] and the introduction), LMS sampling design (see Lemma 3.6.1 in Section 3.6 of Chapter 3), etc.

Suppose that F_y and F_x are superpopulation distribution functions of y and x, respectively. Further, suppose that $Q_y(p)$=inf$\{t \in \mathbb{R} : F_y(t) \geq p\}$ and $Q_x(p)$=inf$\{t \in \mathbb{R} : F_x(t) \geq p\}$ are superpopulation p^{th} quantiles of y and x, respectively, and $\mathbf{V}_i$=$\mathbf{R}_i - \sum_{i=1}^N \mathbf{R}_i/N$ for i=$1,\ldots,N$, where

$$\mathbf{R}_i = (\mathbb{1}_{[Y_i\leq Q_y(p_1)]},\ldots,\mathbb{1}_{[Y_i\leq Q_y(p_k)]},\mathbb{1}_{[X_i\leq Q_x(p_1)]}\ldots,\mathbb{1}_{[X_i\leq Q_x(p_k)]})$$

for $p_1,\ldots,p_k \in (0,1)$ and $k \geq 1$. Moreover, let $\mathbf{T}_V$=$\sum_{i=1}^N \mathbf{V}_i(1-\pi_i)/\sum_{i=1}^N \pi_i(1-\pi_i)$, where $\{\pi_i\}_{i=1}^N$ denote inclusion probabilities. Recall from earlier chapters that all vectors in Euclidean spaces are taken as row vectors and superscript T is used to denote their transpose. Before, we state the main result, let us consider the following assumptions.

Assumption 4.2.1. $n/N \to \lambda$ *as* $\nu \to \infty$, *where* $0 < \lambda < 1$.

Assumption 4.2.2. *The inclusion probabilities* $\{\pi_i\}_{i=1}^N$ *are such that the following hold.*

(i) *Given any* $k \geq 1$ *and* $p_1, \ldots, p_k \in (0,1)$, $(n/N^2)\sum_{i=1}^{N}(\boldsymbol{V}_i - \boldsymbol{T}_V\pi_i)^T(\boldsymbol{V}_i - \boldsymbol{T}_V\pi_i)(\pi_i^{-1} - 1) \to \Gamma$ *as* $\nu \to \infty$ *a.s.* $[\boldsymbol{P}]$ *for some positive definite (p.d.) matrix* Γ.

(ii) *There exist constants* $K_1, K_2 > 0$ *such that for all* i=$1, \ldots, N$ *and* $\nu \geq 1$, $K_1 \leq N\pi_i/n \leq K_2$ *a.s.* $[\boldsymbol{P}]$.

Suppose that $supp(F)$=(a_1, a_2) is the support (see [26]) of any distribution function F, where a_1=sup$\{t \in \mathbb{R} : F(t) = 0\}$ and a_2=inf$\{t \in \mathbb{R} : F(t) = 1\}$. Note that $-\infty \leq a_1 < a_2 \leq \infty$. Then, we consider the following assumption on superpopulation distributions of y and x.

Assumption 4.2.3. *Superpopulation distribution functions* F_y *of* y *and* F_x *of* x *are continuous and are differentiable with positive continuous derivatives* f_y *and* f_x *on* $supp(F_y) \subseteq (-\infty, \infty)$ *and* $supp(F_x) \subseteq (0, \infty)$, *respectively.*

Similar assumptions like Assumptions 4.2.1 and 4.2.2–*(i)* are stated and discussed in Chapter 2 (see the discussion related to Assumptions 2.1.1 and 2.1.4 in Section 2.1 of Chapter 2). It can be shown using SLLN that Assumption 4.2.2-*(i)* holds under SRSWOR, LMS and any πPS sampling designs (see Lemma 4.8.10 in Section 4.8). It can also be shown that Assumption 4.2.2-*(ii)* holds under the aforementioned sampling designs (see Lemma 3.6.1 in Chapter 3). Assumption 4.2.2-*(ii)* was considered earlier in sample survey literature (see (C1) in [7] and Assumption 2–*(i)* in [85]). Assumption 4.2.3 was considered before by [26] (see A2 in [26]). Assumptions 4.2.1 and 4.2.2 are required to show the finite dimensional convergence of the empirical process $\{\sqrt{n}(\hat{F}_u(t) - t) : t \in [0,1]\}$ for $d(i,s)$=$(N\pi_i)^{-1}$ under high entropy sampling designs, where

$$\hat{F}_u(t) = \sum_{i \in s} d(i,s)\mathbb{1}_{[U_i \leq t]} \Big/ \sum_{i \in s} d(i,s) \text{ and } U_i = F_y(Y_i) \tag{4.2.2}$$

for i=$1, \ldots, N$ and $0 \leq t \leq 1$. Here, F_y is as in the paragraph preceding Assumption 4.2.1. On the other hand, Assumptions 4.2.1, 4.2.2-*(ii)* and 4.2.3 are used to establish the tightness of this empirical process under the same sampling designs. Based on the weak convergence of this empirical process, we shall prove the weak convergence of the aforementioned quantile processes under high entropy sampling designs. Suppose that $\tilde{D}[0,1]$ is the class of all right continuous functions defined on $[0,1]$ with finite left limits, and $\tilde{\mathcal{D}}$ is the σ-field on $\tilde{D}[0,1]$ generated by the open balls (ball σ-field) with respect to the sup norm metric. Then, we state the following proposition.

Proposition 4.2.1. *Suppose that Assumptions 4.2.1 and 4.2.3 hold. Then, under* $\boldsymbol{P}^*$*,*

$$\left\{\sqrt{n}(\hat{F}_u(t)-t) : t \in [0,1]\right\} \xrightarrow{\mathcal{L}} \mathbb{H} \text{ as } \nu \to \infty$$

in $(\tilde{D}[0,1], \tilde{\mathcal{D}})$ *with respect to the sup norm metric, for* $d(i,s)$=$(N\pi_i)^{-1}$ *and any high entropy sampling design satisfying Assumption 4.2.2, where* $\mathbb{H}$ *is a mean* 0 *Gaussian process in* $\tilde{D}[0,1]$ *with almost sure continuous paths.*

The weak convergence of the empirical process mentioned in Proposition 4.2.1 is first shown under rejective sampling designs by establishing its tightness and finite dimensional convergence. Then, the weak convergence of this empirical process is shown under high entropy sampling designs using the fact that any high entropy sampling design can be approximated by a rejective sampling design in Kullback-Liebler divergence.

[7] and [43] showed the weak convergence of a similar version of the above-mentioned empirical process under some conditions on sampling designs (e.g., (HT2) in [7], and (F2) and (F3) in [43]). These conditions hold under very few sampling designs (with fixed sample size) like SRSWOR and rejective sampling designs. We are able to dispense with those conditions and show the weak convergence of $\left\{\sqrt{n}(\hat{F}_u(t)-t) : t \in [0,1]\right\}$ for $d(i,s)$=$(N\pi_i)^{-1}$ under any high entropy sampling design satisfying Assumption 4.2.2. Examples of such a sampling design are SRSWOR, LMS and HEπPS sampling designs. Recall from the introduction that a sampling design is called HEπPS sampling design if it is a high entropy as well as a πPS sampling design (e.g., RS sampling design, rejective sampling design, etc.). In particular, we are able to show the weak convergence of the aforementioned empirical process under LMS and HEπPS sampling designs, which are not covered in the earlier literature. Now, we state the following theorem.

Theorem 4.2.1. *Fix any* $0 < \alpha < \beta < 1$*. Suppose that Assumptions 4.2.1 and 4.2.3 hold, and* $E_{\boldsymbol{P}}||\boldsymbol{W}_i||^2 < \infty$ *for* $\boldsymbol{W}_i$=$(X_i, Y_i, X_iY_i, X_i^2)$*. Then, under the probability distribution* $\boldsymbol{P}^*$*,*

$$\{\sqrt{n}(G(p) - Q_{y,N}(p)) : p \in [\alpha,\beta]\} \xrightarrow{\mathcal{L}} \mathbb{Q} \text{ as } \nu \to \infty$$

in $(D[\alpha,\beta], \mathcal{D})$ *with respect to the sup norm metric, for any high entropy sampling design satisfying Assumption 4.2.2, where* $G(p)$ *denotes one of* $\hat{Q}_y(p)$*,* $\hat{Q}_{y,RA}(p)$*,* $\hat{Q}_{y,DI}(p)$ *and* $\hat{Q}_{y,REG}(p)$ *with* $d(i,s)$=$(N\pi_i)^{-1}$*, and* $\mathbb{Q}$ *is a mean* 0 *Gaussian process in* $D[\alpha,\beta]$ *with almost sure continuous path and p.d. covariance kernel*

$$K(p_1,p_2) = \lim_{\nu\to\infty}(n/N^2)E_{\mathbf{P}}\bigg(\sum_{i=1}^{N}(\zeta_i(p_1) - \overline{\zeta}(p_1) - S(p_1)\pi_i)\times \\ (\zeta_i(p_2) - \overline{\zeta}(p_2) - S(p_2)\pi_i)(\pi_i^{-1} - 1)\bigg) \text{ for } p_1, p_2 \in [\alpha, \beta]. \tag{4.2.3}$$

Here, $\overline{\zeta}(p)$=$\sum_{i=1}^{N}\zeta_i(p)/N$, $S(p)$=$\sum_{i=1}^{N}(\zeta_i(p) - \overline{\zeta}(p))(1 - \pi_i)/\sum_{i=1}^{N}\pi_i(1 - \pi_i)$, *and* $\zeta_i(p)$*'s are as in Table 4.1 below.*

TABLE 4.1: Expressions of $\zeta_i(p)$'s appearing in (4.2.3) and (4.2.5) for different $G(p)$'s in the cases of high entropy and RHC sampling designs.

$G(p)$	$\zeta_i(p)$
$\hat{Q}_y(p)$	$\mathbb{1}_{[Y_i\leq Q_y(p)]}/f_y(Q_y(p))$
$\hat{Q}_{y,RA}(p)$	$\mathbb{1}_{[Y_i\leq Q_y(p)]}/f_y(Q_y(p)) - (Q_y(p)/Q_x(p))\mathbb{1}_{[X_i\leq Q_x(p)]}/f_x(Q_x(p))$
$\hat{Q}_{y,DI}(p)$	$\mathbb{1}_{[Y_i\leq Q_y(p)]}/f_y(Q_y(p)) - (E_{\mathbf{P}}(Y_i)/E_{\mathbf{P}}(X_i))\mathbb{1}_{[X_i\leq Q_x(p)]}/f_x(Q_x(p))$
$\hat{Q}_{y,REG}(p)$	$\mathbb{1}_{[Y_i\leq Q_y(p)]}/f_y(Q_y(p)) - (E_{\mathbf{P}}(X_iY_i)/E_{\mathbf{P}}(X_i)^2)\mathbb{1}_{[X_i\leq Q_x(p)]}/f_x(Q_x(p))$

As discussed in the beginning of this chapter, the weak convergence of the quantile processes mentioned in Theorem 4.2.1 are shown under high entropy sampling designs using the weak convergence of empirical process mentioned in Proposition 4.2.1, Hadamard differentiability of the quantile map and the functional delta method. The weak convergence of the quantile process constructed based on the sample quantile for $d(i,s)$=$(N\pi_i)^{-1}$ was considered earlier in [26] under a high entropy sampling design. However, in [26], the author did not provide much details of the derivation of the main weak convergence result. Using dominated convergence theorem (DCT) and Lemma 4.8.10 in Section 4.8, $K(p_1,p_2)$ in (4.2.3) can be expressed in terms of superpopulation moments under SRSWOR, LMS and any HEπPS sampling designs as in Table 4.2 below.

TABLE 4.2: $K(p_1,p_2)$ in (4.2.3) under different high entropy sampling designs.

Sampling design	$K(p_1,p_2)$
SRSWOR and LMS	$(1-\lambda)E_{\mathbf{P}}\big[\zeta_i(p_1) - E_{\mathbf{P}}(\zeta_i(p_1))\big]\big[\zeta_i(p_2) - E_{\mathbf{P}}(\zeta_i(p_2))\big]$
HEπPS	[1] $E_{\mathbf{P}}\big[\zeta_i(p_1) - E_{\mathbf{P}}(\zeta_i(p_1)) + \lambda\chi^{-1}\mu_x^{-1}X_iE_{\mathbf{P}}\big((\zeta_i(p_1) - E_{\mathbf{P}}(\zeta_i(p_1)))X_i\big)\big]\times$ $\big[\zeta_i(p_2) - E_{\mathbf{P}}(\zeta_i(p_2)) + \lambda\chi^{-1}\mu_x^{-1}X_iE_{\mathbf{P}}\big((\zeta_i(p_2) - E_{\mathbf{P}}(\zeta_i(p_2)))X_i\big)\big]\times$ $\big[\mu_xX_i^{-1} - \lambda\big]$

[1] μ_x=$E_{\mathbf{P}}(X_i)$ and χ=$\mu_x - (\lambda E_{\mathbf{P}}(X_i)^2/\mu_x)$.

Next, we shall show the weak convergence of the quantile processes considered in Section 4.1 under RHC sampling design. Recall from the introduction that in RHC sampling design, the finite population $\mathcal{P}$ is first divided randomly into n disjoint groups of sizes $\tilde{N}_1 \cdots, \tilde{N}_n$, respectively, by taking a sample of $\tilde{N}_1$ units from N units with SRSWOR, a sample of $\tilde{N}_2$ units from $N - \tilde{N}_1$ units with SRSWOR and so on. Then, one unit is selected in the sample from each of these groups independently with probability proportional to the size variable x. [66] suggested this sampling design for constructing the well-known RHC estimator of the population mean. Advantages of the RHC estimator are discussed in Section 3.1 of Chapter 3. Before, we state the next theorem, let us consider some assumptions on the superpopulation distribution **P**.

Assumption 4.2.4. *There exists a constant* K *such that* $\max_{1\leq i\leq N} X_i/\min_{1\leq i\leq N} X_i \leq K$ *a.s.* [***P***].

Assumption 4.2.5. *The support of the joint distribution of* (Y_i, X_i) *is not a subset of a straight line in* $\mathbb{R}^2$.

As in the earlier chapters, here also we consider the following assumption.

Assumption 4.2.6. *For the RHC sampling design,* $\{\tilde{N}_r\}_{r=1}^n$ *are such that*

$$\tilde{N}_r = \begin{cases} N/n, \textit{ for } r = 1, \cdots, n, \textit{ when } N/n \textit{ is an integer}, \\ \lfloor N/n \rfloor, \textit{ for } r = 1, \cdots, k, \textit{ and} \\ \lfloor N/n \rfloor + 1, \textit{ for } r = k+1, \cdots, n, \textit{ when } N/n \textit{ is not an integer}, \end{cases} \tag{4.2.4}$$

where k *is such that* $\sum_{r=1}^n \tilde{N}_r = N$. *Here,* $\lfloor N/n \rfloor$ *is the integer part of* N/n.

Assumption 4.2.4 is equivalent to Assumption 4.2.2–*(ii)* under any πPS sampling design. Similar assumptions like Assumptions 4.2.4–4.2.6 are stated and discussed in Chapter 2 (see the discussion related to Assumptions 2.1.6 and 2.2.1 in Chapter 2). These assumptions are required to show the finite dimensional convergence of the empirical process $\{\sqrt{n}(\hat{F}_u(t) - t)) : t \in [0, 1]\}$ for $d(i, s) = (NX_i)^{-1} G_i$ under RHC sampling design, where G_i's are as in the 1^{st} paragraph of Section 4.1. Assumptions 4.2.4 and 4.2.6 are also required to establish the tightness of this empirical process. As in the case of high entropy sampling designs, here also we shall show the weak convergence of several quantile processes based on the weak convergence of the above-mentioned empirical process.

Proposition 4.2.2. *Suppose that* $E_{\boldsymbol{P}}(X_i)^{-1} < \infty$*, and Assumptions 4.2.1 and 4.2.3–4.2.6 hold. Then, the conclusion of Proposition 4.2.1 holds for* $d(i,s)=(NX_i)^{-1}G_i$ *and RHC sampling design.*

Theorem 4.2.2. *Fix any* $0 < \alpha < \beta < 1$*. Suppose that* $E_{\boldsymbol{P}}(X_i)^{-1} < \infty$*,* $E_{\boldsymbol{P}}||\boldsymbol{W}_i||^2 < \infty$ *for* $\boldsymbol{W}_i=(X_i, Y_i, X_iY_i, X_i^2)$*, and Assumptions 4.2.1 and 4.2.3–4.2.6 hold. Then, the conclusion of Theorem 4.2.1 holds for* $d(i,s)=(NX_i)^{-1}G_i$ *and RHC sampling design with p.d. covariance kernel*

$$\begin{aligned} K(p_1,p_2) &= \lim_{\nu\to\infty} n\gamma E_{\boldsymbol{P}}\big[(\overline{X}/N)\sum_{i=1}^{N}(\zeta_i(p_1)-\overline{\zeta}(p_1))(\zeta_i(p_2)-\overline{\zeta}(p_2))X_i^{-1}\big] \\ &= cE_{\boldsymbol{P}}(X_i)E_{\boldsymbol{P}}\big[\big(\zeta_i(p_1)-E_{\boldsymbol{P}}(\zeta_i(p_1))\big)\big(\zeta_i(p_2)-E_{\boldsymbol{P}}(\zeta_i(p_2))\big)X_i^{-1}\big] \end{aligned} \tag{4.2.5}$$

for $p_1, p_2 \in [\alpha, \beta]$.

Here, $\gamma=\sum_{r=1}^{n}\tilde{N}_r(\tilde{N}_r-1)/N(N-1)$*,* $c=\lim_{\nu\to\infty} n\gamma$*, and* $\zeta_i(p)$*'s are as in Table 4.1 above.*

The proof techniques of Proposition 4.2.2 and Theorem 4.2.2 are similar to the proof techniques of Proposition 4.2.1 and Theorem 4.2.1, respectively. It follows from Lemma 2.7.5 in Section 2.7 of Chapter 2 that $c=1-\lambda$ for λ^{-1} an integer, and $c=\lambda\lfloor\lambda^{-1}\rfloor(2-\lambda\lfloor\lambda^{-1}\rfloor-\lambda)$ when λ^{-1} is a non-integer. If we replace $Q_{y,N}$ by Q_y in the quantile processes considered in this section, then the weak convergence of these quantile processes can be shown under high entropy and RHC sampling designs using the key ideas of the proofs of Theorems 4.2.1 and 4.2.2.

4.3. Weak convergence of quantile processes under stratified multistage cluster sampling design

Stratified multistage cluster sampling design with SRSWOR is used instead of single stage sampling designs mentioned in the preceding section, when heterogenity is present in the population values of (y, x). Let us recall the definition of stratified multistage cluster sampling design with SRSWOR from the introduction. Suppose that the finite population $\mathcal{P}$ is divided into H strata or subpopulations, where stratum h consists of M_h clusters for h=$1,\ldots,H$. Further, the j^{th} cluster in stratum h consists of N_{hj} units for j=$1,\ldots,M_h$. For any given h=$1,\ldots,H$, j=$1,\ldots,M_h$ and l=$1,\ldots,N_{hj}$, we assume that the l^{th} unit from cluster j in stratum h is the i^{th} unit in the population $\mathcal{P}$, where $i=\sum_{h'=1}^{h}\sum_{j'=1}^{M_{h'}}N_{h'j'}-\sum_{j'=j}^{M_h}N_{hj'}+l$. In stratified multistage cluster sampling design with SRSWOR, first a sample s_h of m_h ($< M_h$) clusters is selected from stratum h under SRSWOR for each h. Then, a sample s_{hj} of r_h ($< N_{hj}$) units is selected from j^{th} cluster in stratum h if it is selected in the sample of clusters s_h in the first stage for h=$1,\ldots,H$. Thus the resulting sample is s=$\cup_{1\leq h\leq H, j\in s_h} s_{hj}$. The samplings in two stages are done independently across the strata and the clusters. Under the above sampling design, the inclusion probability of the i^{th} population unit is π_i=$m_h r_h/M_h N_{hj}$ if it belongs to the j^{th} cluster of stratum h. Note that stratified multistage cluster sampling design with SRSWOR becomes stratified sampling design with SRSWOR, when N_{hj}=1 for any h=$1,\ldots,H$ and j=$1,\ldots,M_h$. Also, note that stratified multistage cluster sampling design with SRSWOR becomes multistage cluster sampling design with SRSWOR, when H=1.

Suppose that (Y'_{hjl}, X'_{hjl}) denotes the value of (y,x) corresponding to the l^{th} unit from cluster j in stratum h. Note that given any h, j and l, (Y'_{hjl}, X'_{hjl})=(Y_i, X_i), where i=$\sum_{h'=1}^{h}\sum_{j'=1}^{M_{h'}} N_{h'j'} - \sum_{j'=j}^{M_h} N_{hj'} + l$ and (Y_i, X_i) is the value of (y,x) corresponding to the i^{th} population unit. We assume that for any given h=$1,\ldots,H$, $\{(Y'_{hjl}, X'_{hjl}) : l = 1,\ldots,N_{hj}, j = 1,\ldots,M_h\}$ are i.i.d. random vectors defined on $(\Omega, \mathcal{F}, \mathbf{P})$ with marginal distribution functions $F_{y,h}$ and $F_{x,h}$, where $F_{y,h}$'s and $F_{x,h}$'s are not necessarily identical for varying h. We also assume that the population observations on (y,x) in any stratum are independent of the observations in other strata. [35] used a similar superpopulation model set up for studying the asymptotic behavior of sample quantiles. However, they considered all $F_{y,h}$'s to be the same. Note that H, M_h, N_{hj}, m_h and r_h depend on ν, when we consider the sequence of populations $\{P_\nu\}$. However, for simplicity, we omit ν. As in the cases of single stage sampling designs, here also we shall show the weak convergence of various quantile processes based on the weak convergence of the empirical process $\{\sqrt{n}(\hat{F}_u(t) - t) : t \in [0,1]\}$ for $d(i,s)$=$(N\pi_i)^{-1}$.

First, we consider the case, when H is fixed as $\nu \to \infty$ (cf. [10]). In this case, we need the following assumptions to show that the conclusions of Proposition 4.2.1 and Theorem 4.2.1 hold for stratified multistage cluster sampling design with SRSWOR. Let N_h=$\sum_{j=1}^{M_h} N_{hj}$ and n_h=$m_h r_h$ be the number of population units in stratum h and the number of population units sampled from stratum h, respectively, for any h=$1,\ldots,H$.

Assumption 4.3.1. $\sum_{\nu=1}^{\infty} exp(-KM_h) < \infty$, $0 < \underline{\lim}_{\nu\to\infty} m_h/M_h \leq \overline{\lim}_{\nu\to\infty} m_h/M_h < 1$, $\lim_{\nu\to\infty} n_h/n$=$\lambda_h > 0$, $\lim_{\nu\to\infty} N_h/N$=$\Lambda_h > 0$ *and* $0 < \underline{\lim}_{\nu\to\infty} \min_{1\leq j\leq M_h} r_h/N_{hj} \leq \overline{\lim}_{\nu\to\infty} \max_{1\leq j\leq M_h} r_h/N_{hj} < 1$ *for any* h=$1,\ldots,H$ *and* $K > 0$, *and* $\max_{1\leq h\leq H}\sum_{j=1}^{M_h} N_{hj}^4/M_h$=$O(1)$ *and* $\max_{1\leq h\leq H}\sum_{j=1}^{M_h}(N_{hj} - N_h/M_h)^2/M_h \to 0$ *as* $\nu \to \infty$.

Assumption 4.3.2. *For any h=$1,\ldots,H$, the support of the joint distribution of (Y'_{hjl},X'_{hjl}) is not a subset of a straight line in $\mathbb{R}^2$, and $E_{\boldsymbol{P}}||\mathbf{W}'_{hjl}||^2<\infty$, where $\mathbf{W}'_{hjl}$=$(X'_{hjl},Y'_{hjl},X'_{hjl}Y'_{hjl},(X'_{hjl})^2)$.*

Assumption 4.3.3. *$supp(F_{y,h})$=$\mathcal{C}_y$ and $supp(F_{x,h})$=$\mathcal{C}_x$ for any h=$1,\ldots,H$ and some open intervals $\mathcal{C}_y\subseteq(-\infty,\infty)$ and $\mathcal{C}_x\subseteq(0,\infty)$. Moreover, $F_{y,h}$ and $F_{x,h}$ are continuous on $\mathbb{R}$ and are differentiable with positive continuous derivatives $f_{y,h}$ and $f_{x,h}$ on $\mathcal{C}_y$ and $\mathcal{C}_x$, respectively, for any h=$1,\ldots,H$ and $\nu\geq 1$.*

The condition $\sum_{\nu=1}^{\infty}exp(-KM_h)<\infty$ for any h=$1,\ldots,H$ and $K>0$ in Assumption 4.3.1 holds, when the number of clusters in each stratum is a strictly increasing function of ν. This condition implies that $M_1,\ldots,M_H$ grow infinitely as $\nu\rightarrow\infty$. The condition $\max_{1\leq h\leq H}\sum_{j=1}^{M_h}N_{hj}^4/M_h$=$O(1)$ as $\nu\rightarrow\infty$ in Assumption 4.3.1 holds, when cluster sizes in any stratum are not arbitrarily large. The condition $\max_{1\leq h\leq H}\sum_{j=1}^{M_h}(N_{hj}-N_h/M_h)^2/M_h\rightarrow 0$ as $\nu\rightarrow\infty$ in Assumption 4.3.1 implies that the variation among cluster sizes in each stratum is negligible. The rest of the conditions in Assumption 4.3.1 are often used in sample survey literature (see [62], [77] and references therein). Assumptions 4.3.1–4.3.3 are required to establish the finite dimensional convergence of the empirical process $\{\sqrt{n}(\hat{F}_u(t)-t):t\in[0,1]\}$ for $d(i,s)$=$(N\pi_i)^{-1}$ under stratified multistage cluster sampling design with SRSWOR, whereas Assumptions 4.3.1 and 4.3.3 are required to show the tightness of this empirical process under the same sampling design.

Next, we consider the case, when $H\rightarrow\infty$ as $\nu\rightarrow\infty$ (cf. [35], [77], etc.). In this case, we replace Assumption 4.3.1 by Assumption 4.3.4 and Assumption 4.3.2 by Assumption 4.3.5 given below, and consider some further assumptions to show that the conclusions of Proposition 4.2.1 and Theorem 4.2.1 hold for stratified multistage cluster sampling design with SRSWOR.

Assumption 4.3.4. *$\sum_{\nu=1}^{\infty}exp(-KH)<\infty$ for any $K>0$, and $\max_{1\leq h\leq H,1\leq j\leq M_h}nM_hN_{hj}/m_hN=O(1)$, $\sum_{h=1}^{H}M_h^4/H$=$O(1)$ and $\max_{1\leq h\leq H}\sum_{j=1}^{M_h}(N_{hj}-N_h/M_h)^2/M_h\rightarrow 0$ as $\nu\rightarrow\infty$.*

Next, suppose that $F_{y,H}(t)$=$\sum_{h=1}^{H}(N_h/N)F_{y,h}(t)$ and $F_{x,H}(t)$=$\sum_{h=1}^{H}(N_h/N)F_{x,h}(t)$, and $Q_{y,H}$ and $Q_{x,H}$ are quantile functions corresponding to $F_{y,H}$ and $F_{x,H}$, respectively. Let

$$\mathbf{R}'_{hjl}=(\mathbb{1}_{[Y'_{hjl}\leq Q_{y,H}(p_1)]},\ldots,\mathbb{1}_{[Y'_{hjl}\leq Q_{y,H}(p_k)]},\mathbb{1}_{[X'_{hjl}\leq Q_{x,H}(p_1)]},\ldots,\mathbb{1}_{[X'_{hjl}\leq Q_{x,H}(p_k)]})$$

for any $k \geq 1$ and $p_1, \ldots, p_k \in (0,1)$, and Γ_h=$E_{\mathbf{P}}(\mathbf{R}'_{hjl} - E_{\mathbf{P}}(\mathbf{R}'_{hjl}))^T\ (\mathbf{R}'_{hjl} - E_{\mathbf{P}}(\mathbf{R}'_{hjl}))$. Then, we consider the following assumptions.

Assumption 4.3.5. *Given any $k \geq 1$ and $p_1, \ldots, p_k \in (0,1)$, $\sum_{h=1}^{H} N_h(N_h - n_h)\Gamma_h/n_h N \rightarrow \Gamma_1$ and $\sum_{h=1}^{H} N_h \Gamma_h/N \rightarrow \Gamma_2$ as $\nu \rightarrow \infty$ for some positive definite matrices Γ_1 and Γ_2. Moreover, $\sum_{h=1}^{H}\sum_{j=1}^{M_h}\sum_{l=1}^{N_{hj}} \boldsymbol{W}'_{hjl}/N \rightarrow \Theta$=$(\Theta_1, \ldots, \Theta_4)$ and $\sum_{h=1}^{H}\sum_{j=1}^{M_h}\sum_{l=1}^{N_{hj}} ||\boldsymbol{W}'_{hjl}||^2/N$= $O(1)$ a.s. $[\boldsymbol{P}]$ as $\nu \rightarrow \infty$, where $\Theta_1 > 0$.*

Further, suppose that $f_{y,H}(t)$=$dF_{y,H}/dt$ and $f_{x,H}(t)$=$dF_{x,H}/dt$, and consider the following assumptions.

Assumption 4.3.6. *$supp(F_{y,h})$=$\mathcal{C}_y$ and $supp(F_{x,h})$=$\mathcal{C}_x$ for any h=$1, \ldots, H$ and some open intervals $\mathcal{C}_y \subseteq (-\infty, \infty)$ and $\mathcal{C}_x \subseteq (0, \infty)$. Further, there exists a distribution function $\tilde{F}_y$ with $supp(\tilde{F}_y)$=$\mathcal{C}_y$ and positive continuous derivative $\tilde{f}_y$ such that $F_{y,H}(t) \rightarrow \tilde{F}_y(t)$ for any $t \in \mathbb{R}$ and $\sup_{\mathcal{C}_y} |f_{y,H}(t) - \tilde{f}_y(t)| \rightarrow 0$ as $\nu \rightarrow \infty$. There also exists a distribution function $\tilde{F}_x$ with $supp(\tilde{F}_x)$=$\mathcal{C}_x$ and positive continuous derivative $\tilde{f}_x$ such that $F_{x,H}(t) \rightarrow \tilde{F}_x(t)$ for any $t \in \mathbb{R}$ and $\sup_{\mathcal{C}_x} |f_{x,H}(t) - \tilde{f}_x(t)| \rightarrow 0$ as $\nu \rightarrow \infty$.*

The condition $\max_{1\leq h\leq H, 1\leq j\leq M_h} nM_h N_{hj}/m_h N$=$O(1)$ as $\nu \rightarrow \infty$ in Assumption 4.3.4 was considered earlier in the literature (cf. [77]). This condition and Assumption 4.2.1 imply that cluster sizes in any stratum cannot be arbitrarily large. The condition $\sum_{h=1}^{H} M_h^4/H$=$O(1)$ as $\nu \rightarrow \infty$ in Assumption 4.3.4 holds, when the number of clusters in any stratum is not arbitrarily large. Assumption 4.3.6 implies that $F_{y,H}$ and $F_{x,H}$ can be approximated by the distribution functions $\tilde{F}_y$ and $\tilde{F}_x$, respectively, when $H \rightarrow \infty$ as $\nu \rightarrow \infty$. This assumption also implies that $f_{y,H}$ and $f_{x,H}$ can be approximated uniformly by the density functions of $\tilde{F}_y$ and $\tilde{F}_x$, respectively, when $H \rightarrow \infty$ as $\nu \rightarrow \infty$. Assumptions 4.3.4 and 4.3.5 are required to show the finite dimensional convergence of the empirical process $\{\sqrt{n}(\hat{F}_u(t) - t) : t \in [0,1]\}$ for $d(i,s)$=$(N\pi_i)^{-1}$ under stratified multistage cluster sampling design with SRSWOR, and Assumptions 4.3.4 and 4.3.6 are required to establish the tightness of this empirical process under the same sampling design. Now, we state the following results.

Proposition 4.3.1. *(i) Suppose that H is fixed as $\nu \rightarrow \infty$, and Assumptions 4.2.1 and 4.3.1–4.3.3 hold. Then, the conclusion of Proposition 4.2.1 holds for stratified multistage cluster sampling design with SRSWOR.*

(ii) Further, if $H \rightarrow \infty$ as $\nu \rightarrow \infty$, and Assumptions 4.2.1 and 4.3.3–4.3.6 hold, then the same result holds.

Theorem 4.3.1. *(i) Suppose that H is fixed as $\nu \to \infty$, and Assumptions 4.2.1 and 4.3.1–4.3.3 hold. Then, the conclusion of Theorem 4.2.1 holds for stratified multistage cluster sampling design with SRSWOR with p.d. covariance kernel*

$$K(p_1,p_2) = \lim_{\nu\to\infty}(n/N^2)\sum_{h=1}^{H} N_h(N_h-n_h)E_{\boldsymbol{P}}\big(\zeta'_{hjl}(p_1) - E_{\boldsymbol{P}}(\zeta'_{hjl}(p_1))\big)\times \big(\zeta'_{hjl}(p_2) - E_{\boldsymbol{P}}(\zeta'_{hjl}(p_2))\big)/n_h \text{ for } p_1,p_2 \in [\alpha,\beta]. \quad (4.3.1)$$

Here, $\zeta'_{hjl}(p)$'s are as in Table 4.3 below.

(ii) Further, if $H \to \infty$ as $\nu \to \infty$, and Assumption 4.2.1 and 4.3.3–4.3.6 hold, then the same result holds.

TABLE 4.3: Expressions of $\zeta'_{hjl}(p)$'s appearing in (4.3.1) for different $G(p)$'s in the case of stratified multistage cluster sampling design with SRSWOR.

	$G(p)$	$\zeta'_{hjl}(p)$
H is fixed as $\nu \to \infty$	$\hat{Q}_y(p)$	$\mathbb{1}_{[Y'_{hjl}\le Q_{y,H}(p)]}/f_{y,H}(Q_{y,H}(p))$
	$\hat{Q}_{y,RA}(p)$	$\mathbb{1}_{[Y'_{hjl}\le Q_{y,H}(p)]}/f_{y,H}(Q_{y,H}(p)) - (Q_{y,H}(p)/Q_{x,H}(p))\times$ $\mathbb{1}_{[X'_{hjl}\le Q_{x,H}(p)]}/f_{x,H}(Q_{x,H}(p))$
	$\hat{Q}_{y,DI}(p)$	$\mathbb{1}_{[Y'_{hjl}\le Q_{y,H}(p)]}/f_{y,H}(Q_{y,H}(p)) - (\sum_{h=1}^{H}(N_h/N)E_{\mathbf{P}}(Y'_{hjl})/$ ${}^1\sum_{h=1}^{H}(N_h/N)E_{\mathbf{P}}(X'_{hjl}))\mathbb{1}_{[X'_{hjl}\le Q_{x,H}(p)]}/f_{x,H}(Q_{x,H}(p))$
	$\hat{Q}_{y,REG}(p)$	$\mathbb{1}_{[Y'_{hjl}\le Q_{y,H}(p)]}/f_{y,H}(Q_{y,H}(p)) - (\sum_{h=1}^{H}(N_h/N)E_{\mathbf{P}}(X'_{hjl}Y'_{hjl})/$ $\sum_{h=1}^{H}(N_h/N)E_{\mathbf{P}}(X'_{hjl})^2)\mathbb{1}_{[X'_{hjl}\le Q_{x,H}(p)]}/f_{x,H}(Q_{x,H}(p))$
$H \to \infty$ as $\nu \to \infty$	$\hat{Q}_y(p)$	$\mathbb{1}_{[Y'_{hjl}\le Q_{y,H}(p)]}/f_{y,H}(Q_{y,H}(p))$
	$\hat{Q}_{y,RA}(p)$	$\mathbb{1}_{[Y'_{hjl}\le Q_{y,H}(p)]}/f_{y,H}(Q_{y,H}(p)) - (Q_{y,H}(p)/Q_{x,H}(p))\times$ $\mathbb{1}_{[X'_{hjl}\le Q_{x,H}(p)]}/f_{x,H}(Q_{x,H}(p))$
	$\hat{Q}_{y,DI}(p)$	$\mathbb{1}_{[Y'_{hjl}\le Q_{y,H}(p)]}/f_{y,H}(Q_{y,H}(p)) - ({}^2\Theta_2/{}^2\Theta_1)\times$ $\mathbb{1}_{[X'_{hjl}\le Q_{x,H}(p)]}/f_{x,H}(Q_{x,H}(p))$
	$\hat{Q}_{y,REG}(p)$	$\mathbb{1}_{[Y'_{hjl}\le Q_{y,H}(p)]}/f_{y,H}(Q_{y,H}(p)) - ({}^2\Theta_3/{}^2\Theta_4)\times$ $\mathbb{1}_{[X'_{hjl}\le Q_{x,H}(p)]}/f_{x,H}(Q_{x,H}(p))$

[2] Θ_1, Θ_2, Θ_3 and Θ_4 are as in Assumption 4.3.5 in Section 4.3.

The proof techniques of Proposition 4.3.1 and Theorem 4.3.1 are similar to the proof techniques of Proposition 4.2.1 and Theorem 4.2.1, respectively. The weak convergence of the above quantile processes with $Q_{y,N}$ replaced by Q_y can be shown using the key ideas of the proof of Theorem 4.3.1. When H is fixed as $\nu \to \infty$, the expression of $K(p_1,p_2)$ in (4.3.1) can be further simplified (cf. (4.7.42) in Section 4.7) because $N_h/N \to \Lambda_h$ and $(n/N^2)N_h(N_h-n_h)/n_h \to \lambda\Lambda_h(\Lambda_h/\lambda\lambda_h - 1)$ as $\nu \to \infty$ for any h=1, . . . , H by Assumptions 4.2.1 and 4.3.1.

4.4. Functions of quantile processes

In this section, we derive the asymptotic normality of the smooth L-estimators as well as the estimators of smooth functions of finite population quantiles, which are based on the sample quantile, and the ratio, the difference and the regression estimators of the population quantiles, under sampling designs considered in the preceding two sections. The smooth L-estimators include the estimators of the population α-trimmed means, whereas the estimators of smooth functions of population quantiles include the estimators of any specific quantile, the interquartile range and the quantile based measure of skewness in the population. Note that non smooth L-estimators are also special cases of these latter estimators. Some asymptotic normality results related to the estimators of the population quantiles are available in sample survey literature. [53] showed that the ratio estimator of the population median is asymptotically normal under SRSWOR. [35] derived the asymptotitc normality of the sample quantile under stratified cluster sampling design with SRSWOR based on superpopulation model assumptions. [85] derived the asymptotitc normality of the sample quantile under some conditions on sampling deigns. [18] derived the asymptotic normality of the sample median under SRSWOR based on superpopulation model assumptions. [77] derived the asymptotic normality of smooth and non smooth L-estimators, which are constructed based on the sample quantile, under stratified multistage cluster sampling design with SRSWOR. However, there is no result present in the existing literature related to the asymptotic normality of the smooth L-estimators and the estimators of smooth functions of population quantiles, which are based on the ratio, the difference and the regression estimators of the population quantile. There is also no result present in the available literature related to the asymptotic normality of the above estimators under high entropy and RHC sampling designs.

Let us fix $0 < \alpha < \beta < 1$ and consider the finite population parameter $\int_{[\alpha,\beta]} Q_{y,N}(p)J(p)dp$ for some known smooth function J on $[0,1]$. It follows from the definition of $Q_{y,N}(p)$ that the above parameter coincides with the population α-trimmed mean

$$\tau_{\alpha,N} = \left(\sum_{i=\lfloor N\alpha \rfloor+2}^{N-\lfloor N\alpha \rfloor-1} Y_{(i)} + (1-\{N\alpha\})(Y_{(\lfloor N\alpha \rfloor+1)} + Y_{(N-\lfloor N\alpha \rfloor)}) \right) \Big/ N(1-2\alpha)$$

when $0 < \alpha < 1/2$, β=$1-\alpha$ and $J(p)$=$1/(1-2\alpha)$, $p \in [0,1]$. Here, $Y_{(1)}, \ldots, Y_{(N)}$ are the ordered population values of y, and $[N\alpha]$ and $\{N\alpha\}$ are, respectively, the integer and the fractional parts of $N\alpha$. Several estimators of $\int_{[\alpha,\beta]} Q_{y,N}(p)J(p)dp$ can be constructed by plugging $\hat{Q}_y(p), \hat{Q}_{y,RA}(p), \hat{Q}_{y,DI}(p)$ and $\hat{Q}_{y,DI}(p)$ into $\int_{[\alpha,\beta]} Q_{y,N}(p)J(p)dp$. Note that these estimators can be expressed as weighted linear combinations of the ordered sample observations on y, where the weights are mainly generated by the smooth function J. That is why these estimators are called smooth L-estimators (cf. [77]).

Next, suppose that $k \geq 1$, $p_1, \ldots, p_k \in (0,1)$, $f : \mathbb{R}^k \to \mathbb{R}$ is a smooth function, and $f(Q_{y,N}(p_1), \ldots, Q_{y,N}(p_k))$ is a finite population parameter. Some examples of such a parameter are given in Table 4.4 below. Several estimators of $f(Q_{y,N}(p_1), \ldots, Q_{y,N}(p_k))$ can be constructed by plugging $\hat{Q}_y(p)$, $\hat{Q}_{y,RA}(p)$, $\hat{Q}_{y,DI}(p)$ and $\hat{Q}_{y,DI}(p)$ in $f(Q_{y,N}(p_1), \ldots, Q_{y,N}(p_k))$.

TABLE 4.4: Examples of $f(Q_{y,N}(p_1), \ldots, Q_{y,N}(p_k))$.

Parameter	k	$p_1, \ldots, p_k$	f
Median	1	p_1=0.5	$f(t)$=t
Interquartile range	2	p_1=0.25, p_2=0.75	$f(t_1, t_2)$=$t_2 - t_1$
Bowley's measure of skewness	3	p_1=0.25, p_2=0.5, p_3=0.75	$f(t_1, t_2)$=$(t_1 + t_3 - 2t_2)/(t_3 - t_1)$

Now, we state the asymptotic normality results for the above estimators.

Theorem 4.4.1. *(i) Fix $0 < \alpha < \beta < 1$. Suppose that the conclusion of Theorem 4.2.1 holds and $K(p_1, p_2)$ in* (4.2.3) *is continuous on $[\alpha, \beta] \times [\alpha, \beta]$. Then, under* ***P*****,*

$$\begin{aligned} &\sqrt{n}\left(\int_{[\alpha,\beta]} G(p)J(p)dp - \int_{[\alpha,\beta]} Q_{y,N}(p)J(p)dp\right) \xrightarrow{\mathcal{L}} N(0, \sigma_1^2) \text{ and} \\ &\sqrt{n}\left(f(G(p_1), \ldots, G(p_k)) - f(Q_{y,N}(p_1), \ldots, Q_{y,N}(p_k))\right) \xrightarrow{\mathcal{L}} N(0, \sigma_2^2) \text{ as } \nu \to \infty \end{aligned} \tag{4.4.1}$$

for any high entropy sampling design, where $k \geq 1$, $p_1, \ldots, p_k \in [\alpha, \beta]$, and $G(p)$ is one of $\hat{Q}_y(p)$, $\hat{Q}_{y,RA}(p)$, $\hat{Q}_{y,DI}(p)$ and $\hat{Q}_{y,REG}(p)$ with $d(i,s)$=$(N\pi_i)^{-1}$. Here,

$$\sigma_1^2 = \int_\alpha^\beta \int_\alpha^\beta K(p_1, p_2)J(p_1)J(p_2)dp_1dp_2,\ \sigma_2^2 = a\Delta a^T, \tag{4.4.2}$$

Δ is a $k \times k$ matrix such that

$$((\Delta))_{ij} = K(p_i, p_j) \text{ for } 1 \leq i, j \leq k, \text{ and } a = \lim_{\nu\to\infty} \nabla f(Q_{y,N}(p_1), \ldots, Q_{y,N}(p_k)) \tag{4.4.3}$$

a.s. [***P***].

(ii) Further, if the assumptions of Theorem 4.2.2 hold, then the results in (4.4.1) *hold for $d(i,s)$=$(NX_i)^{-1}G_i$ in the case of RHC sampling design.*

It can be shown using the expressions in Table 4.2 and Assumption 4.2.3 that $K(p_1, p_2)$ in (4.2.3) is continuous on $[\alpha, \beta] \times [\alpha, \beta]$ under SRSWOR, LMS and πPS sampling designs. Next, we state the following theorem.

Theorem 4.4.2. *(i) Fix $0 < \alpha < \beta < 1$. Suppose that H is fixed as $\nu \to \infty$, and Assumptions 4.2.1 and 4.3.1–4.3.3 hold, then the results in* (4.4.1) *of Theorem 4.4.1 hold for $d(i, s)$=$(N\pi_i)^{-1}$= $M_h N_{hj}/N m_h r_h$ under stratified multistage cluster sampling design with SRSWOR.*
(ii) On the other hand, if $H \to \infty$ as $\nu \to \infty$, Assumptions 4.2.1 and 4.3.3–4.3.6 hold, and $K(p_1, p_2)$ in (4.3.1) *is continuous on $[\alpha, \beta] \times [\alpha, \beta]$, then the same results hold.*

When $H \to \infty$ as $\nu \to \infty$ in the case of stratified multistage cluster sampling design with SRSWOR, it can be shown that $K(p_1, p_2)$ in (4.3.1) is continuous on $[\alpha, \beta] \times [\alpha, \beta]$ if the limit in the expression of $K(p_1, p_2)$ in (4.3.1) exists uniformly over $[\alpha, \beta] \times [\alpha, \beta]$.

4.4.1 Estimation of asymptotic covariance kernels and confidence intervals

Suppose that

$$\theta_1 = \int_{[\alpha,\beta]} Q_{y,N}(p)J(p)dp, \theta_2 = f(Q_{y,N}(p_1), \ldots, Q_{y,N}(p_k)),$$
$$\hat{\theta}_1 = \int_{[\alpha,\beta]} G(p)J(p)dp \text{ and } \hat{\theta}_2 = f(G(p_1), \ldots, G(p_k)),$$

where $G(p)$ is one of $\hat{Q}_y(p)$, $\hat{Q}_{y,RA}(p)$, $\hat{Q}_{y,DI}(p)$ and $\hat{Q}_{y,REG}(p)$. Then, $\sqrt{n}(\hat{\theta}_i - \theta_i) \xrightarrow{\mathcal{L}} N(0, \sigma_i^2)$ for i=1, 2, where σ_i^2's are as in Theorem 4.4.1. Further, suppose that $\hat{\sigma}_i \xrightarrow{p} \sigma_i$ for some estimator $\hat{\sigma}_i$ of σ_i, i=1, 2. Then, a $100(1-\eta)\%$ confidence interval for θ_i can be constructed as

$$[\hat{\theta}_i - Z_{\eta/2}\hat{\sigma}_i/\sqrt{n}, \hat{\theta}_i + Z_{\eta/2}\hat{\sigma}_i/\sqrt{n}] \text{ for } i = 1, 2,$$

where $Z_{\eta/2}$ is the $(1-\eta/2)^{th}$ quantile of the standard normal distribution. We now discuss the estimation of the asymptotic covariance kernels mentioned in (4.2.3), (4.2.5) and (4.3.1) based on which consistent estimators of σ_i^2's will be constructed.

Following the approach of [16], $K(p_1, p_2)$, for $d(i, s)$=$(N\pi_i)^{-1}$, under any high entropy sampling design (see (4.2.3)) can be estimated by

$$\hat{K}(p_1,p_2) = (n/N^2)\sum_{i\in s}(\hat{\zeta}_i(p_1) - \hat{\bar{\zeta}}(p_1) - \hat{S}(p_1)\pi_i)\times \\ (\hat{\zeta}_i(p_2) - \hat{\bar{\zeta}}(p_2) - \hat{S}(p_2)\pi_i)(\pi_i^{-1} - 1)\pi_i^{-1}, \tag{4.4.4}$$

where $\hat{\bar{\zeta}}(p)=\sum_{i\in s}(N\pi_i)^{-1}\hat{\zeta}_i(p)$ and $\hat{S}(p)=\sum_{i\in s}(\hat{\zeta}_i(p) - \hat{\bar{\zeta}}(p))(\pi_i^{-1} - 1)/\sum_{i\in s}(1-\pi_i)$. Here, $\hat{\zeta}_i(p)$ is obtained by replacing the superpopulation parameters involved in the expression of $\zeta_i(p)$ in Table 4.1 by their estimators under high entropy sampling designs (see Table 4.5 below). Note that $\sqrt{n}(\hat{Q}_y(p+1/\sqrt{n}) - \hat{Q}_y(p-1/\sqrt{n}))/2$ was considered as an estimator of $1/f_y(Q_y(p))$ earlier in [77].

Next, $K(p_1,p_2)$, for $d(i,s)=(NX_i)^{-1}G_i$ under RHC sampling design (see (4.2.5)), can be estimated as

$$\hat{K}(p_1,p_2) = n\gamma(\overline{X}/N)\sum_{i\in s}G_i(\hat{\zeta}_i(p_1) - \hat{\bar{\zeta}}(p_1))(\hat{\zeta}_i(p_2) - \hat{\bar{\zeta}}(p_2))X_i^{-2}, \tag{4.4.5}$$

where $\hat{\bar{\zeta}}(p)=\sum_{i\in s}(NX_i)^{-1}G_i\hat{\zeta}_i(p)$. Here, $\hat{\zeta}_i(p)$ is obtained by replacing the superpopulation parameters involved in the expression of $\zeta_i(p)$ in Table 4.1 by their estimators under RHC sampling design (see Table 4.5 below).

TABLE 4.5: Estimators of various superpopulation parameters involved in the expression of $\zeta_i(p)$ in Table 4.1 for high entropy and RHC sampling designs.

Parameters	Estimators	
	High entropy sampling designs	RHC sampling design
$Q_y(p)$	$\hat{Q}_y(p)$ with $d(i,s)=(N\pi_i)^{-1}$	$\hat{Q}_y(p)$ with $d(i,s)=(NX_i)^{-1}G_i$
$Q_x(p)$	$\hat{Q}_x(p)$ with $d(i,s)=(N\pi_i)^{-1}$	$\hat{Q}_x(p)$ with $d(i,s)=(NX_i)^{-1}G_i$
$1/f_y(Q_y(p))$	$\sqrt{n}(\hat{Q}_y(p+1/\sqrt{n})-$ $\hat{Q}_y(p-1/\sqrt{n}))/2$	$\sqrt{n}(\hat{Q}_y(p+1/\sqrt{n})-$ $\hat{Q}_y(p-1/\sqrt{n}))/2$
$1/f_x(Q_y(p))$	$\sqrt{n}(\hat{Q}_x(p+1/\sqrt{n})-$ $\hat{Q}_x(p-1/\sqrt{n}))/2$	$\sqrt{n}(\hat{Q}_x(p+1/\sqrt{n})-$ $\hat{Q}_x(p-1/\sqrt{n}))/2$
$E_{\mathbf{P}}(Y_i)$	$\sum_{i\in s}(N\pi_i)^{-1}Y_i$	$\sum_{i\in s}(NX_i)^{-1}G_iY_i$
$E_{\mathbf{P}}(X_i)$	$\sum_{i\in s}(N\pi_i)^{-1}X_i$	$\sum_{i\in s}N^{-1}G_i$
$E_{\mathbf{P}}(X_iY_i)$	$\sum_{i\in s}(N\pi_i)^{-1}X_iY_i$	$\sum_{i\in s}N^{-1}G_iY_i$
$E_{\mathbf{P}}(X_i)^2$	$\sum_{i\in s}(N\pi_i)^{-1}X_i^2$	$\sum_{i\in s}N^{-1}G_iX_i$

Given an estimator $\hat{K}(p_1,p_2)$ of $K(p_1,p_2)$, an estimator of σ_1^2 can be constructed as $\hat{\sigma}_1^2$ $=\int_\alpha^\beta\int_\alpha^\beta \hat{K}(p_1,p_2)J(p_1)J(p_2)\,dp_1dp_2$, whereas an estimator σ_2^2 can be constructed as $\hat{\sigma}_2^2=\hat{a}\hat{\Delta}\hat{a}^T$.

Here, $\hat{a}$=$\nabla f(\hat{Q}_y(p_1),\ldots,\hat{Q}_y(p_k))$, $k\geq 1$, $p_1,\ldots,p_k\in[\alpha,\beta]$, and $\hat{\Delta}$ is a $k\times k$ matrix such that $((\hat{\Delta}))_{ij}$=$\hat{K}(p_i,p_j)$ for $1\leq i,j\leq k$. In the following theorem, we assert that the above estimators of σ_1^2 and σ_2^2 are consistent.

Theorem 4.4.3. *(i) Fix $0<\alpha<\beta<1$. Suppose that the assumptions of Theorem 4.2.1 hold, $K(p_1,p_2)$ is as in* (4.2.3)*, and $\hat{K}(p_1,p_2)$ is as in* (4.4.4)*. Then, under $\boldsymbol{P}^*$,*

$$\hat{\sigma}_i^2\xrightarrow{p}\sigma_i^2 \textit{ as } \nu\rightarrow\infty \textit{ for } i=1,2 \tag{4.4.6}$$

and any high entropy sampling design satisfying Assumption 4.2.2.
(ii) Further, if the assumptions of Theorem 4.2.2 hold, $K(p_1,p_2)$ is as in (4.2.5)*, and $\hat{K}(p_1,p_2)$ is as in* (4.4.5)*. Then, the result in* (4.4.6) *hold under RHC sampling design.*

Next, for the case of stratified multistage cluster sampling design with SRSWOR , $E_{\mathbf{P}}(\zeta'_{hjl}(p_1)-E_{\mathbf{P}}(\zeta'_{hjl}(p_1)))(\zeta'_{hjl}(p_2)-E_{\mathbf{P}}(\zeta'_{hjl}(p_2)))$ in the expression of $K(p_1,p_2)$ in (4.3.1) can be estimated as

$$\sum_{j\in s_h}\sum_{l\in s_{hj}} M_hN_{hj}(\hat{\zeta}_{hjl}(p_1)-\hat{\bar{\zeta}}_h(p_1))(\hat{\zeta}_{hjl}(p_2)-\hat{\bar{\zeta}}_h(p_2))/m_hr_hN_h,$$

where $\hat{\bar{\zeta}}_h(p)$=$\sum_{j\in s_h}\sum_{l\in s_{hj}}M_hN_{hj}\hat{\zeta}_{hjl}(p)/m_hr_hN_h$, and h=$1,\ldots,H$. Here, $\hat{\zeta}_{hjl}(p)$ is obtained by replacing the superpopulation parameters involved in the expression of $\zeta'_{hjl}(p)$ in Table 4.3 by their estimators as mentioned in Table 4.6 below. Thus an estimator of $K(p_1,p_2)$ in (4.3.1)

TABLE 4.6: Estimators of various superpopulation parameters involved in the expression of $\zeta'_{hjl}(p)$ in Table 4.3 for stratified multistage cluster sampling design with SRSWOR.

Parameters	Estimators
$Q_{y,H}(p)$	$\hat{Q}_y(p)$ with $d(i,s)$=$(N\pi_i)^{-1}$=M_hN_{hj}/Nm_hr_h
$Q_{x,H}(p)$	$\hat{Q}_x(p)$ with $d(i,s)$=$(N\pi_i)^{-1}$=M_hN_{hj}/Nm_hr_h
$1/f_{y,H}(Q_{y,H}(p))$	$\sqrt{n}(\hat{Q}_y(p+1/\sqrt{n})-\hat{Q}_y(p-1/\sqrt{n}))/2$
$1/f_{x,H}(Q_{x,H}(p))$	$\sqrt{n}(\hat{Q}_x(p+1/\sqrt{n})-\hat{Q}_x(p-1/\sqrt{n}))/2$
$\sum_{h=1}^H(N_h/N)E_{\mathbf{P}}(X'_{hjl})$ as well as [2]Θ_1	$\sum_{h=1}^H\sum_{j\in s_h,l\in s_{hj}}M_hN_{hj}X'_{hjl}/m_hr_hN$
$\sum_{h=1}^H(N_h/N)E_{\mathbf{P}}(Y'_{hjl})$ as well as [2]Θ_2	$\sum_{h=1}^H\sum_{j\in s_h,l\in s_{hj}}M_hN_{hj}Y'_{hjl}/m_hr_hN$
$\sum_{h=1}^H(N_h/N)E_{\mathbf{P}}(X'_{hjl}Y'_{hjl})$ as well as [2]Θ_3	$\sum_{h=1}^H\sum_{j\in s_h,l\in s_{hj}}M_hN_{hj}X'_{hjl}Y'_{hjl}/m_hr_hN$
$\sum_{h=1}^H(N_h/N)E_{\mathbf{P}}(X'_{hjl})^2$ as well as [2]Θ_4	$\sum_{h=1}^H\sum_{j\in s_h,l\in s_{hj}}M_hN_{hj}(X'_{hjl})^2/m_hr_hN$

[2] Θ_1, Θ_2, Θ_3 and Θ_4 are as in Assumption 4.3.5 in Section 4.3.

under stratified multistage cluster sampling design with SRSWOR is obtained as

$$\hat{K}(p_1,p_2) = (n/N^2)\sum_{h=1}^{H}(N_h^2/n_h - N_h)\sum_{j\in s_h}\sum_{l\in s_{hj}} M_h N_{hj}(\hat{\zeta}_{hjl}(p_1) - \hat{\bar{\zeta}}_h(p_1))\times \\ (\hat{\zeta}_{hjl}(p_2) - \hat{\bar{\zeta}}_h(p_2))/m_h r_h N_h. \tag{4.4.7}$$

Given $\hat{K}(p_1,p_2)$, estimators of σ_1^2 and σ_2^2 can be constructed under stratified multistage cluster sampling design with SRSWOR in the same way as in the case of single stage sampling designs discussed in the paragraph preceding Theorem 4.4.3. Now, we state the following theorem.

Theorem 4.4.4. *Fix $0 < \alpha < \beta < 1$. Suppose that the assumptions of Theorem 4.3.1 hold, $K(p_1,p_2)$ is as in* (4.3.1)*, and $\hat{K}(p_1,p_2)$ is as in* (4.4.7)*. Then, the result in* (4.4.6) *of Theorem 4.4.3 hold under stratified multistage cluster sampling design with SRSWOR.*

4.5. Comparison of different estimators

4.5.1 Comparison of the estimators of functions of quantiles

In this section, we shall first compare different estimators of the finite population parameter $\int_{[\alpha,\beta]} Q_{y,N}(p)J(p)dp$ as well as $f(Q_{y,N}(p_1),\cdots,Q_{y,N}(p_k))$ (see Section 4.4) under each of SRSWOR, RHC and any HEπPS sampling designs in terms of their asymptotic variances given in Theorem 4.4.1. Here, $0 < \alpha < \beta < 1$, $k \geq 1$ and $p_1,\ldots,p_k \in (0,1)$. Recall from Section 4.4 that these parameters include the median, the α-trimmed mean, the interquartile range and the quantile based measure of skewness. Let us assume that $P(s,\omega)$ is one of SRSWOR, RHC and any HEπPS sampling designs. Let us also assume that $K_1(p_1,p_2)$, $K_2(p_1,p_2)$, $K_3(p_1,p_2)$ and $K_4(p_1,p_2)$ are the asymptotic covariance kernels of the quantile processes constructed based on $\hat{Q}_y(p)$, $\hat{Q}_{y,RA}(p)$, $\hat{Q}_{y,DI}(p)$ and $\hat{Q}_{y,REG}(p)$ under $P(s,\omega)$, respectively (see Table 4.2 and (4.2.5)), and $\{\Delta_i : 1 \leq i \leq 4\}$ is a $k \times k$ matrix such that

$$((\Delta_i))_{jl} = K_i(p_j,p_l) \text{ for } 1 \leq j,l \leq k \text{ and } 1 \leq i \leq 4. \tag{4.5.1}$$

Then, we have the following theorem.

Theorem 4.5.1. *Suppose that $X_i \leq b$ a.s. $[\boldsymbol{P}]$ for some $b > 0$, $E_{\boldsymbol{P}}(X_i)^{-1} < \infty$, Assumption 4.2.1 holds with $0 < \lambda < E_{\boldsymbol{P}}(X_i)/b$, and Assumptions 4.2.4, 4.2.5 and 4.2.6 hold. Then, we have* *the following results.*

(i) Under $P(s,\omega)$, the asymptotic variance of the estimator of $\int_{[\alpha,\beta]} Q_{y,N}(p)J(p)dp$ based on the sample quantile is smaller than the asymptotic variances of its estimators based on the ratio, the difference and the regression estimators of the finite population quantile if and only if

$$\max_{2\leq i\leq 4}\left\{\int_{\alpha}^{\beta}\int_{\alpha}^{\beta}\big(K_1(p_1,p_2)-K_i(p_1,p_2)\big)J(p_1)J(p_2)dp_1dp_2\right\}<0 \qquad (4.5.2)$$

(ii) Under $P(s,\omega)$, the asymptotic variance of the estimator of $f(Q_{y,N}(p_1),\cdots,Q_{y,N}(p_k))$ based on the sample quantile is smaller than the asymptotic variances of its estimators based on the ratio, the difference and the regression estimators of the finite population quantile if and only if

$$\max_{2\leq i\leq 4} a(\Delta_1-\Delta_i)a^T<0, \qquad (4.5.3)$$

where a=$\nabla f(Q_y(p_1),\cdots,Q_y(p_k))$ is the gradient of f at $\big(Q_y(p_1),\cdots,Q_y(p_k)\big)$.

Next, we shall compare the performances of each of the estimators of $\int_{[\alpha,\beta]} Q_{y,N}(p)J(p)dp$ as well as $f(Q_{y,N}(p_1),\cdots,Q_{y,N}(p_k))$ considered in Section 4.4 under SRSWOR, RHC and any HEπPS sampling designs in terms of their asymptotic variances (see Theorem 4.4.1). Let us assume that $G(p)$ is one of $\hat{Q}_y(p)$, $\hat{Q}_{y,RA}(p)$, $\hat{Q}_{y,DI}(p)$ and $\hat{Q}_{y,REG}(p)$, $K_1^*(p_1,p_2)$, $K_2^*(p_1,p_2)$ and $K_3^*(p_1,p_2)$ denote asymptotic covariance kernels of $\{\sqrt{n}(G(p)-Q_{y,N}(p)):p\in[\alpha,\beta]\}$ under SRSWOR, RHC and any HEπPS sampling designs (see Table 4.2 and (4.2.5)), respectively, and $\{\Delta_i^*:1\leq i\leq 3\}$ are $k\times k$ matrices such that

$$((\Delta_i^*))_{jl}=K_i^*(p_j,p_l) \text{ for } 1\leq j,l\leq k \text{ and } 1\leq i\leq 3. \qquad (4.5.4)$$

Then, we have the following theorem.

Theorem 4.5.2. *Suppose that $X_i\leq b$ a.s. $[\boldsymbol{P}]$ for some $b>0$, $E_{\boldsymbol{P}}(X_i)^{-1}<\infty$, Assumption 4.2.1 holds with $0<\lambda<E_{\boldsymbol{P}}(X_i)/b$, and Assumptions 4.2.4, 4.2.5 and 4.2.6 hold. Then, we have the following results.*

(i) The asymptotic variance of the estimator of $\int_{[\alpha,\beta]} Q_{y,N}(p)J(p)dp$ based on $G(p)$ under SRSWOR is smaller than its asymptotic variance under RHC as well as any HEπPS sampling design, which uses auxiliary information, if and only if

$$\max_{2\leq i\leq 3}\left\{\int_{\alpha}^{\beta}\int_{\alpha}^{\beta}\big(K_1^*(p_1,p_2)-K_i^*(p_1,p_2)\big)J(p_1)J(p_2)dp_1dp_2\right\}<0. \qquad (4.5.5)$$

(ii) The asymptotic variance of the estimator of $f(Q_{y,N}(p_1),\cdots,Q_{y,N}(p_k))$ based on $G(p)$ under SRSWOR is smaller than its asymptotic variance under RHC as well as any HEπPS sampling design if and only if

$$\max_{2\le i\le 3} a(\Delta_1^* - \Delta_i^*)a^T < 0, \tag{4.5.6}$$

where a=$\nabla f(Q_y(p_1),\cdots,Q_y(p_k))$ is the gradient of f at $\big(Q_y(p_1),\cdots,Q_y(p_k)\big)$.

The conditions that $X_i \le b$ *a.s.* $[\mathbf{P}]$ for some $b>0$, and $0<\lambda<E_{\mathbf{P}}(X_i)/b$ are discussed in Chapter 2 (see the discussion related to Assumption 2.2.1 in Chapter 2). Now, we consider some examples where the conditions (4.5.2) and (4.5.3) hold, and some examples where these conditions fail to hold. Suppose that Y_i's have a normal distribution with mean $\mu\in\{-10+j\}_{j=0}^{20}$ and s.d. σ=1, X_i=e^{Y_i} for i=1, $\ldots,N$, and λ=0.05. Then, the conditions (4.5.2) and (4.5.3) are discussed in Table 4.7 below in the cases of various finite population parameters and sampling designs. Next, we consider some examples where the conditions (4.5.5) and (4.5.6) hold, and some examples where these conditions fail to hold. Suppose that Y_i's have a normal distribution with mean μ=10 and s.d. $\sigma\in\{j/100\}_{j=1}^{10}\cup\{j/10\}_{j=1}^{20}$, X_i=e^{Y_i} for i=1, $\ldots,N$, and λ=0.05. Then, the conditions (4.5.5) and (4.5.6) are discussed in Table 4.8 below in the cases of various finite population parameters and their estimators. The above conditions depend on superpopulation quantiles, moments and densities. In practice, one can check these conditions by estimating the above-mentioned superpopulation parameters (see Table 4.5 in the preceding section) based on a pilot survey.

Theorem 4.5.1 shows that in the case of the estimation of $\int_{[\alpha,\beta]} Q_{y,N}(p)J(p)dp$ and $f(Q_{y,N}(p_1),\cdots,Q_{y,N}(p_k))$, the use of the auxiliary information in the estimation stage may have an adverse effect on the performances of their estimators based on the ratio, the difference and the regression estimators under each of SRSWOR, RHC and any HEπPS sampling designs. This is in striking contrast to the case of the estimation of the finite population mean, where the use of the auxiliary information in the estimation stage improves the performance of the GREG estimator under these sampling designs (see Chapaters 2 and 3). On the other hand, Theorem 4.5.2 implies that the performance of each of the estimators of the above parameters considered in this chapter may become worse under RHC and any HEπPS sampling deigns, which use the auxiliary information, than their performances under SRSWOR.

TABLE 4.7: Discussion of the conditions (4.5.2) and (4.5.3).

		Sampling design		
Parameter	The condition	SRSWOR	RHC	HEπPS
Median	(4.5.3) holds for	$\mu \leq -2$ & $\mu \geq 8$	$\mu \leq -2$ & $\mu \geq 8$	$\mu \leq -2$ & $\mu \geq 8$
	(4.5.3) does not hold for	$-1 \leq \mu \leq 7$	$-1 \leq \mu \leq 7$	$-1 \leq \mu \leq 7$
α-trimmed mean with α=0.1	(4.5.2) holds for	μ=1	μ=1	μ=1
	(4.5.2) does not hold for	$\mu \neq 1$	$\mu \neq 1$	$\mu \neq 1$
α-trimmed mean with α=0.3	(4.5.2) holds for	$\mu \leq -2$ & $\mu \geq 8$	$\mu \leq -2$ & $\mu \geq 9$	$\mu \leq -2$ & $\mu \geq 9$
	(4.5.2) does not hold for	$-1 \leq \mu \leq 7$	$-1 \leq \mu \leq 8$	$-1 \leq \mu \leq 8$
Inter-quartile range	(4.5.3) holds for	$\mu \leq -2$ & $\mu \geq 4$	$\mu \leq -2$ & $\mu \geq 4$	$\mu \leq -2$ & $\mu \geq 4$
	(4.5.3) does not hold for	$-1 \leq \mu \leq 3$	$-1 \leq \mu \leq 3$	$-1 \leq \mu \leq 3$
Bowley's measure of skewness	(4.5.3) holds for	$\mu \leq -2$ & $\mu \geq 5$	$\mu \leq -2$ & $\mu \geq 7$	$\mu \leq -2$ & $\mu \geq 7$
	(4.5.3) does not hold for	$-1 \leq \mu \leq 4$	$-1 \leq \mu \leq 6$	$-1 \leq \mu \leq 6$

TABLE 4.8: Discussion of the conditions (4.5.5) and (4.5.6).

		Estimator based on			
Parameter	The condition	$\hat{Q}_y(p)$	$\hat{Q}_{y,RA}(p)$	$\hat{Q}_{y,DI}(p)$	$\hat{Q}_{y,REG}(p)$
Median	(4.5.6) holds for	$\sigma \geq 0.2$	$\sigma \geq 0.2$	$\sigma \geq 0.2$	$\sigma \geq 0.2$
	(4.5.6) does not hold for	$\sigma \leq 0.1$	$\sigma \leq 0.1$	$\sigma \leq 0.1$	$\sigma \leq 0.1$
α-trimmed mean with α=0.1	(4.5.5) holds for	$\sigma \geq 1.2$	$\sigma \geq 1.3$	$\sigma \geq 1.6$	$\sigma \geq 1.2$
	(4.5.5) does not hold for	$\sigma \leq 1.1$	$\sigma \leq 1.2$	$\sigma \leq 1.5$	$\sigma \leq 1.1$
α-trimmed mean with α=0.3	(4.5.5) holds for	$\sigma \geq 0.2$	$\sigma \geq 0.2$	$\sigma \geq 0.2$	$\sigma \geq 0.2$
	(4.5.5) does not hold for	$\sigma \leq 0.1$	$\sigma \leq 0.1$	$\sigma \leq 0.1$	$\sigma \leq 0.1$
Inter-quartile range	(4.5.6) holds for	$\sigma \geq 0.06$	$\sigma \geq 0.06$	$\sigma \geq 1.1$	$\sigma \geq 1$
	(4.5.6) does not hold for	$\sigma \leq 0.05$	$\sigma \leq 0.05$	$\sigma \leq 1$	$\sigma \leq 0.9$
Bowley's measure of skewness	(4.5.6) holds for	$\sigma \geq 0.03$	$\sigma \geq 0.03$	$\sigma \geq 0.1$	$0.1 \leq \sigma \leq 0.6$ & $\sigma \geq 1.2$
	(4.5.6) does not hold for	$\sigma \leq 0.02$	$\sigma \leq 0.02$	$\sigma \leq 0.09$	$\sigma \leq 0.09$ & $0.7 \leq \sigma \leq 1.1$

4.5.1.1 Comparison of the estimators of quantile based location, spread and skewness

It follows from Theorem 4.5.1 that in the cases of the median, the interquartile range and the Bowley's measure of skewness, the estimator based on the sample median becomes more efficient than the estimators based on the ratio, the difference and the regression estimators of the finite population quantile under $P(s,\omega)$ if and only if (4.5.3) holds with k, $p_1,\ldots,p_k$ and a as in Table 4.9 below. Here, $P(s,\omega)$ is one of SRSWOR, RHC and any HEπPS sampling designs. On the other hand, it follows from Theorem 4.5.2 that in the cases of the above parameters the performance of the estimator based on $G(p)$ becomes worse under RHC and any HEπPS sampling deigns, which use the auxiliary information, than its performance under SRSWOR if and only if (4.5.6) holds with k, $p_1, \ldots, p_k$ and a as in Table 4.9 below. Here, $G(p)$ is one of $\hat{Q}_y(p)$, $\hat{Q}_{y,RA}(p)$, $\hat{Q}_{y,DI}(p)$ and $\hat{Q}_{y,REG}(P)$.

TABLE 4.9: k, $p_1, \ldots, p_k$ and a in (4.5.3) and (4.5.6) for different parameters.

Parameter	k	$p_1, \ldots, p_k$	a
Median	1	p_1=0.5	1
Interquartile range	2	p_1=0.25, p_2=0.75	$(-1, 1)$
Bowley's measure of skewness	3	p_1=0.25, p_2=0.5, p_3=0.75	$2\big(Q_y(p_3) - Q_y(p_2), Q_y(p_1) - Q_y(p_3), Q_y(p_2) - Q_y(p_1)\big)/\big(Q_y(p_3) - Q_y(p_1)\big)^2$

4.5.2 Comparison of the sample mean, the sample median and the GREG estimator

Here, we compare the GREG estimator, say $\hat{\overline{Y}}_{GREG}$ (see [24] and references therein), of the finite population mean $\overline{Y}$=$\sum_{i=1}^{N} Y_i/N$, the sample mean $\overline{y}$=$\sum_{i\in s} Y_i/n$, and the sample median $\hat{Q}_y(0.5)$ under SRSWOR in terms of asymptotic variances of $\sqrt{n}(\hat{\overline{Y}}_{GREG} - E_{\mathbf{P}}(Y_i))$, $\sqrt{n}(\overline{y} - E_{\mathbf{P}}(Y_i))$ and $\sqrt{n}(\hat{Q}_y(0.5) - Q_y(0.5))$, when the superpopulation median $Q_y(0.5)$ and the superpopulation mean $E_{\mathbf{P}}(Y_i)$ are same.

Theorem 4.5.3. *Suppose that $Q_y(0.5)$=$E_{\mathbf{P}}(Y_i)$, and Assumptions 4.2.1 and 4.2.3 hold. Then, under SRSWOR, the asymptotic variance of the sample median is smaller than that of the sample mean and the asymptotic variance of the GREG estimator of the mean is smaller than that of the sample median if and only if*

$$\sigma_y^2 > 1/4\sigma_y^2 f_y^2(Q_y(0.5)), \text{ and} \tag{4.5.7}$$

$$\rho_{xy}^2 > (1-\lambda)^{-1}(1 - 1/4\sigma_y^2 f_y^2(Q_y(0.5))), \tag{4.5.8}$$

respectively. Here, σ_y^2 is the superpopulation variance of y, and ρ_{xy} is the superpopulation correlation coefficient between x and y.

The conditions (4.5.7) and (4.5.8) are discussed below for different superpopulation distributions of Y_i's and X_i's, and different values of λ (see Tables 4.10 and 4.11 below). As mentioned in the cases of (4.5.2), (4.5.3), (4.5.5) and (4.5.6) in the preceding section, the conditions (4.5.7) and (4.5.8) can also be checked using a pilot survey.

TABLE 4.10: Discussion of the condition (4.5.7).

Superpopulation distribution of Y_i's	The condition (4.5.7) holds iff
Exponential power distribution with location $\mu \in \mathbb{R}$, scale $\sigma > 0$ and shape $\alpha > 0$	[3] $\alpha^2\Gamma(3/\alpha) > \Gamma^3(1/\alpha)$
Student's t-distribution with degrees of freedom $m > 2$	[3] $4\Gamma^2((m+1)/2) > (m-2)\pi\Gamma^2(m/2)$

[3] Here, $\Gamma(\cdot)$ denotes the gamma function.

TABLE 4.11: Discussion of the condition (4.5.8).

Superpopulation distribution of Y_i's	Superpopulation distribution of X_i's	λ
Normal distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$	Any distribution supported on $(0, \infty)$	The condition (4.5.8) holds for any $\lambda \in (0, 1)$
Standard Laplace distribution	X_i=$\max\{Y_i, 0\}$ for i=$1, \ldots, N$	The condition (4.5.8) holds iff $\lambda \in (0, 0.25)$

Theorem 4.5.3 implies that under SRSWOR, the performance of the sample mean is worse than that of the sample median and the performance of the sample median is worse than that of the GREG estimator if and only if (4.5.7) and (4.5.8) hold, respectively. In the case of a finite population, if the population observations on y are generated from heavy-tailed distributions (e.g., exponential power, student's t, etc.) and SRSWOR is used, the sample median becomes more efficient than the sample mean. It is well-known that a similar result holds in the classical set up involving i.i.d. sample observations. However, the GREG estimator of the mean becomes more efficient than the sample median under SRSWOR, whenever y and x are highly correlated. This is in striking contrast to what happens in the case of i.i.d. observations.

4.6. Demonstration using real data

In this section, we use the data on irrigated land area for the state of West Bengal in India from the District Census Handbook (2011) available in Office of the Registrar General and Census Commissioner, India (*https://censusindia.gov.in/nada/index.php/catalog/1362.*). In West Bengal, lands are irrigated by different sources like canals, wells, waterfalls, lakes, etc., and irrigated land area (in Hectares) in different villages are reported in the above data set. We consider the population of 14224 villages having lands irrigated by canals in this state. We use this data set to demonstrate the accuracy of the asymptotic normal approximations for the distributions of several estimators of several parameters under single stage sampling designs like SRSWOR, LMS and RHC sampling designs.

We use the same data set to demonstrate the accuracy of the asymptotic approximations for the distributions of different estimators of different parameters under stratified multistage cluster sampling design with SRSWOR. Note that the above-mentioned population can be divided into 18 districts, and every district can further be divided into sub districts consisting of villages. We consider districts as strata, sub districts as clusters and villages as population units. Boxplots of number of clusters, number of population units, 4^{th} order moments of cluster sizes, and variance of cluster sizes in different strata are given in Figure 4.1 below. Descriptive statistics of number of clusters, number of population units, 4^{th} order moments of cluster sizes and variances of cluster sizes are given in Table 4.12 below.

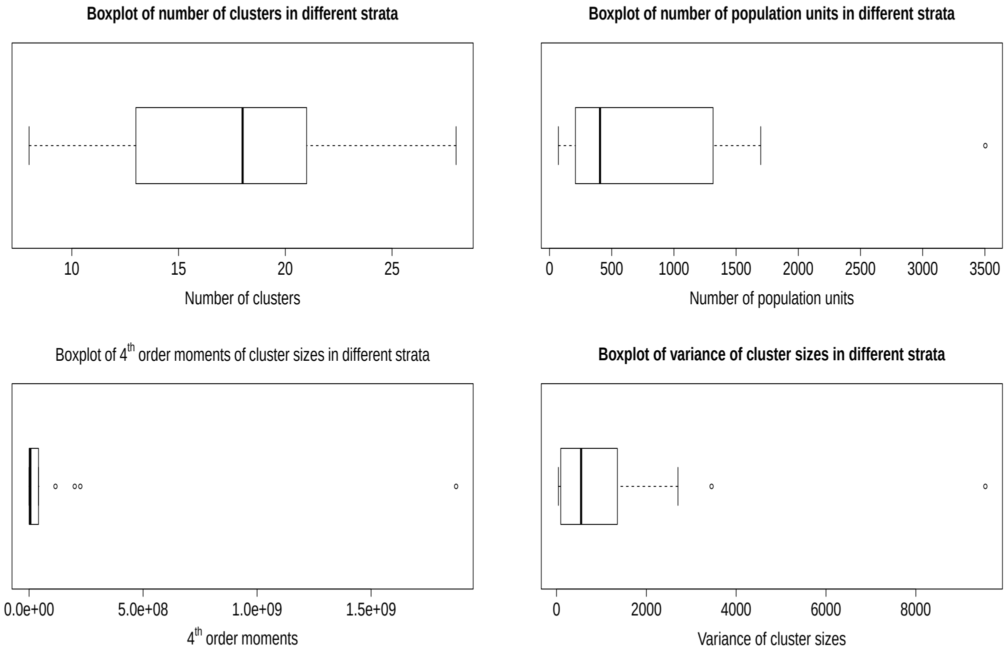

FIGURE 4.1: Boxplots of number of clusters, number of population units, maximum cluster sizes, and variance of cluster sizes in different strata.

We choose the land area irrigated by canals as the study variable y, and the total irrigated land area as the auxiliary variable x. We are interested in the estimation of the median and the α-trimmed means of y, where α=0.1 and 0.3. We are also interested in the estimation of the interquartile range and the Bowley's measure of skewness of y. For each of the aforementioned

TABLE 4.12: Descriptive statistics of number of clusters, number of population units, 4^{th} order moments of cluster sizes and variances of cluster sizes.

	1^{st} quartile	Median	3^{rd} quartile
Number of clusters	13	18	21
Number of population units	208.5	406	1252
4^{th} order moment of cluster sizes	391394.8	5414937	37339619
Variance of cluster sizes	114.33	547.91	1269.26

parameters, we compute relative biases of the estimators, which are based on the sample quantile, and the ratio, the difference and the regression estimators of the population quantile. We consider I=1000 samples each of size n=200 and n=500 selected using single stage sampling designs mentioned in the first paragraph of this section. Further, we consider I=1000 samples each of size n=108 (a sample of 6 clusters from each stratum and a sample of 1 village from each selected cluster) and n=216 (a sample of 6 clusters from each stratum and a sample of 2 villages from each selected cluster) selected using stratified multistage cluster sampling design with SRSWOR. Suppose that $\hat{\theta}$ is an estimator of the parameter θ, and $\hat{\theta}_k$ is the estimate of θ computed based on the k^{th} sample using a sampling design $P(s)$ for k=$1,\ldots,I$. The relative bias of $\hat{\theta}$ under $P(s)$ (cf. [7]) is computed as

$$RB(\hat{\theta}, P) = \sum_{k=1}^{I}(\hat{\theta}_k - \theta_0)/I\theta_0, \tag{4.6.1}$$

where θ_0 is the true value of θ in the population. Note that θ_0 is known because we have all the population values available for y and x in the above-mentioned dataset. We use the R software for drawing samples as well as computing estimators. For sample quantiles, we use weighted.quantile function in R. The plots of relative biases for different parameters, estimators, sampling designs and sample sizes are presented in Figures 4.2–4.9 below. Also, boxplots of relative biases for different parameters and estimators in the cases of single stage sampling designs and stratified multistage cluster sampling design with SRSWOR are given in Figure 4.10 below.

Next, we compute asymptotic MSEs of the estimators following the procedure described below. Recall from Section 4.4 the expressions of the asymptotic covariance kernels $K(p_1, p_2)$ of several quantile processes considered in this chapter. Note that $K(p_1, p_2)$=$\lim_{\nu\to\infty} E_{\mathbf{P}}(\sigma_1(p_1, p_2))$ for $d(i, s)$=$(N\pi_i)^{-1}$ under high entropy sampling designs, $K(p_1, p_2)$=$\lim_{\nu\to\infty} E_{\mathbf{P}}(\sigma_2(p_1, p_2))$ for $d(i, s)$=$(NX_i)^{-1}G_i$ under RHC sampling design, and $K(p_1, p_2)$=$\lim_{\nu\to\infty}(n/N^2)\sum_{h=1}^{H} N_h(N_h - n_h)\sigma_h(p_1, p_2)/n_h$ for $d(i, s)$=$(N\pi_i)^{-1}$ under stratified multistage cluster sampling design with

SRSWOR, where

$$\begin{aligned}
&\sigma_1(p_1,p_2) = (n/N^2)\sum_{i=1}^{N}(\zeta_i(p_1)-\overline{\zeta}(p_1)-S(p_1)\pi_i)(\zeta_i(p_2)-\overline{\zeta}(p_2)-S(p_2)\pi_i)\times \\
&(\pi_i^{-1}-1), \\
&\sigma_2(p_1,p_2) = (n\gamma)(\overline{X}/N)\sum_{i=1}^{N}(\zeta_i(p_1)-\overline{\zeta}(p_1))(\zeta_i(p_2)-\overline{\zeta}(p_2))/X_i, \text{ and} \\
&\sigma_h(p_1,p_2) = E_{\mathbf{P}}(\zeta'_{hjl}(p_1)-E_{\mathbf{P}}(\zeta'_{hjl}(p_1)))(\zeta'_{hjl}(p_2)-E_{\mathbf{P}}(\zeta'_{hjl}(p_2)))
\end{aligned} \tag{4.6.2}$$

for h=$1,\ldots,H$. Here, $\zeta_i(p)$'s, $\zeta'_{hjl}(p)$'s, $\overline{\zeta}(p)$, $S(p)$ and γ are as in Sections 4.3 and 4.4, and N_h and n_h are as in the paragraph preceding Assumption 4.3.1 in Section 4.3. Note that $\zeta_i(p)$'s in $\sigma_1(p_1,p_2)$ and $\sigma_2(p_1,p_2)$ involve superpopulation parameters like $E_{\mathbf{P}}(X_i)$, $E_{\mathbf{P}}(Y_i)$, $E_{\mathbf{P}}(X_iY_i)$, $E_{\mathbf{P}}(X_i^2)$, $f_y(Q_y(p))$ and $f_x(Q_x(p))$ (see Table 4.1). We approximate $E_{\mathbf{P}}(X_i)$, $E_{\mathbf{P}}(Y_i)$, $E_{\mathbf{P}}(X_iY_i)$ and $E_{\mathbf{P}}(X_i^2)$ by their finite population versions $\overline{X}$, $\overline{Y}$, $\sum_{i=1}^{N}X_iY_i/N$ and $\sum_{i=1}^{N}X_i^2/N$, respectively. We also approximate $1/f_y(Q_y(p))$ and $1/f_x(Q_x(p))$ by

$$\begin{aligned}
&\sqrt{N}(Q_{y,N}(p+1/\sqrt{N})-Q_{y,N}(p-1/\sqrt{N}))/2 \text{ and} \\
&\sqrt{N}(Q_{x,N}(p+1/\sqrt{N})-Q_{x,N}(p-1/\sqrt{N}))/2,
\end{aligned} \tag{4.6.3}$$

respectively, following the ideas in [77]. Next, we approximate the superpopulation covariance $\sigma_h(p_1,p_2)$ between $\zeta'_{hjl}(p_1)$ and $\zeta'_{hjl}(p_2)$ by

$$\sum_{j=1}^{M_h}\sum_{l=1}^{N_{hj}}(\zeta'_{hjl}(p_1)-\overline{\zeta}'_h(p_1))(\zeta'_{hjl}(p_2)-\overline{\zeta}'_h(p_2))/N_h, \tag{4.6.4}$$

where $\overline{\zeta}'_h(p)$=$\sum_{j=1}^{M_h}\sum_{l=1}^{N_{hj}}\zeta'_{hjl}(p)/N_h$. Further, we approximate $\sum_{h=1}^{H}(N_h/N)E_{\mathbf{P}}(X'_{hjl})$ (as well as Θ_1), $\sum_{h=1}^{H}(N_h/N)E_{\mathbf{P}}(Y'_{hjl})$ (as well as Θ_2), $\sum_{h=1}^{H}(N_h/N)\ E_{\mathbf{P}}(X'_{hjl}Y'_{hjl})$ (as well as Θ_3), $\sum_{h=1}^{H}(N_h/N)E_{\mathbf{P}}\ (X'_{hjl})^2$ (as well as Θ_4), $1/f_{y,H}(Q_{y,H}(p))$ and $1/f_{x,H}(Q_{x,H}(p))$ involved in the expressions of $\zeta'_{hjl}(p)$'s (see Table 4.3) in the same way as we approximate $E_{\mathbf{P}}(X_i)$, $E_{\mathbf{P}}(Y_i)$, $E_{\mathbf{P}}(X_iY_i)$, $E_{\mathbf{P}}(X_i^2)$, $f_y(Q_y(p))$ and $f_x(Q_x(p))$ in the case of single stage sampling designs. Let $\tilde{\sigma}_1(p_1,p_2)$, $\tilde{\sigma}_2(p_1,p_2)$ and $\tilde{\sigma}_h(p_1,p_2)$ denote the approximated $\sigma_1(p_1,p_2)$, $\sigma_2(p_1,p_2)$ and $\sigma_h(p_1,p_2)$, respectively. Then, asymptotic MSEs of several estimators of the parameters considered in this section are computed by replacing $K(p_1,p_2)$ in the expressions of σ_1^2 and σ_2^2 (see Theorem 4.4.1) by $\tilde{\sigma}_1(p_1,p_2)/n$, $\tilde{\sigma}_2(p_1,p_2)/n$ and $(1/N^2)\sum_{h=1}^{H}N_h(N_h-n_h)\tilde{\sigma}_h(p_1,p_2)/n_h$. We approximate the double integral in the expression of σ_1^2 by sum after dividing $[\alpha,1-\alpha]$ into 100 sub intervals of equal width.

Based on the asymptotic MSE, we compute the bias relative to the standard error of the single sample estimates for the estimator $\hat{\theta}$ of θ under a sampling design $P(s)$ as

$$I^{-1}\sum_{k=1}^{I}(\hat{\theta}_k-\theta_0)\Big/\big(nAMSE(\hat{\theta})\big)^{1/2}, \tag{4.6.5}$$

where $AMSE(\hat{\theta})$ denotes the asymptotic MSE of $\hat{\theta}$ under $P(s)$, and $(nAMSE(\hat{\theta}))^{1/2}$ denotes the standard error of the single sample estimates. The plots of ratios of biases and $(n$ asymptotic MSE$)^{1/2}$'s for different parameters, estimators, sampling designs and sample sizes are presented in Figures 4.11–4.18 below. Also, boxplots of ratios of biases and $(n$ asymptotic MSE$)^{1/2}$'s for different parameters and estimators in the cases of single stage sampling designs and stratified multistage cluster sampling design with SRSWOR are given in Figure 4.19 below.

Next, we compute ratios of asymptotic and true MSEs for different parameters, estimators and sampling designs considered in this section. The true MSE of an estimator $\hat{\theta}$ of θ under a sampling design $P(s)$ is estimated as

$$MSE(\hat{\theta},P)=\sum_{k=1}^{I}(\hat{\theta}_k-\theta_0)^2/I, \tag{4.6.6}$$

where θ_0 is the true value of θ, and $\hat{\theta}_k$ is the estimate of θ computed based on the k^{th} sample using the sampling design $P(s)$ for k=1, $\ldots, I$. The plots of ratios of asymptotic and true MSEs for different parameters, estimators, sampling designs and sample sizes are presented in Figures 4.20–4.27 below. Also, boxplots of ratios of asymptotic and true MSEs for different parameters and estimators in the cases of single stage sampling designs and stratified multistage cluster sampling design with SRSWOR are given in Figure 4.28 below.

Finally, we compute coverage probabilities of nominal 90% and 95% confidence intervals (see Section 4.4.1) of the parameters discussed in this section. While computing coverage probabilities, we consider the estimators

$$\hat{\sigma}_1^2=\int_{\alpha}^{\beta}\int_{\alpha}^{\beta}\hat{K}(p_1,p_2)J(p_1)J(p_2)dp_1dp_2 \text{ and } \hat{\sigma}_2^2=\hat{a}\hat{\Delta}\hat{a}^T \tag{4.6.7}$$

discussed in the paragraph preceding Theorem 4.4.3. We compute coverage probabilities of nominal 90% and 95% confidence intervals of a parameter by taking the proportion of times confidence intervals constructed based on I=1000 samples include the true value of the parameter. We also compute the magnitude of the Monte Carlo standard errors of these coverage probabilities.

The plots of observed coverage probabilities of nominal 90% and 95% confidence intervals for different parameters, estimators, sampling designs and sample sizes are presented in Figures 4.29–4.44 below. Also, boxplots of observed coverage probabilities of nominal 90% and 95% confidence intervals for different parameters and estimators in the cases of single stage sampling designs and stratified multistage cluster sampling design with SRSWOR are given in Figures 4.45 and 4.46 below.

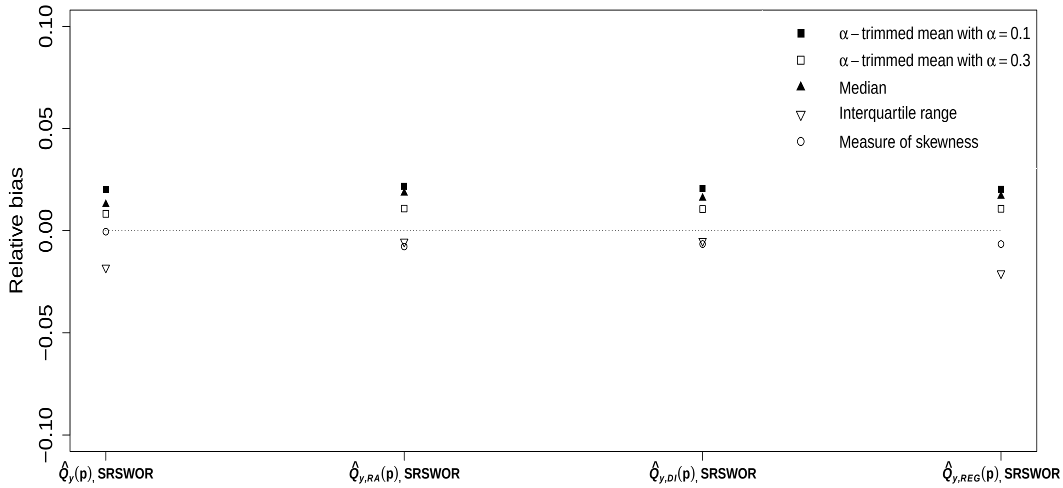

FIGURE 4.2: Relative biases of different estimators for n=500 in the case of SRSWOR.

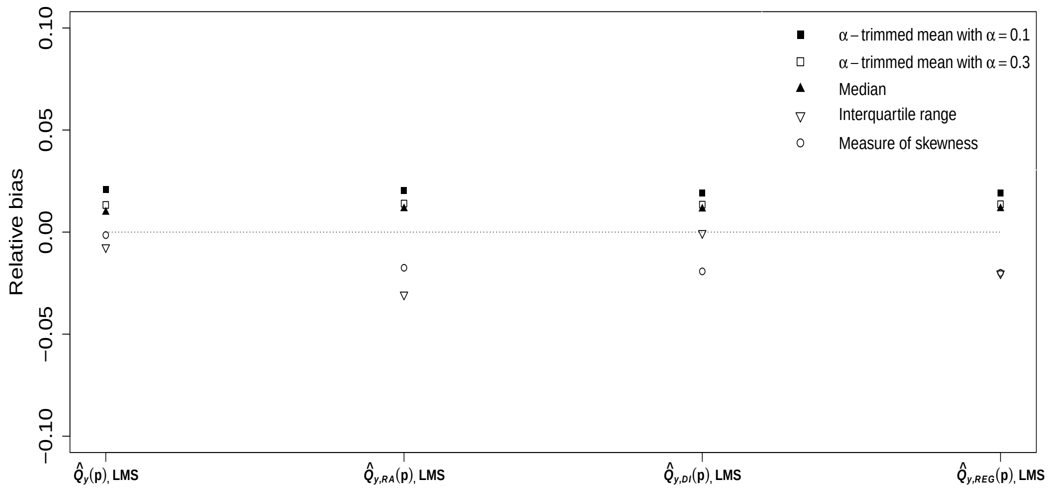

FIGURE 4.3: Relative biases of different estimators for n=500 in the case of LMS sampling design.

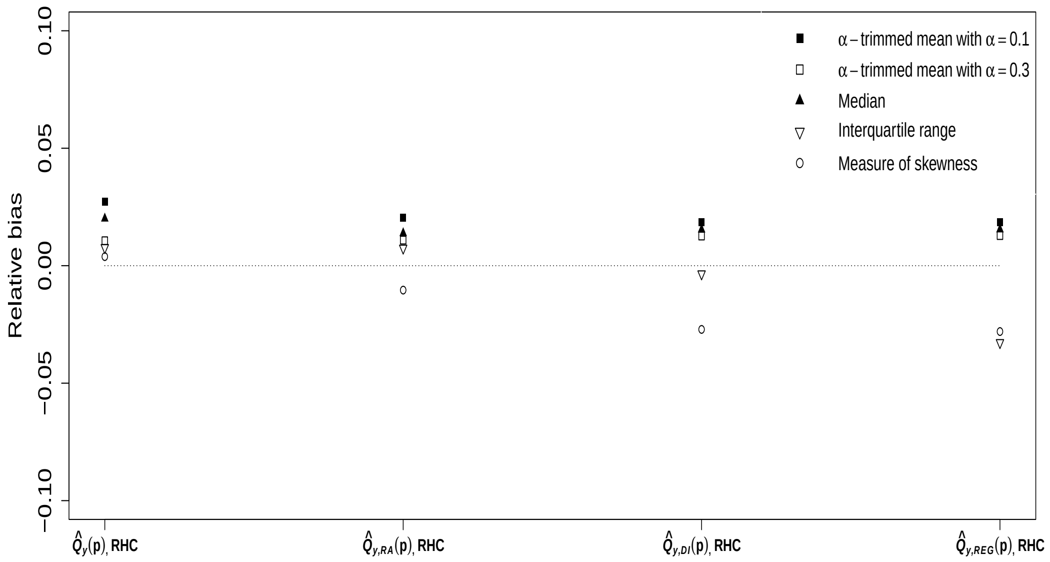

FIGURE 4.4: Relative biases of different estimators for n=500 in the case of RHC sampling design.

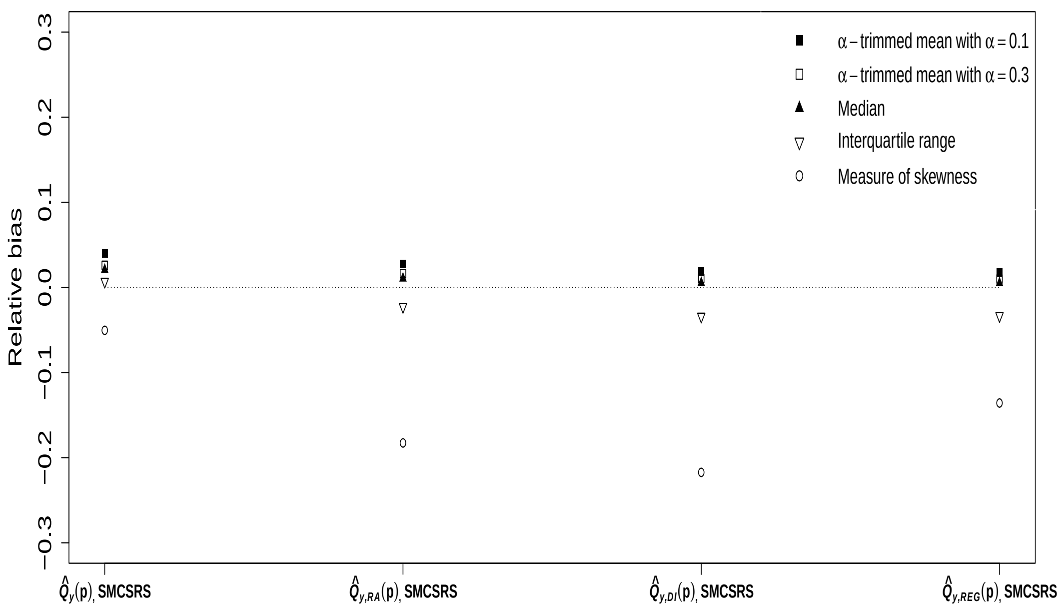

FIGURE 4.5: Relative biases of different estimators for n=216 in the case of SMCSRS. In this figure, SMCSRS stands for stratified multistage cluster sampling design with SRSWOR.

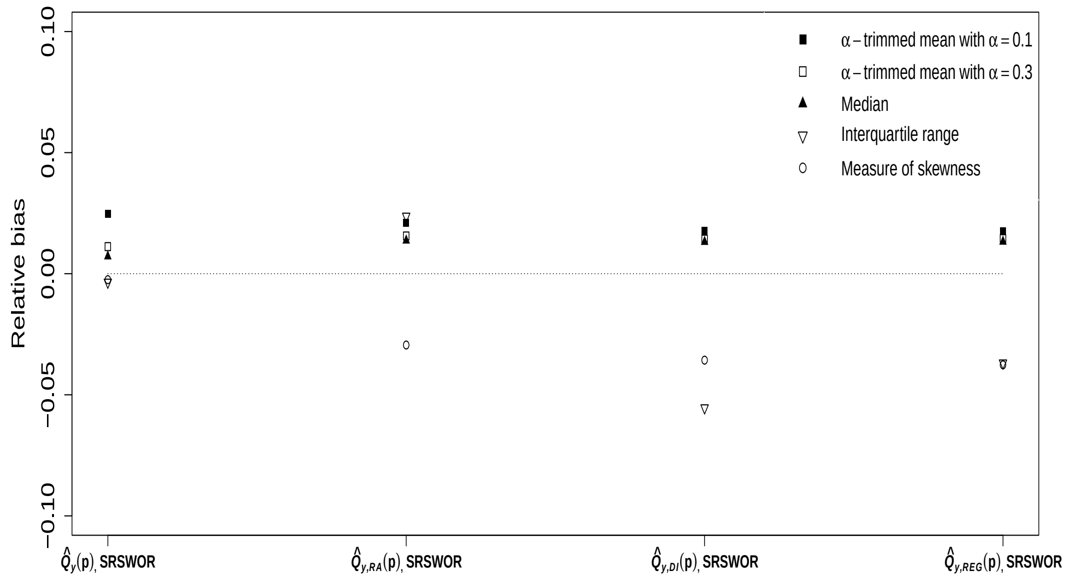

FIGURE 4.6: Relative biases of different estimators for n=200 in the case of SRSWOR.

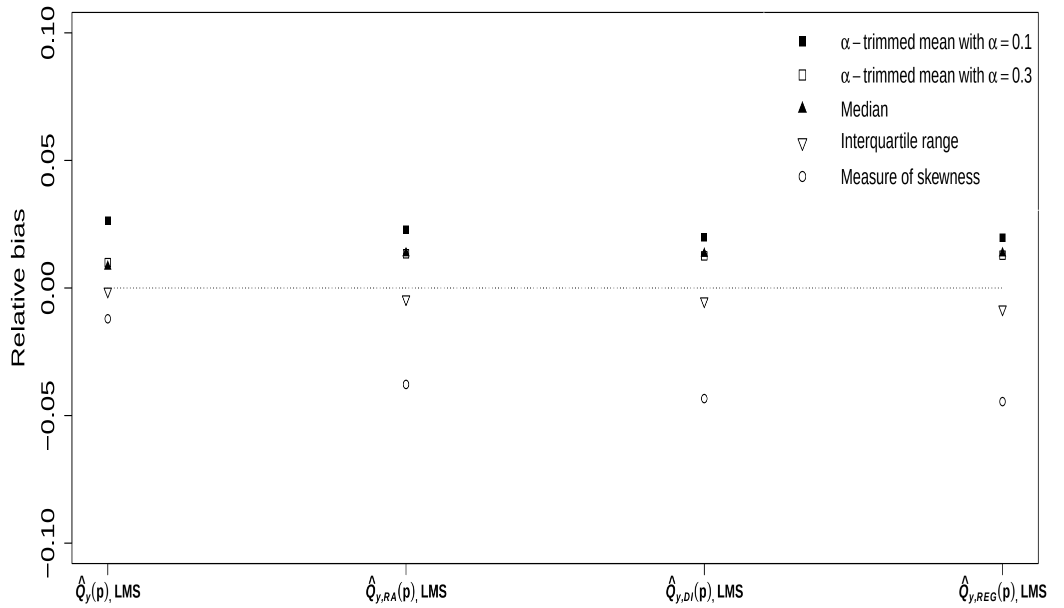

FIGURE 4.7: Relative biases of different estimators for n=200 in the case of LMS sampling design.

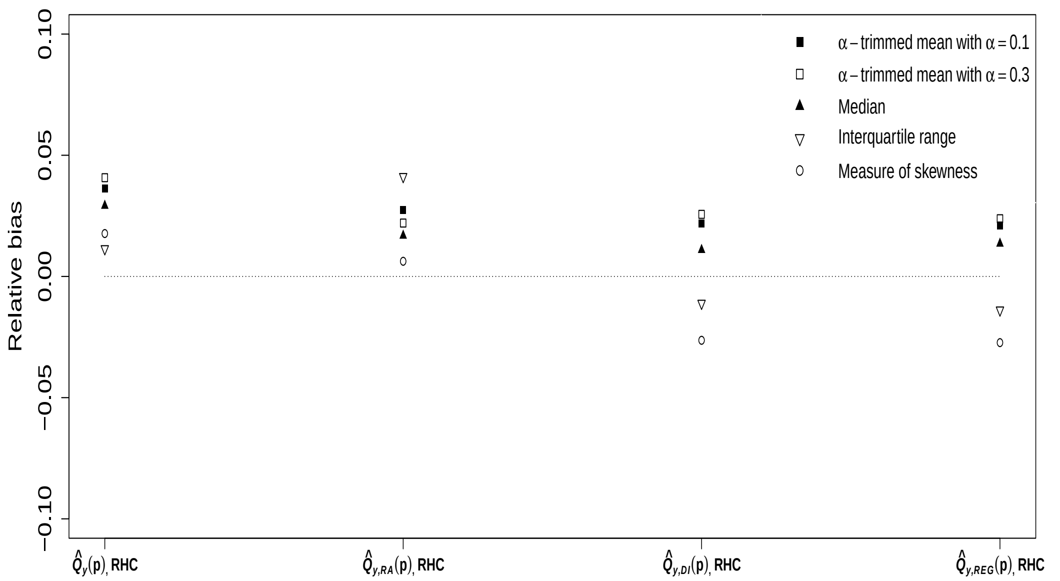

FIGURE 4.8: Relative biases of different estimators for n=200 in the case of RHC sampling design.

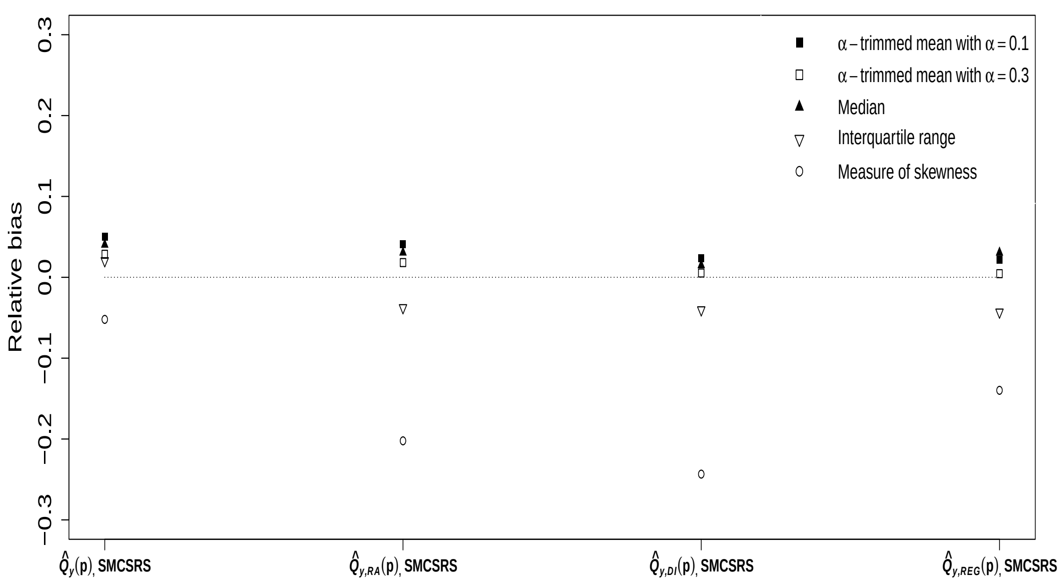

FIGURE 4.9: Relative biases of different estimators for n=108 in the case of SMCSRS. In this figure, SMCSRS stands for stratified multistage cluster sampling design with SRSWOR.

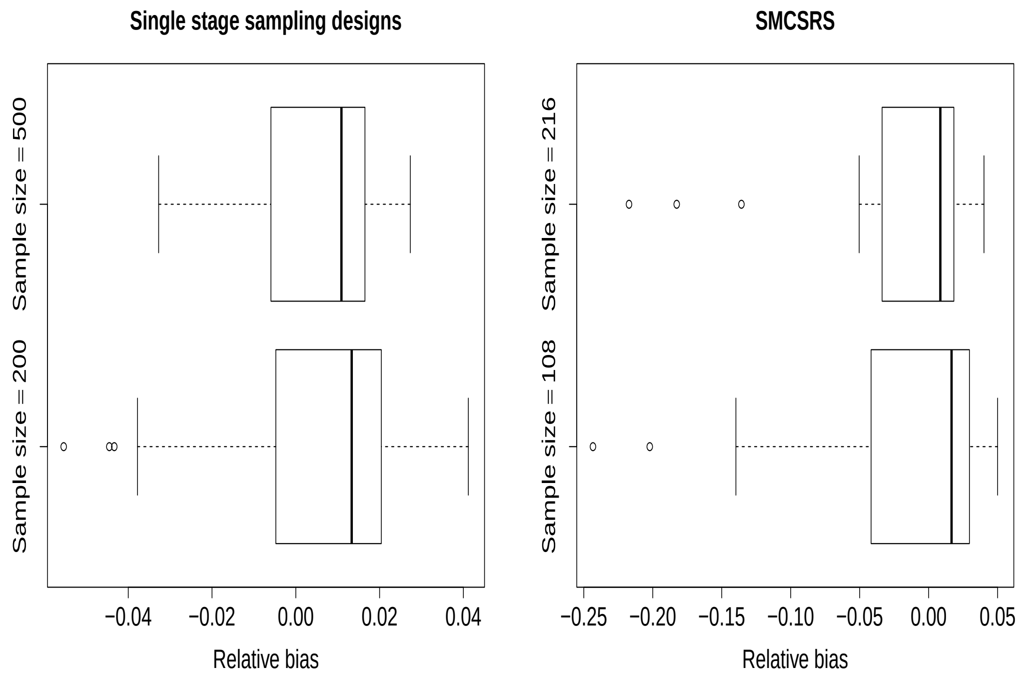

FIGURE 4.10: Boxplots of relative biases for different parameters and estimators in the cases of single stage sampling designs and SMCSRS. In this figure, SMCSRS stands for stratified multistage cluster sampling design with SRSWOR.

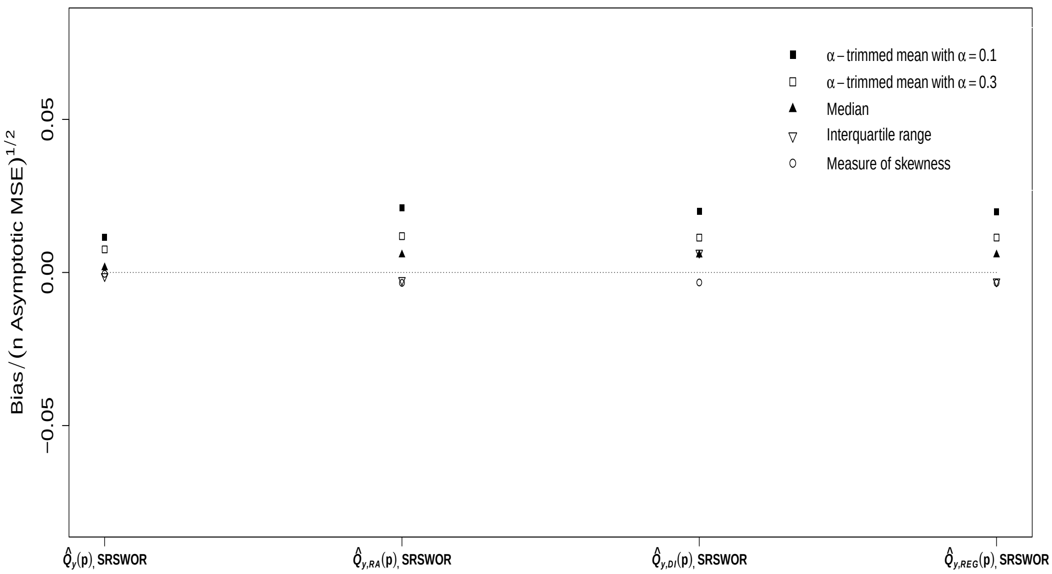

FIGURE 4.11: Ratios of biases and (n asymptotic MSE)$^{1/2}$'s for different estimators under SRSWOR in the case of n=500.

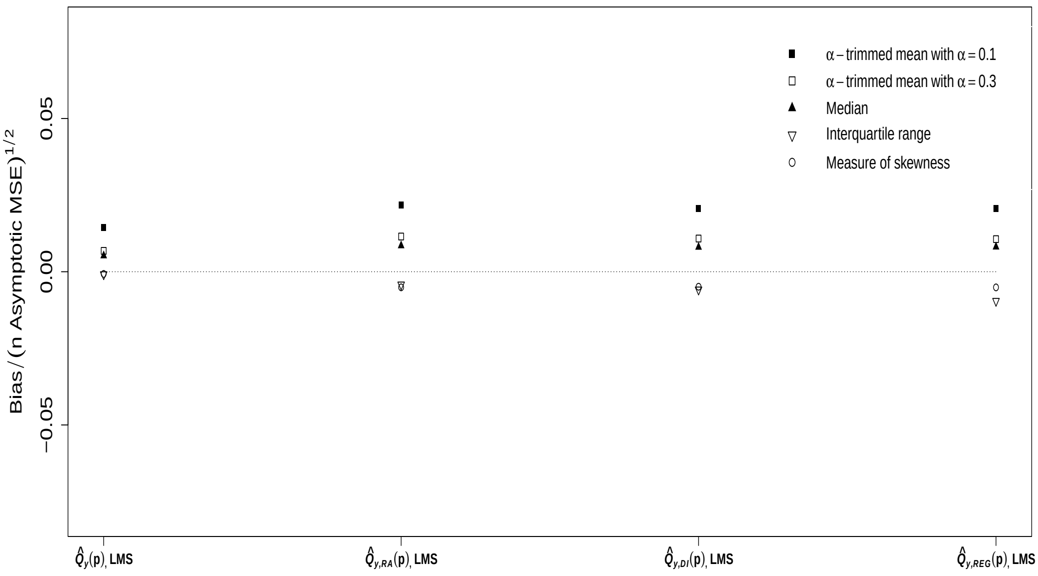

FIGURE 4.12: Ratios of biases and $(n \text{ asymptotic MSE})^{1/2}$'s for different estimators under LMS sampling design in the case of n=500.

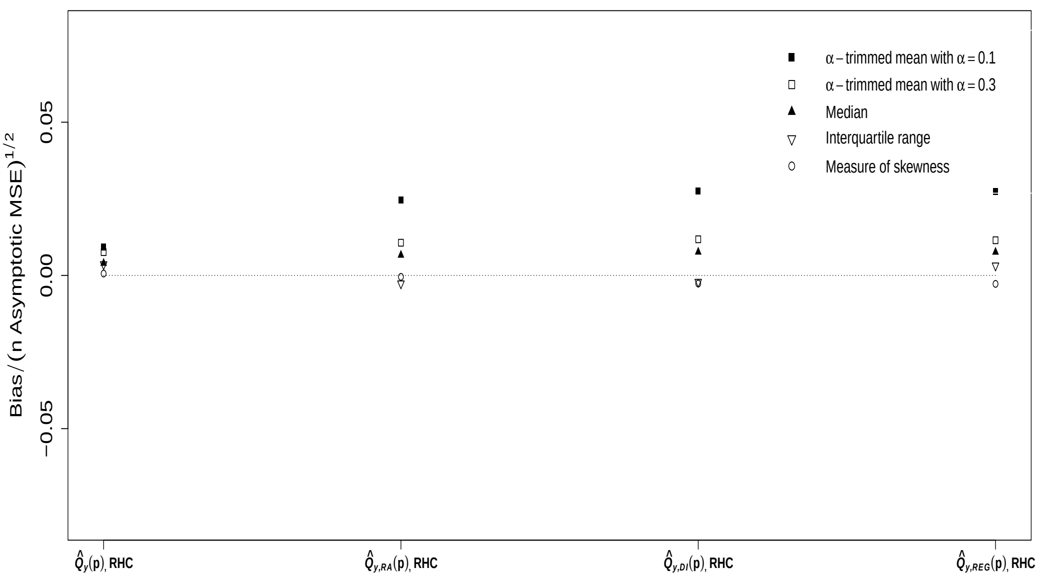

FIGURE 4.13: Ratios of biases and $(n \text{ asymptotic MSE})^{1/2}$'s for different estimators under RHC sampling design in the case of n=500.

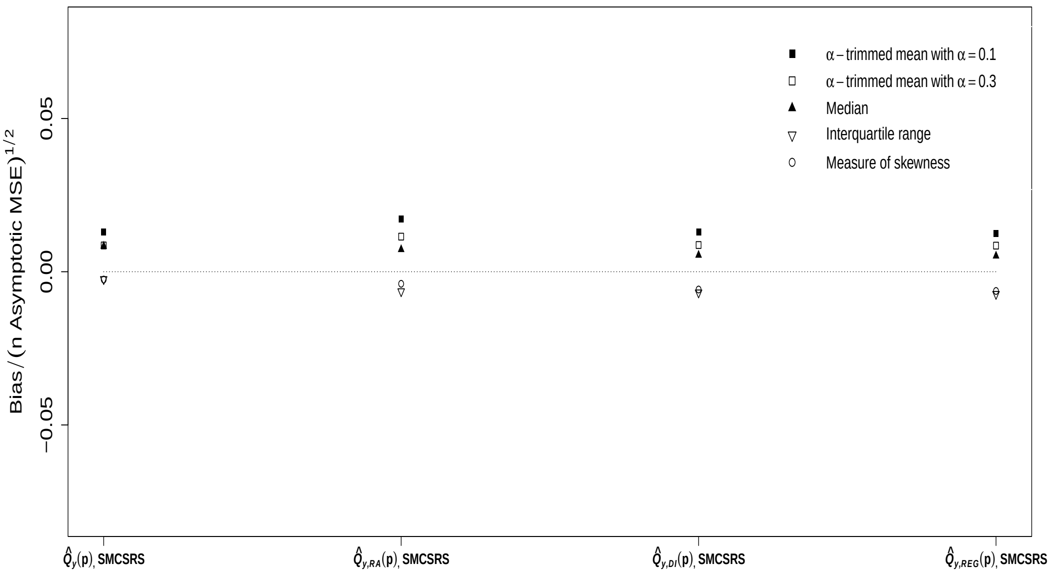

FIGURE 4.14: Ratios of biases and (n asymptotic MSE)$^{1/2}$'s for different estimators under SMCSRS in the case of n=216. In this figure, SMCSRS stands for stratified multistage cluster sampling design with SRSWOR.

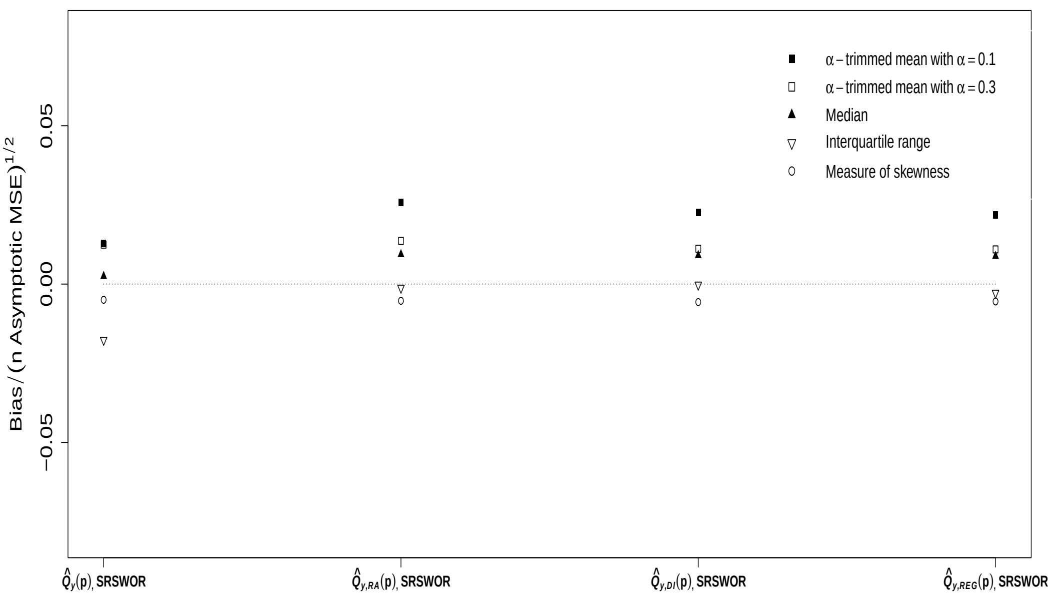

FIGURE 4.15: Ratios of biases and (n asymptotic MSE)$^{1/2}$'s for different estimators under SRSWOR in the case of n=200.

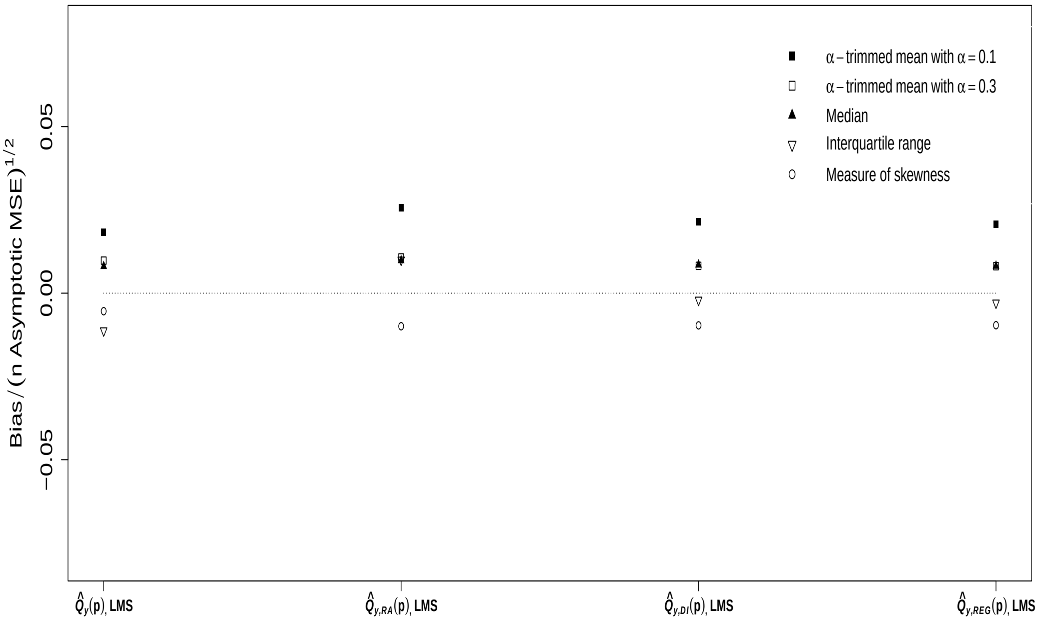

FIGURE 4.16: Ratios of biases and $(n$ asymptotic MSE$)^{1/2}$'s for different estimators under LMS sampling design in the case of n=200.

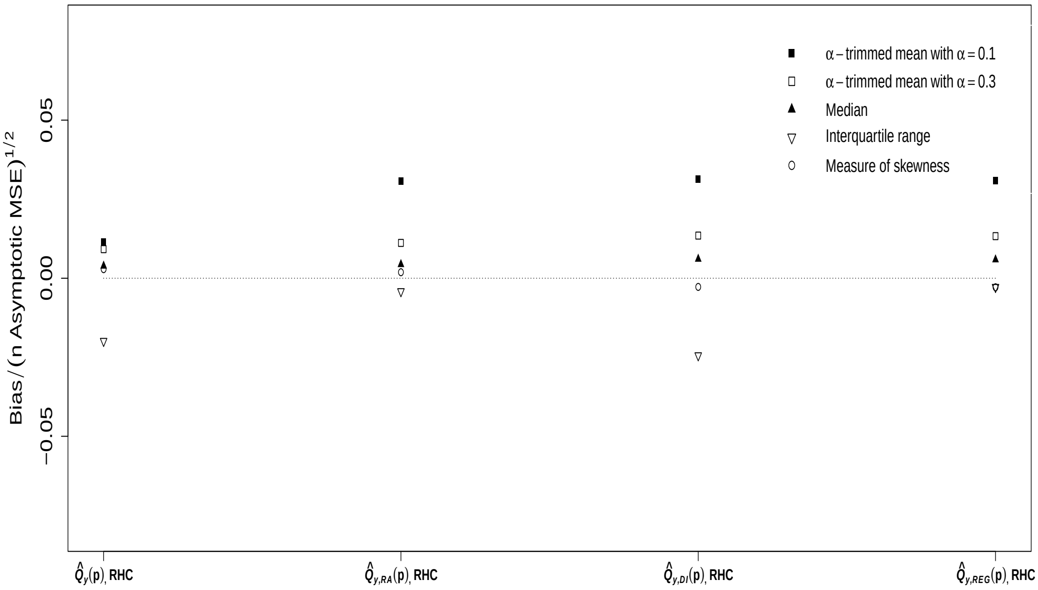

FIGURE 4.17: Ratios of biases and $(n$ asymptotic MSE$)^{1/2}$'s for different estimators under RHC sampling design in the case of n=200.

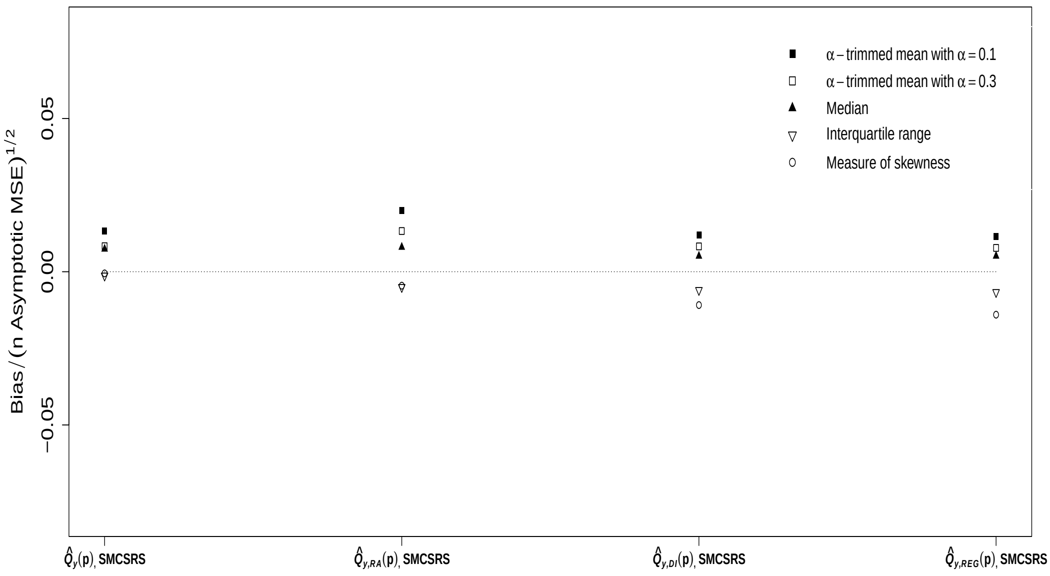

FIGURE 4.18: Ratios of biases and $(n$ asymptotic MSE$)^{1/2}$'s for different estimators under SMCSRS in the case of n=108. In this figure, SMCSRS stands for stratified multistage cluster sampling design with SRSWOR.

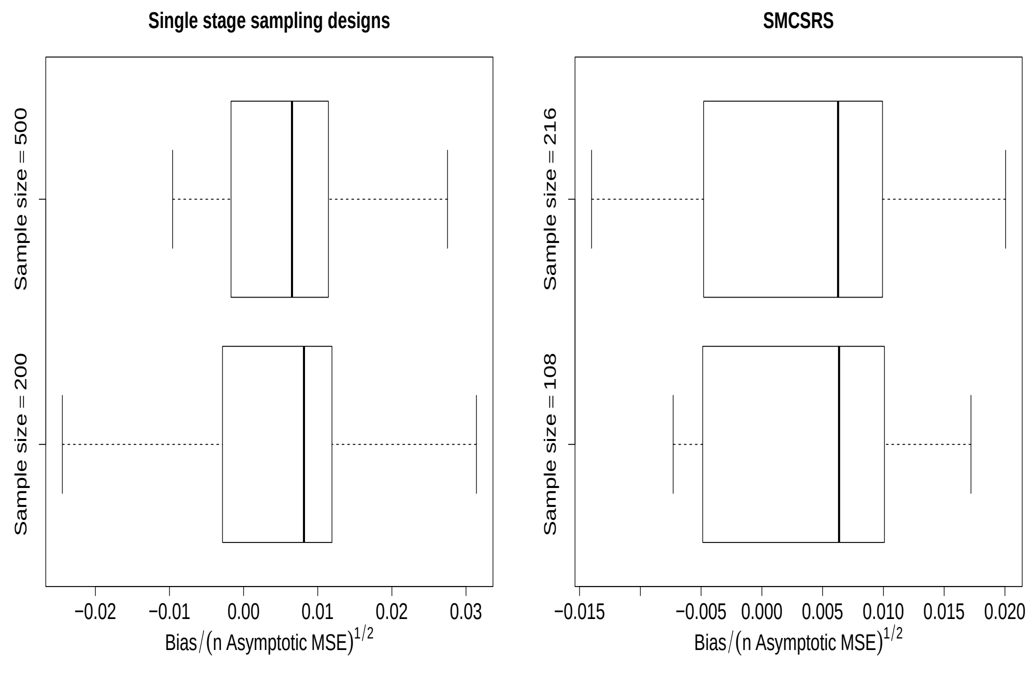

FIGURE 4.19: Boxplots of ratios of biases and $(n$ asymptotic MSE$)^{1/2}$'s for different parameters and estimators in the cases of single stage sampling designs and SMCSRS. In this figure, SMCSRS stands for stratified multistage cluster sampling design with SRSWOR.

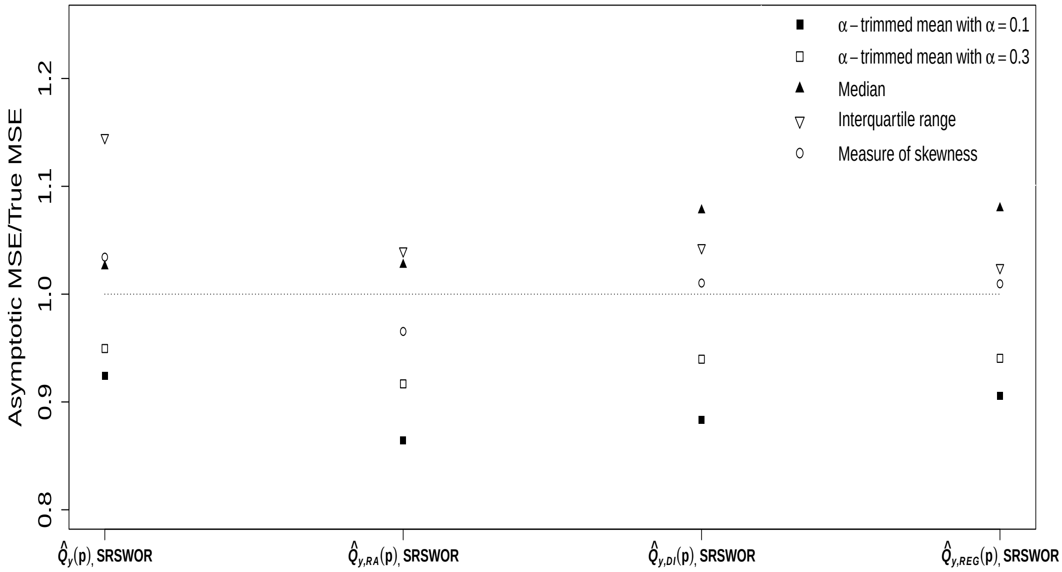

FIGURE 4.20: Ratios of asymptotic and true MSEs of different estimators for n=500 in the case of SRSWOR.

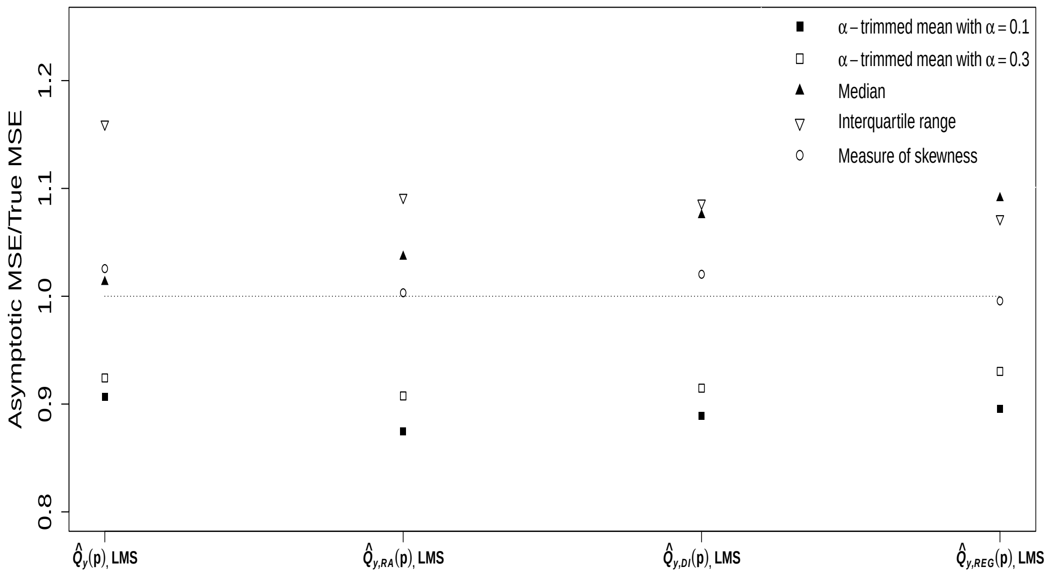

FIGURE 4.21: Ratios of asymptotic and true MSEs of different estimators for n=500 in the case of LMS sampling design.

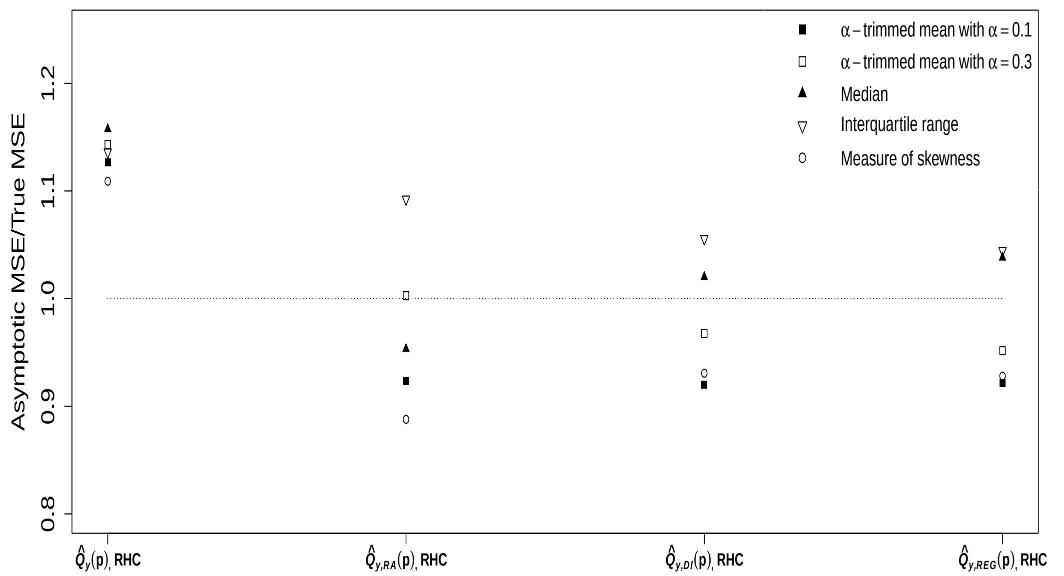

FIGURE 4.22: Ratios of asymptotic and true MSEs of different estimators for n=500 in the case of RHC sampling design.

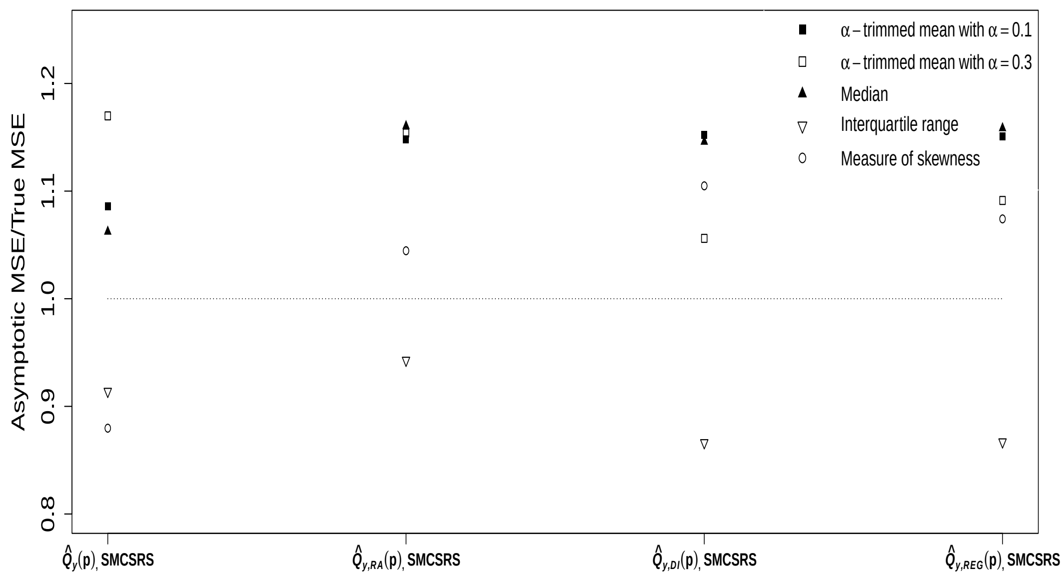

FIGURE 4.23: Ratios of asymptotic and true MSEs of different estimators for n=216 in the case of SMCSRS. In this figure, SMCSRS stands for stratified multistage cluster sampling design with SRSWOR.

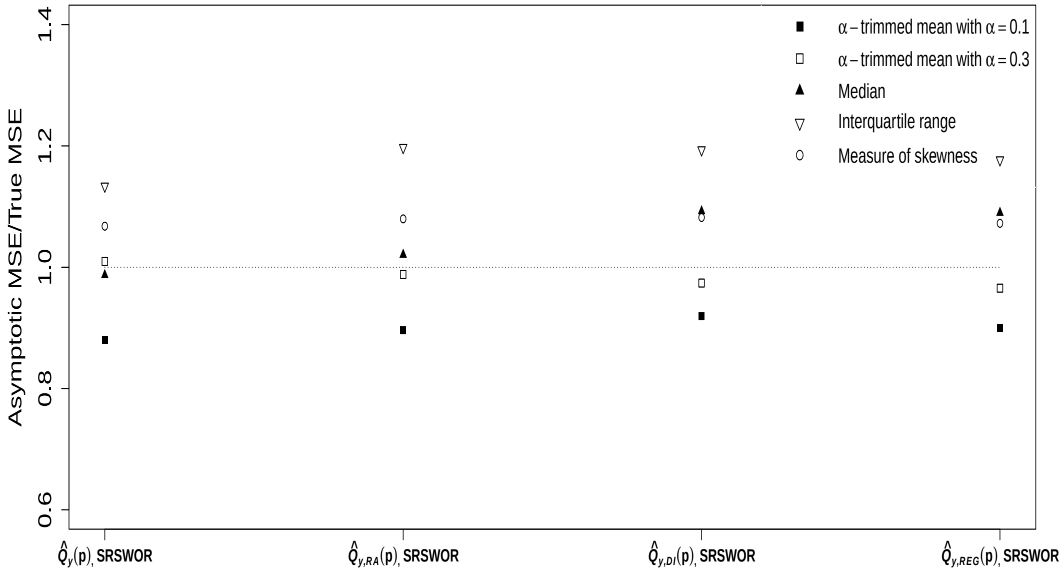

FIGURE 4.24: Ratios of asymptotic and true MSEs of different estimators for n=200 in the case of SRSWOR.

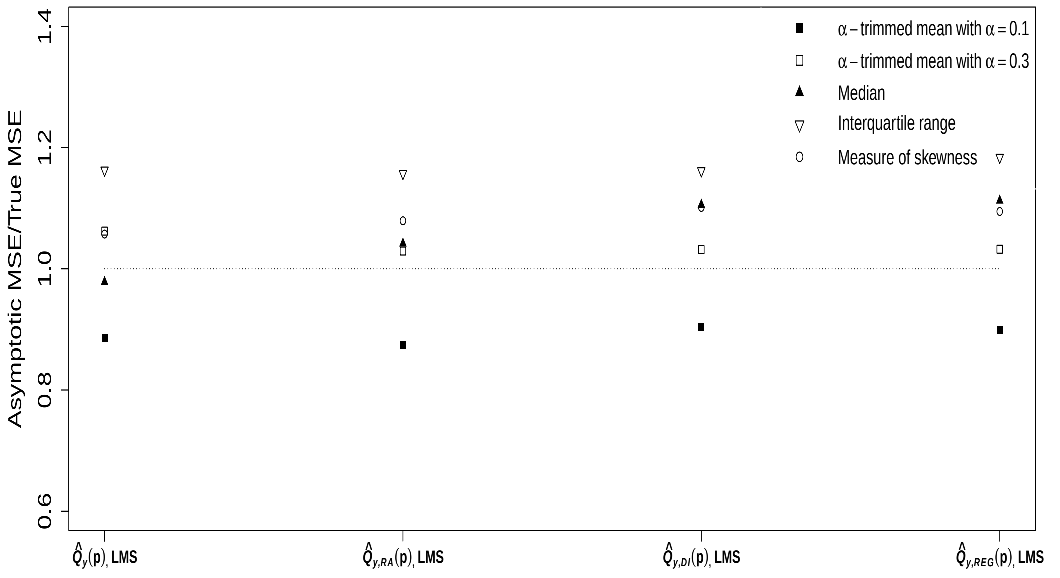

FIGURE 4.25: Ratios of asymptotic and true MSEs of different estimators for n=200 in the case of LMS sampling design.

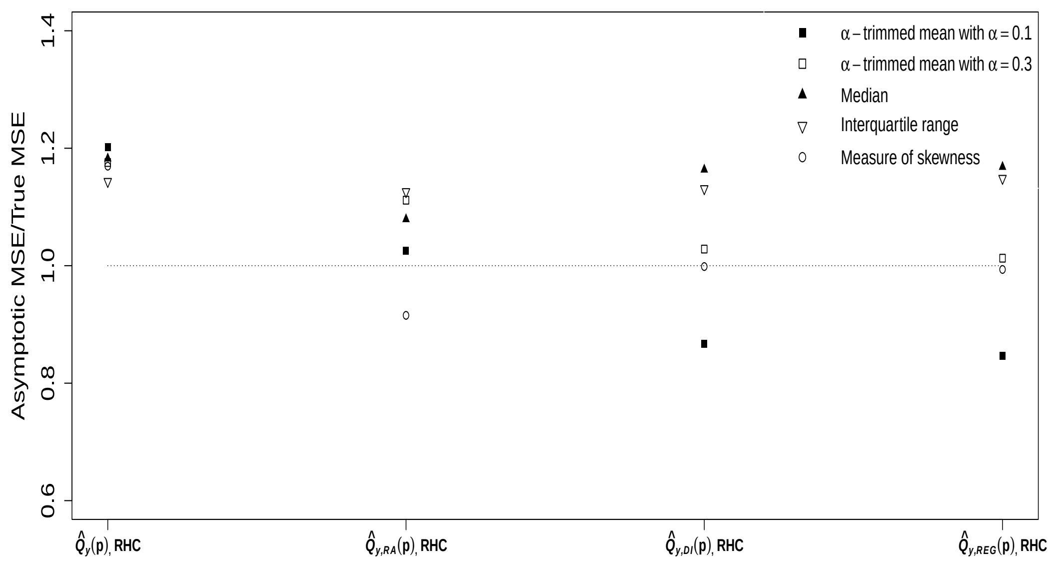

FIGURE 4.26: Ratios of asymptotic and true MSEs of different estimators for n=200 in the case of RHC sampling design.

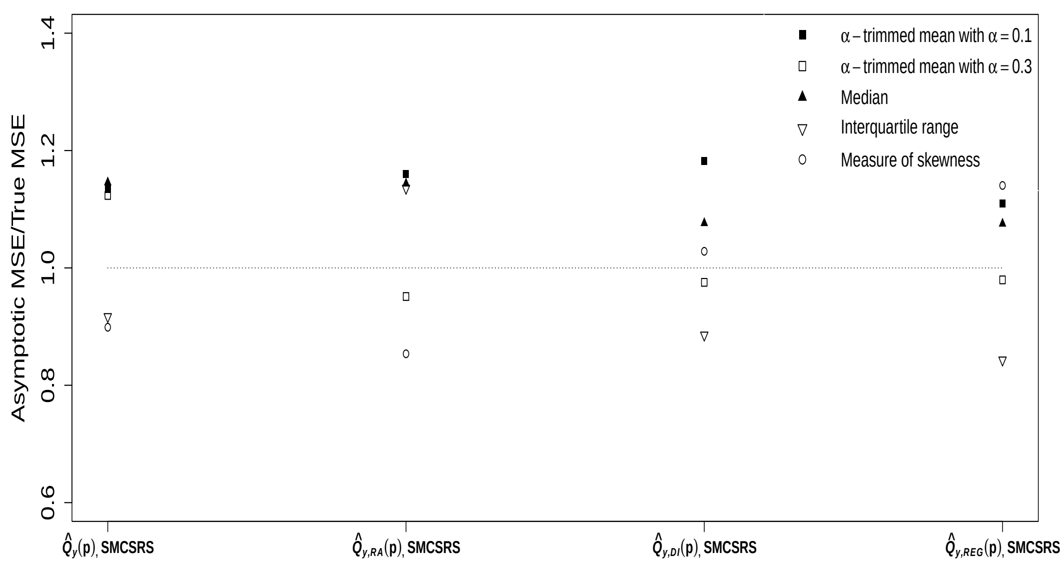

FIGURE 4.27: Ratios of asymptotic and true MSEs of different estimators for n=108 in the case of SMCSRS. In this figure, SMCSRS stands for stratified multistage cluster sampling design with SRSWOR.

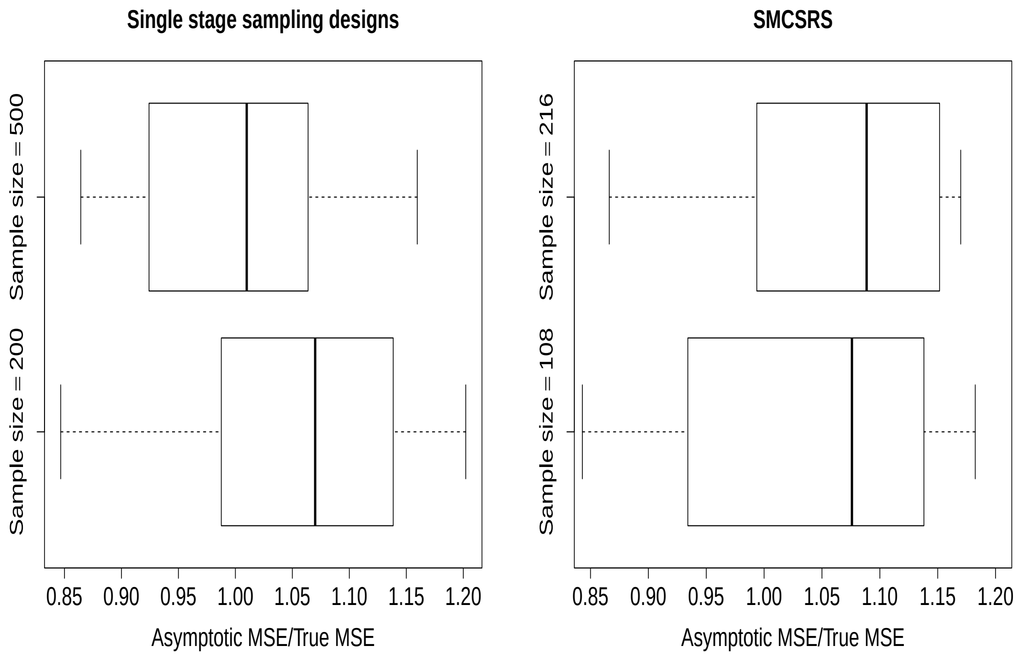

FIGURE 4.28: Boxplots of ratios of asymptotic and true MSEs for different estimators and parameters in the cases of single stage sampling designs and SMCSRS. In this figure, SMCSRS stands for stratified multistage cluster sampling design with SRSWOR.

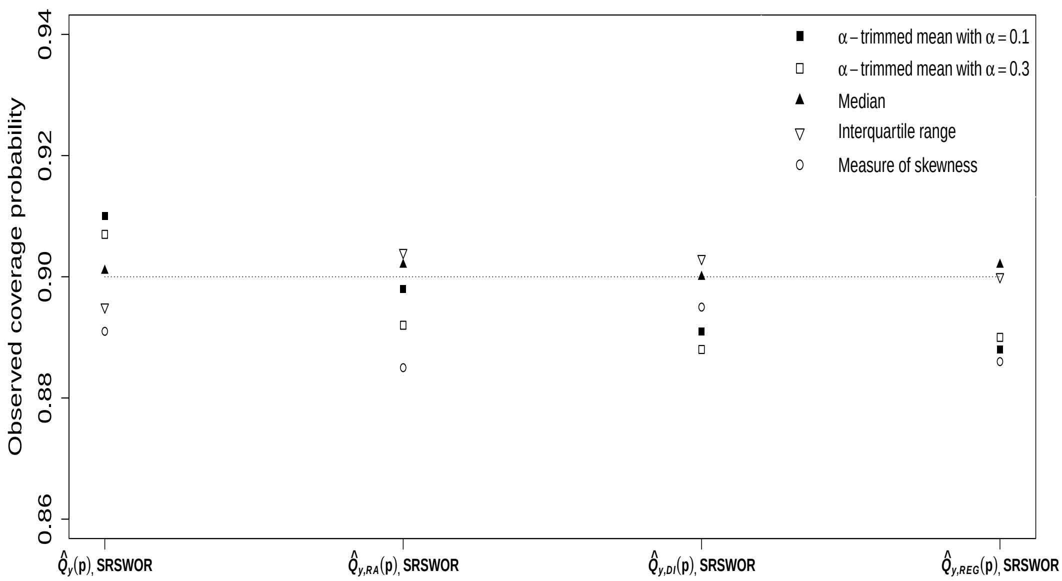

FIGURE 4.29: Observed coverage probabilities of nominal 90% confidence intervals for n=500 in the case of SRSWOR (the number of simulation iterations is 1000 and the magnitude of the Monte Carlo standard error for observed coverage probabilities is 0.009).

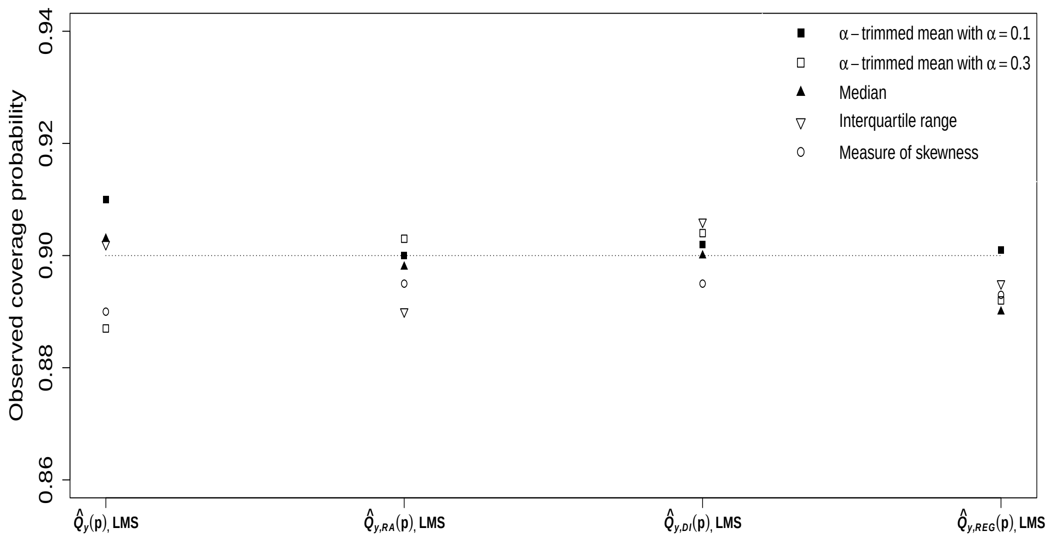

FIGURE 4.30: Observed coverage probabilities of nominal 90% confidence intervals for n=500 in the case of LMS sampling design (the number of simulation iterations is 1000 and the magnitude of the Monte Carlo standard error for observed coverage probabilities is 0.009).

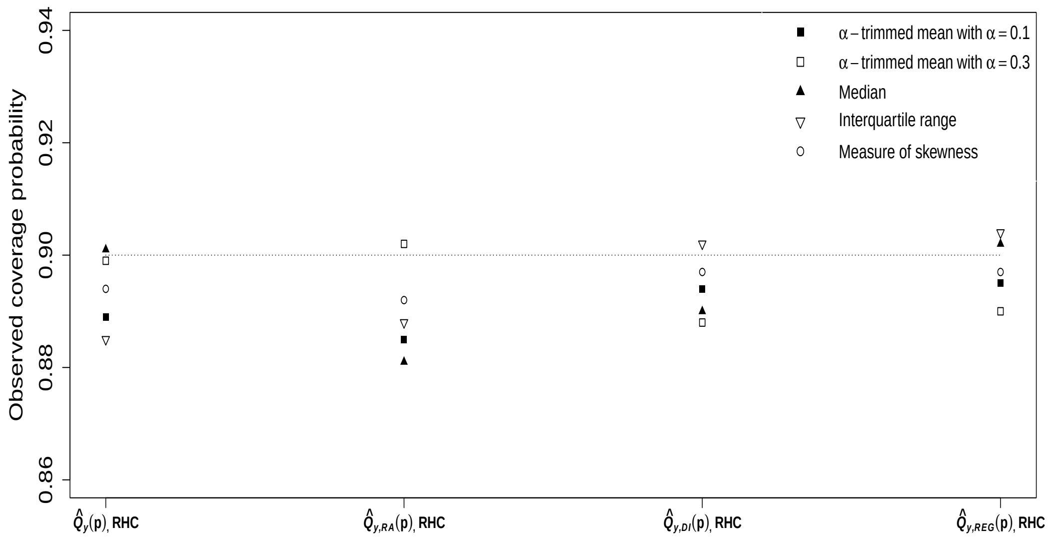

FIGURE 4.31: Observed coverage probabilities of nominal 90% confidence intervals for n=500 in the case of RHC sampling design (the number of simulation iterations is 1000 and the magnitude of the Monte Carlo standard error for observed coverage probabilities is 0.009).

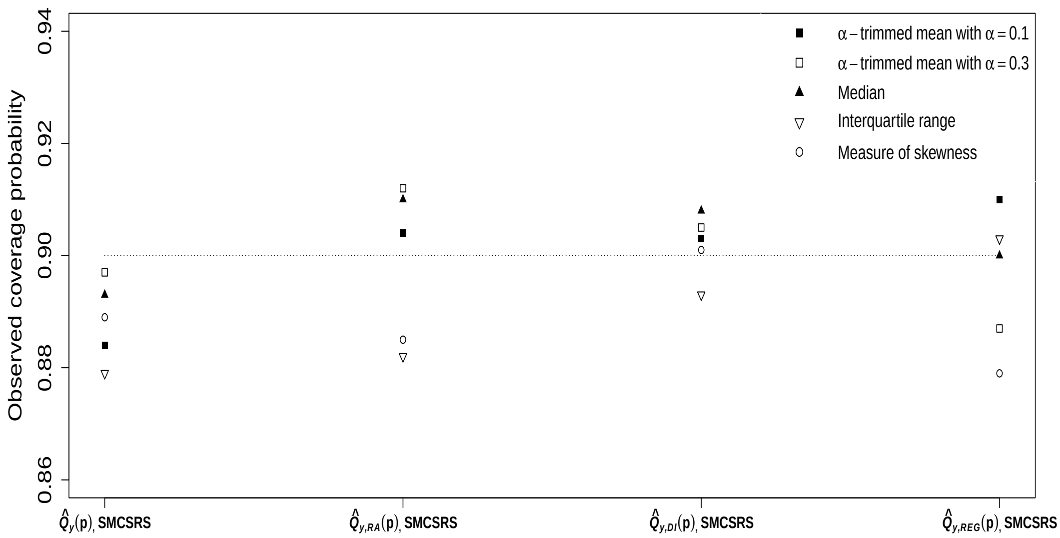

FIGURE 4.32: Observed coverage probabilities of nominal 90% confidence intervals for n=216 in the case of SMCSRS (the number of simulation iterations is 1000 and the magnitude of the Monte Carlo standard error for observed coverage probabilities is 0.009). In this figure, SMCSRS stands for stratified multistage cluster sampling design with SRSWOR.

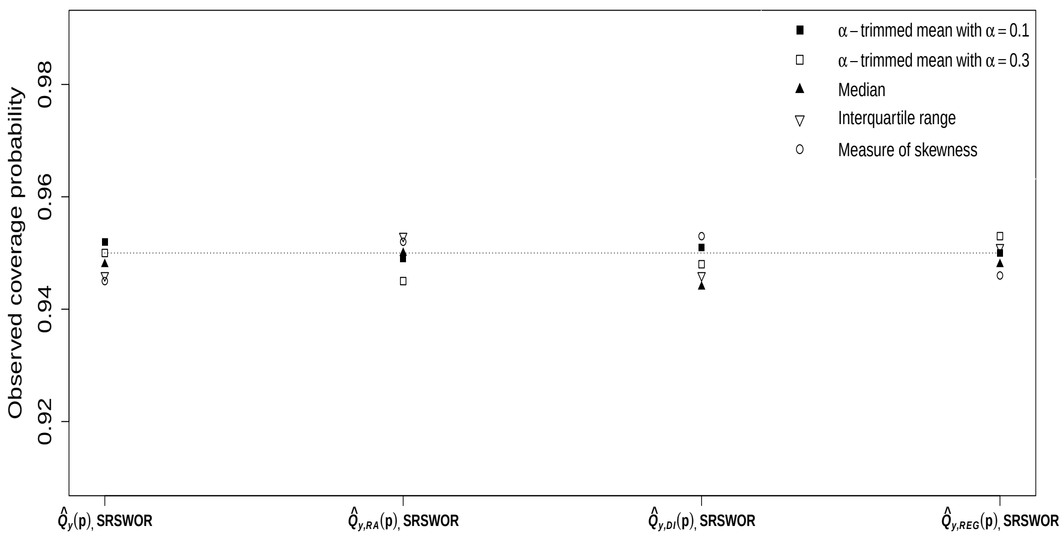

FIGURE 4.33: Observed coverage probabilities of nominal 95% confidence intervals for n=500 in the case of SRSWOR (the number of simulation iterations is 1000 and the magnitude of the Monte Carlo standard error for observed coverage probabilities is 0.007).

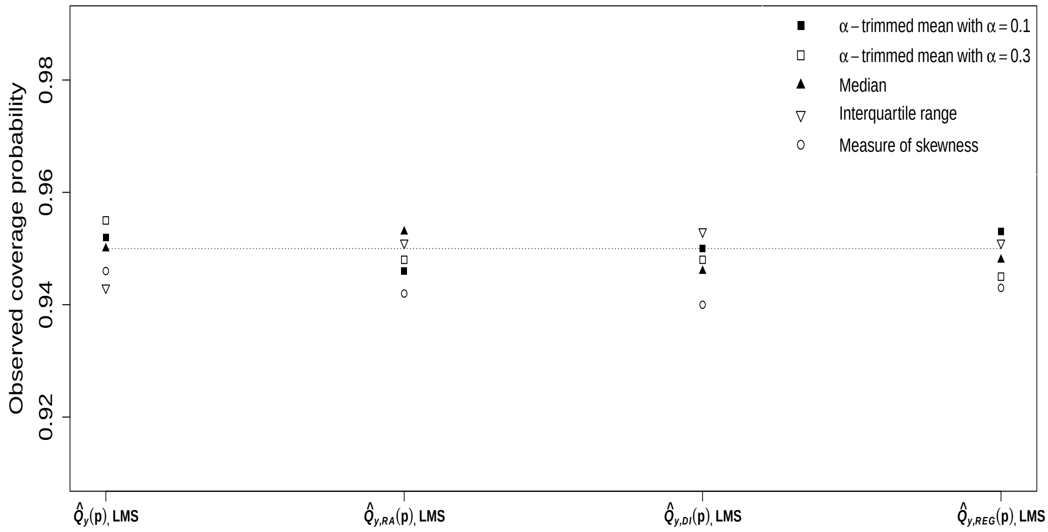

FIGURE 4.34: Observed coverage probabilities of nominal 95% confidence intervals for n=500 in the case of LMS sampling design (the number of simulation iterations is 1000 and the magnitude of the Monte Carlo standard error for observed coverage probabilities is 0.007).

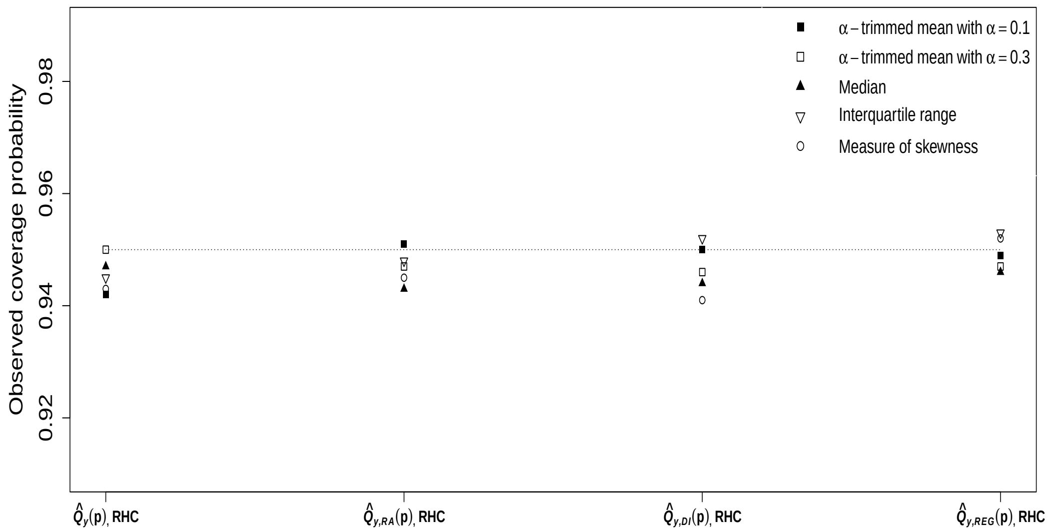

FIGURE 4.35: Observed coverage probabilities of nominal 95% confidence intervals for n=500 in the case of RHC sampling design (the number of simulation iterations is 1000 and the magnitude of the Monte Carlo standard error for observed coverage probabilities is 0.007).

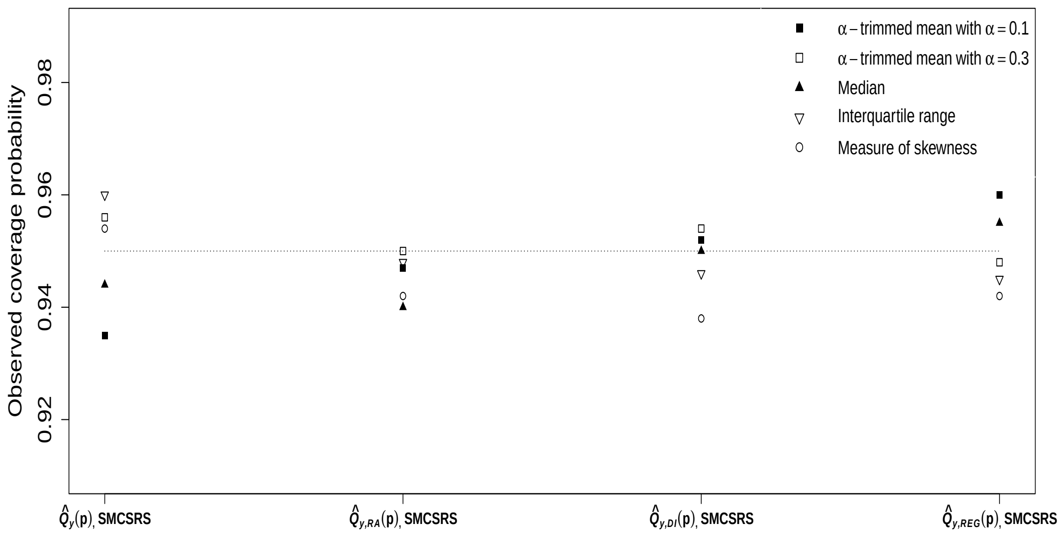

FIGURE 4.36: Observed coverage probabilities of nominal 95% confidence intervals for n=216 in the case of SMCSRS (the number of simulation iterations is 1000 and the magnitude of the Monte Carlo standard error for observed coverage probabilities is 0.007). In this figure, SMCSRS stands for stratified multistage cluster sampling design with SRSWOR.

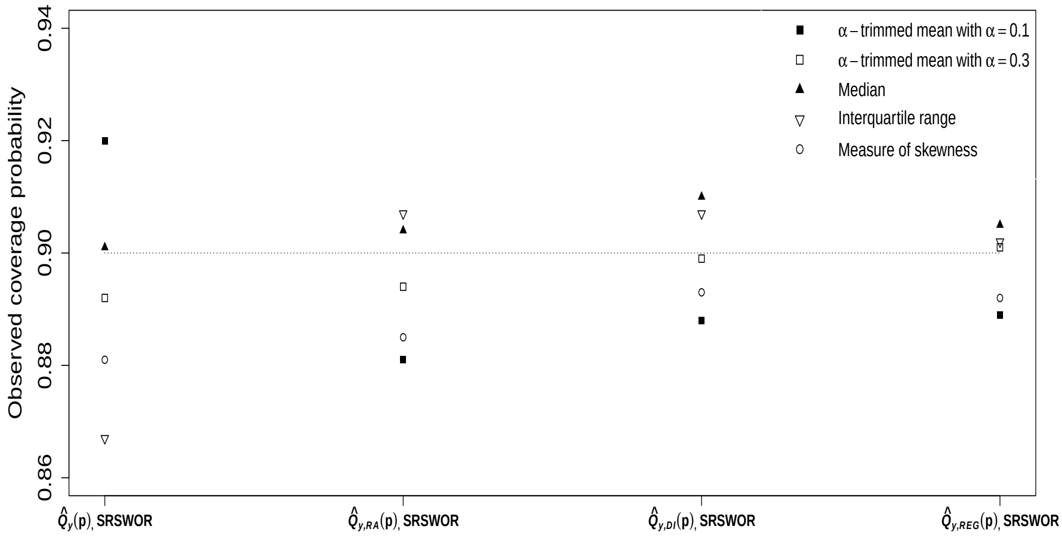

FIGURE 4.37: Observed coverage probabilities of nominal 90% confidence intervals for n=200 in the case of SRSWOR (the number of simulation iterations is 1000 and the magnitude of the Monte Carlo standard error for observed coverage probabilities is 0.009).

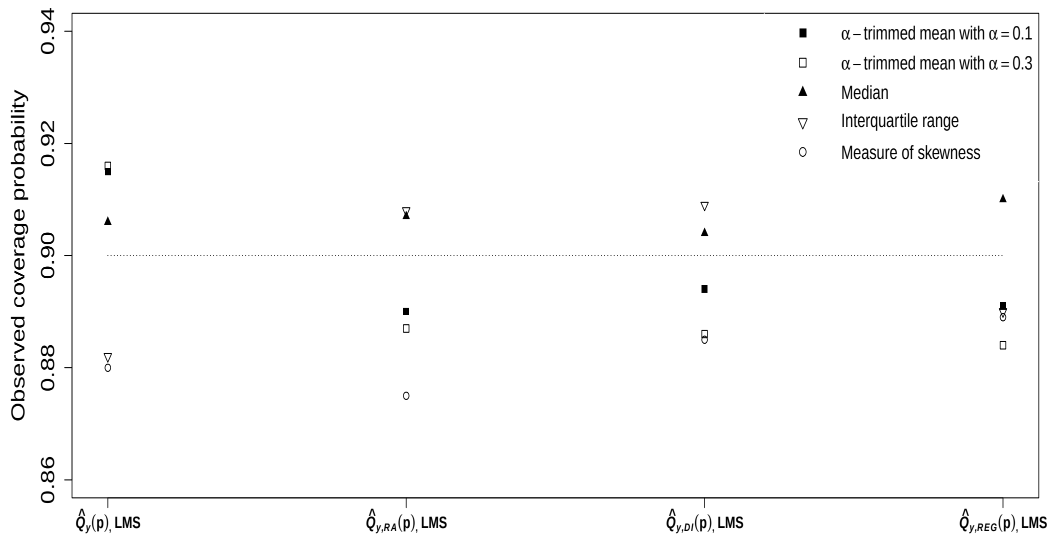

FIGURE 4.38: Observed coverage probabilities of nominal 90% confidence intervals for n=200 in the case of LMS sampling design (the number of simulation iterations is 1000 and the magnitude of the Monte Carlo standard error for observed coverage probabilities is 0.009).

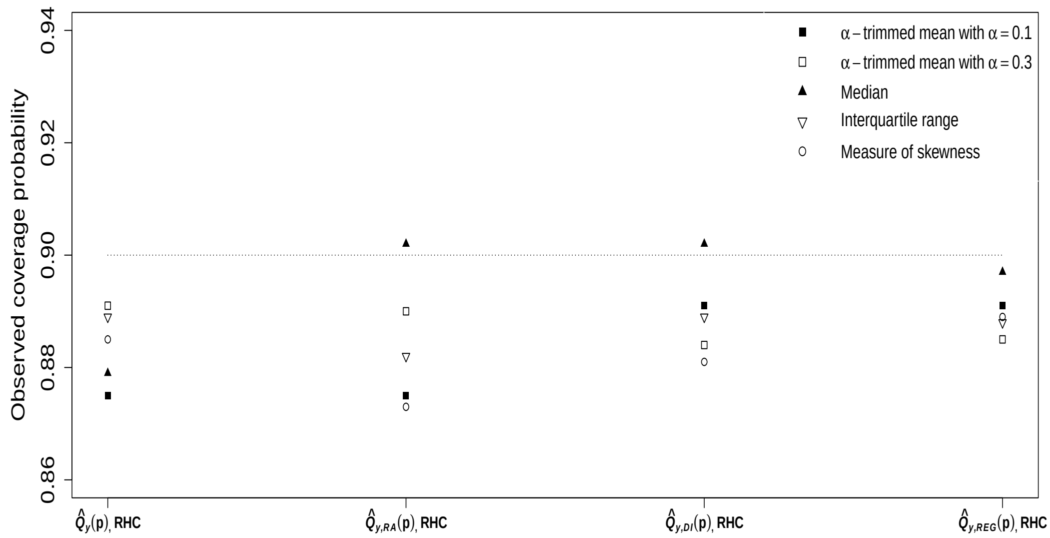

FIGURE 4.39: Observed coverage probabilities of nominal 90% confidence intervals for n=200 in the case of RHC sampling design (the number of simulation iterations is 1000 and the magnitude of the Monte Carlo standard error for observed coverage probabilities is 0.009).

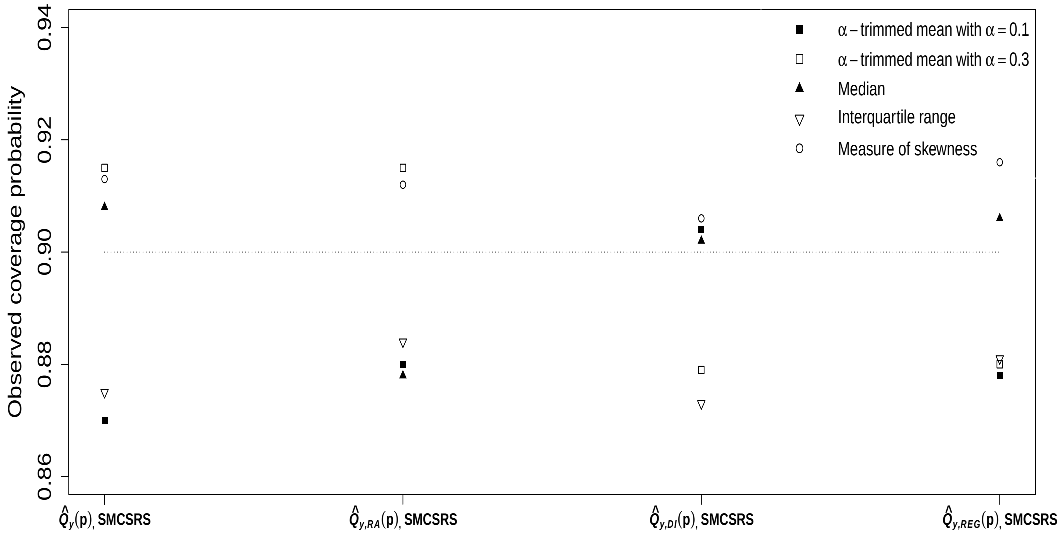

FIGURE 4.40: Observed coverage probabilities of nominal 90% confidence intervals for n=108 in the case of SMCSRS (the number of simulation iterations is 1000 and the magnitude of the Monte Carlo standard error for observed coverage probabilities is 0.009). In this figure, SMCSRS stands for stratified multistage cluster sampling design with SRSWOR.

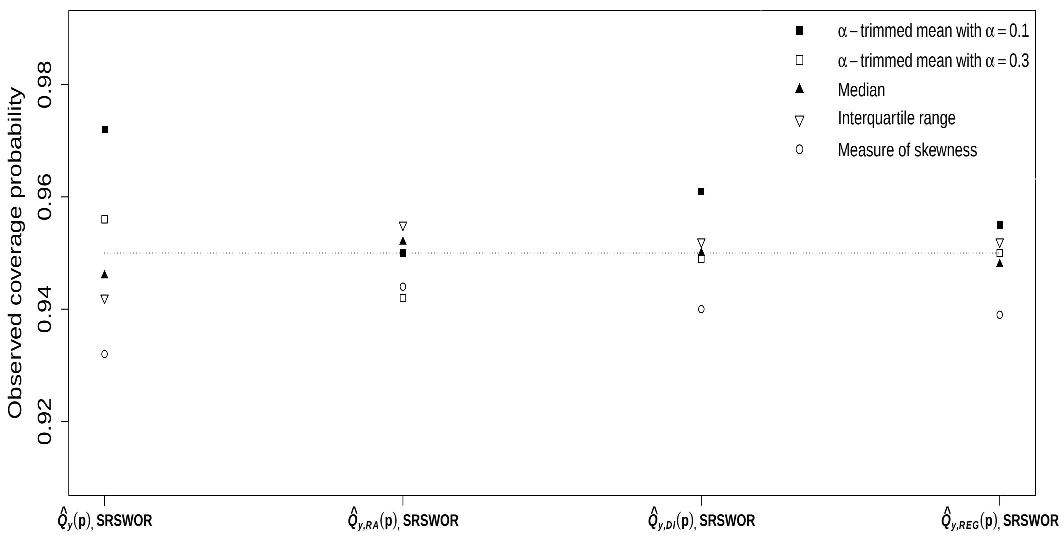

FIGURE 4.41: Observed coverage probabilities of nominal 95% confidence intervals for n=200 in the case of SRSWOR (the number of simulation iterations is 1000 and the magnitude of the Monte Carlo standard error for observed coverage probabilities is 0.007).

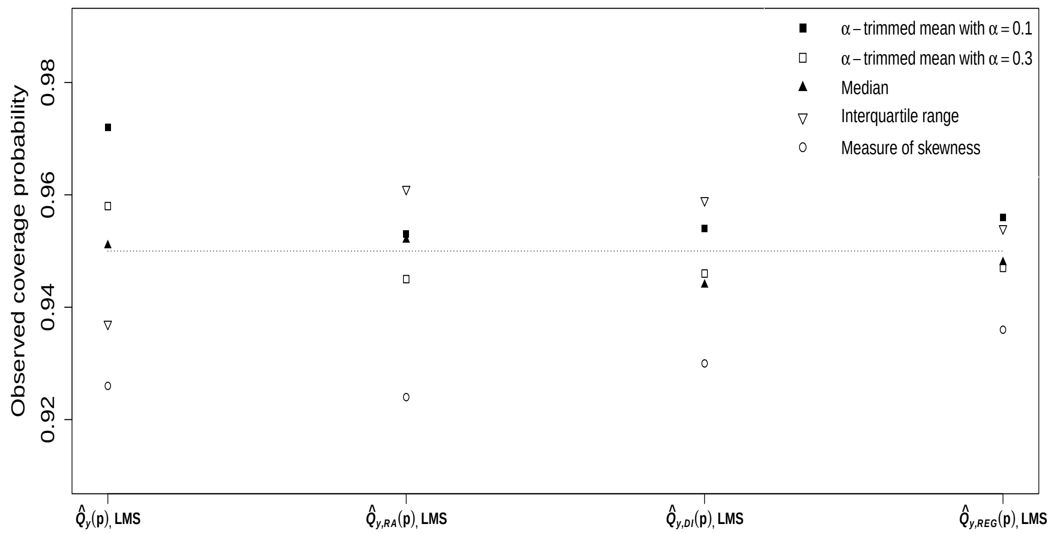

FIGURE 4.42: Observed coverage probabilities of nominal 95% confidence intervals for n=200 in the case of LMS sampling design (the number of simulation iterations is 1000 and the magnitude of the Monte Carlo standard error for observed coverage probabilities is 0.007).

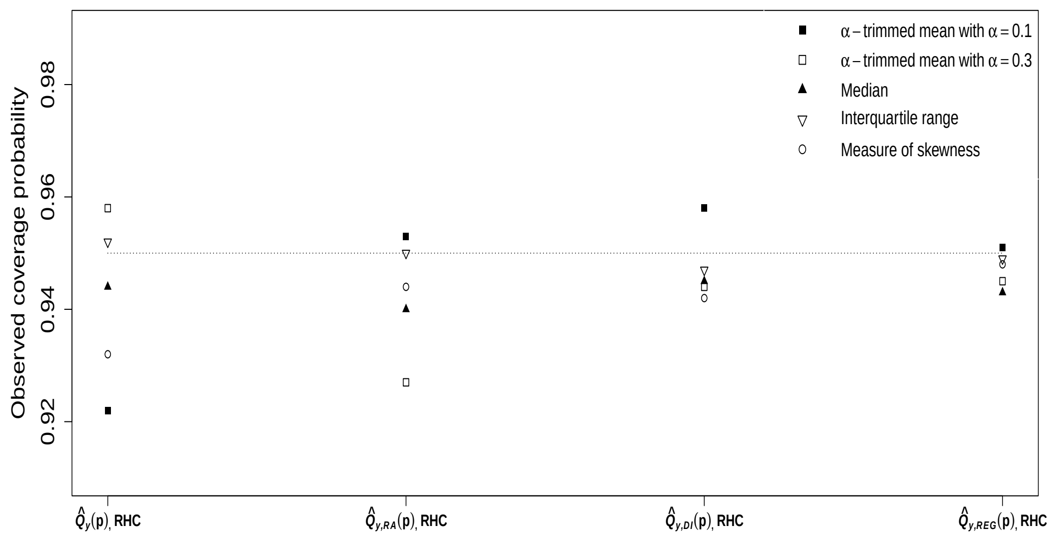

FIGURE 4.43: Observed coverage probabilities of nominal 95% confidence intervals for n=200 in the case of RHC sampling design (the number of simulation iterations is 1000 and the magnitude of the Monte Carlo standard error for observed coverage probabilities is 0.007).

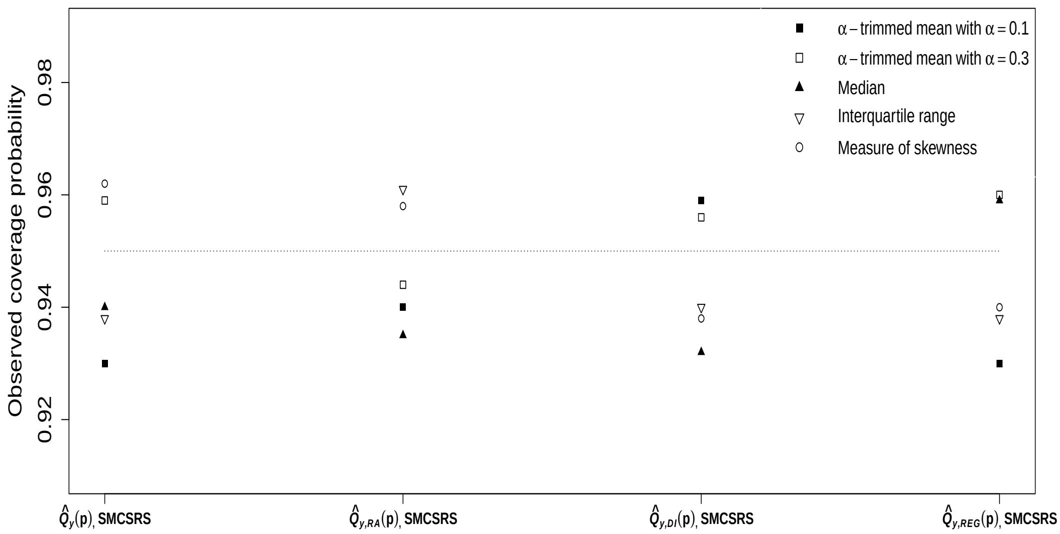

FIGURE 4.44: Observed coverage probabilities of nominal 95% confidence intervals for n=108 in the case of SMCSRS (the number of simulation iterations is 1000 and the magnitude of the Monte Carlo standard error for observed coverage probabilities is 0.007). In this figure, SMCSRS stands for stratified multistage cluster sampling design with SRSWOR.

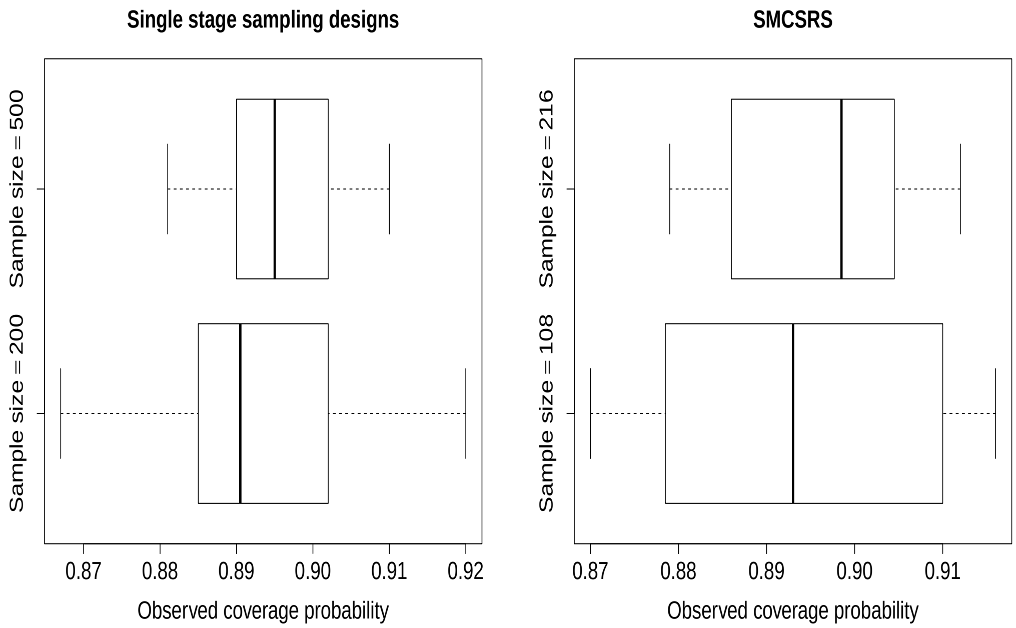

FIGURE 4.45: Boxplots of observed coverage probabilities of nominal 90% confidence intervals for different estimators and parameters in the cases of single stage sampling designs and SMCSRS. In this figure, SMCSRS stands for stratified multistage cluster sampling design with SRSWOR.

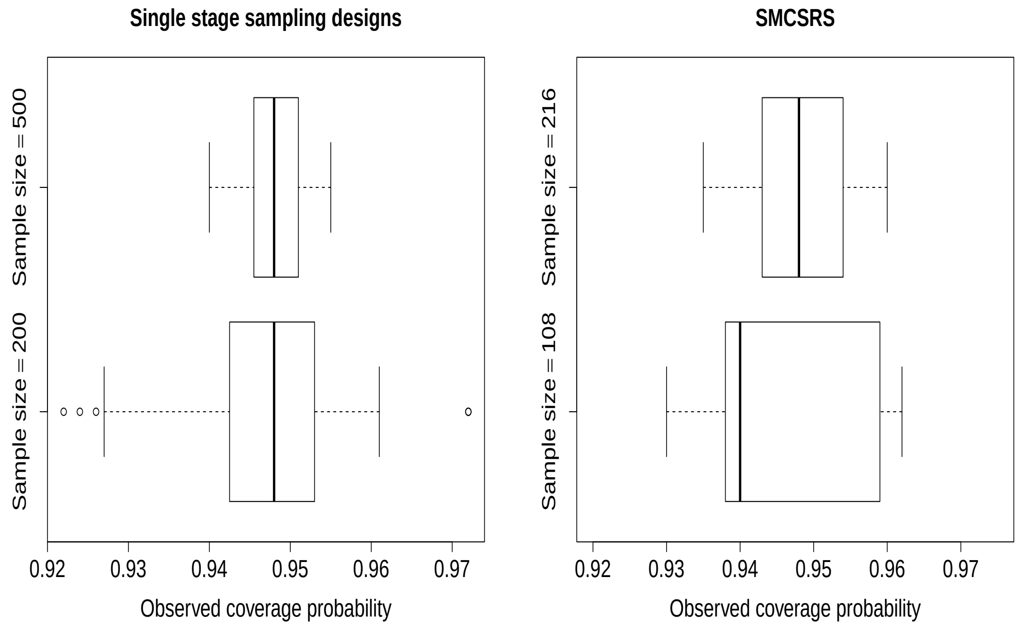

FIGURE 4.46: Boxplots of observed coverage probabilities of nominal 95% confidence intervals for different estimators and parameters in the cases of single stage sampling designs and SMCSRS. In this figure, SMCSRS stands for stratified multistage cluster sampling design with SRSWOR.

The results obtained from the above data analysis are summarised as follows.

(i) It follows from Figures 4.2–4.9 above (see also the boxplot in Figure 4.10 above) that for different parameters, estimators, sampling designs and sample sizes considered in this section, relative biases are quite close to 0 except for the following cases. Figures 4.6 and 4.7 that for n=200, the estimator of the interquartile range based on difference estimator under SRSWOR and the estimators of measure of skewness based on ratio, difference and regression estimators under LMS sampling design have somewhat large negative biases compared to the other estimators. Also, Figures 4.5 and 4.9 shows that the estimators of measure of skewness based on ratio, difference and regression estimators under stratified multistage cluster sampling design with SRSWOR have relatively large negative biases compared to the other estimators for both n=108 and n=216.

(ii) It can be seen from Figures 4.11–4.18 above (see also the boxplot in Figure 4.19 above) that for different parameters, estimators, sampling designs and sample sizes considered in this section, biases relative to $(n \text{ asymptotic MSE})^{1/2}$'s are close to 0.

(iii) It follows from Figures 4.20–4.27 above (see also the boxplot in Figure 4.28 above) that ratios of asymptotic and true MSEs for different parameters, estimators and sampling designs become closer to 1 as the sample size increases from n=200 to n=500.

(iv) Figures 4.29–4.44 above (see also the boxplots in Figures 4.45 and 4.46 above) show that for different parameters, estimators, sampling designs and sample sizes, observed coverage probabilities of nominal 90% and 95% confidence intervals are quite close to 90% and 95%, respectively, except for the following case. Observed coverage probability of nominal 95% confidence interval of α-trimmed mean with α=0.1 based on the sample quantile under SRSWOR and sample size n=200 is 97.2%.

(v) Overall, the asymptotic approximations of the distributions of different estimators of different parameters considered in this chapter seem to work well in finite sample situations. Also, the accuracy of the asymptotic approximations increases as the sample size increases.

4.7. Proofs of the main results

Before we give the proof of Proposition 4.2.1, suppose that $P(s,\omega)$ denotes a high entropy sampling design satisfying Assumption 4.2.2, and $Q(s,\omega)$ denotes a rejective sampling design having inclusion probabilities equal to those of $P(s,\omega)$. Such a rejective sampling design always exists (see [4]).

Proof of Proposition 4.2.1. We shall first show that the conclusion of Proposition 4.2.1 holds for $Q(s,\omega)$. Let us define

$$F_{u,N}(t)=\sum_{i=1}^{N}\mathbb{1}_{[U_i\leq t]}/N \text{ and } \mathbb{U}_n(t)=\sqrt{n}\sum_{i\in s}(N\pi_i)^{-1}(\mathbb{1}_{[U_i\leq t]}-F_{u,N}(t)) \qquad (4.7.1)$$

for $0\leq t\leq 1$. Then, for $d(i,s)$=$(N\pi_i)^{-1}$, we have

$$\mathbb{H}_n:=\left\{\sqrt{n}(\hat{F}_u(t)-t):t\in[0,1]\right\}=\mathbb{Z}_n+\sqrt{n/N}\mathbb{W}_N \text{ with } \mathbb{Z}_n=\left\{\mathbb{U}_n(t)\Big/\sum_{i\in s}(N\pi_i)^{-1}:t\in[0,1]\right\} \text{ and } \mathbb{W}_N=\left\{\sqrt{N}(F_{u,N}(t)-t):t\in[0,1]\right\}. \qquad (4.7.2)$$

Next, define

$$B_{u,N}(t_1,t_2) = F_{u,N}(t_2) - F_{u,N}(t_1) \text{ and } \mathbb{B}_n(t_1,t_2) = \mathbb{U}_n(t_2) - \mathbb{U}_n(t_1) \quad (4.7.3)$$

for $0 \le t_1 < t_2 \le 1$. Then, by Lemma 4.8.2 in Section 4.8, we have $E\big[\big(\mathbb{B}_n(t_1,t_2)\big)^2 \times \big(\mathbb{B}_n(t_2,t_3)\big)^2\big] \le K_1\big(B_{u,N}(t_1,t_3)\big)^2$ for all dyadic rational numbers $0 \le t_1 < t_2 < t_3 \le 1$ *a.s.* $[\mathbf{P}]$, where $K_1 > 0$ is some constant and $\nu \ge 1$. This further implies that

$$E\big[\big(\mathbb{B}_n(t_1,t_2)\big)^2\big(\mathbb{B}_n(t_2,t_3)\big)^2\big] \le K_1\big(B_{u,N}(t_1,t_3)\big)^2 \text{ for any } 0 \le t_1 < t_2 < t_3 \le 1 \quad (4.7.4)$$

a.s. $[\mathbf{P}]$, where $\nu \ge 1$. Suppose that

$$w_n(1/r) = \sup_{|t-u|\le 1/r} |\mathbb{U}_n(t) - \mathbb{U}_n(u)| \text{ and } B = \{s \in \mathcal{S} : w_n(1/r) \ge \delta\} \quad (4.7.5)$$

for r=$1, 2, \ldots$. Here, $w_n(1/r)$ is the modulus of continuity of $\{\mathbb{U}_n(t) : t \in [0,1]\}$. Then, by using (4.7.4) above and imitating the proof of Lemma 2.3.1 in [79] (see p. 49), we obtain

$$\sum_{s\in B} Q(s,\omega) \le \delta^{-4}\Big(\sum_{j=1}^{r} E\big\{\mathbb{B}_n\big((j-1)/r, j/r\big)\big\}^4 + K_2 B_{u,N}(0,1) \max_{1\le j\le r} B_{u,N}\big((j-1)/r, j/r\big)\Big) \quad (4.7.6)$$

a.s. $[\mathbf{P}]$ for any $\delta > 0$, $r \ge 1, \nu \ge 1$ and some constant $K_2 > 0$. Next, it follows from (4.7.6) that

$$\overline{\lim}_{\nu\to\infty} E\big\{\mathbb{B}_n\big((j-1)/r, j/r\big)\big\}^4 \le K_3(1/r)^2 \quad (4.7.7)$$

a.s. $[\mathbf{P}]$ for any j=$1, \ldots, r$, $r \ge 1$ and some constant $K_3 > 0$ by Lemma 4.8.2 in Section 4.8. Now, note that $\{U_i\}_{i=1}^N$ are i.i.d. uniform random variables supported on $(0,1)$ since F_y is continuous by Assumption 4.2.3. Then, $B_{u,N}(t_1,t_2) \to t_2 - t_1$ *a.s.* $[\mathbf{P}]$ by SLLN. Therefore, in view of (4.7.6) and (4.7.7), we have

$$\overline{\lim}_{\nu\to\infty} \sum_{s\in B} Q(s,\omega) \le \delta^{-4}(K_2/r + K_3/r) \text{ a.s. } [\mathbf{P}] \quad (4.7.8)$$

for any $\delta > 0$ and $r \ge 1$. Since, $\sum_{s\in B} Q(s,\omega)$ is bounded, by taking expectation of left hand side in (4.7.8) w.r.t. $\mathbf{P}$ and applying an extended version of Fatou's lemma, we obtain that

$$\overline{\lim}_{\nu\to\infty} \mathbf{P}^*\{w_n(1/r) \ge \delta\} \le \delta^{-4}(K_2/r + K_3/r) \quad (4.7.9)$$

for any $\delta > 0$ and $r \geq 1$. This further implies that $\overline{\lim}_{\nu\to\infty}\mathbf{P}^*\{w_n(1/r) \geq \delta\} \to 0$ for any δ as $r \to \infty$. Then by Theorem 2.3.2 in [79] (see p. 46), $\{\mathbb{U}_n : \nu \geq 1\}$ is weakly/relatively compact in $(\tilde{D}[0,1], \tilde{\mathcal{D}})$ with respect to the sup norm metric under $\mathbf{P}^*$. In other words, given any subsequence $\{\nu_k\}$, there exists a further subsequence $\{\nu_{k_l}\}$ such that $E_{\mathbf{P}^*}(f(\mathbb{U}_n)) \to E(f(\mathbb{U}))$ along the subsequence $\{\nu_{k_l}\}$ for any bounded continuous (with respect to the sup norm metric) and $\tilde{\mathcal{D}}$-measurable function f, and for some random function $\mathbb{U}$ in $(\tilde{D}[0,1], \tilde{\mathcal{D}})$ (see p. 44 in [79]).

Now, under $Q(s,\omega)$, $\mathbf{m}(\mathbb{U}_n(t_1), \ldots, \mathbb{U}_n(t_k))^T \xrightarrow{\mathcal{L}} N(0, \mathbf{m}\Gamma_3\mathbf{m}^T)$ as $\nu \to \infty$ *a.s.* [**P**] by Lemma 4.8.1 in Section 4.8, where $k \geq 1$, $t_1, \ldots, t_k \in (0,1)$, $m \in \mathbb{R}^k$, $\mathbf{m} \neq 0$ and Γ_3 is a p.d. matrix. Moreover, Γ_3=$\lim_{\nu\to\infty} nN^{-2}\sum_{i=1}^{N}(\mathbf{U}_i - \mathbf{T}_U\pi_i)^T(\mathbf{U}_i - \mathbf{T}_U\pi_i)(\pi_i^{-1} - 1)$ *a.s.* [**P**], where $\mathbf{U}_i$=$(\mathbb{1}_{[U_i \leq t_1]} - F_{u,N}(t_1), \ldots, \mathbb{1}_{[U_i \leq t_k]} - F_{u,N}(t_k))$ and $\mathbf{T}_U$=$\sum_{i=1}^{N}\mathbf{U}_i(1-\pi_i)/\sum_{i=1}^{N}\pi_i(1-\pi_i)$. Note that $\sum_{i=1}^{N}||\mathbf{U}_i||^2/N$ is bounded. Also, note that Assumption 4.2.2-*(ii)* holds under $Q(s,\omega)$ because $P(s,\omega)$ and $Q(s,\omega)$ have same inclusion probabilities, and Assumption 4.2.2-*(ii)* holds under $P(s,\omega)$. Then, we have

$$\Gamma_3 = \lim_{\nu\to\infty} E_{\mathbf{P}}(nN^{-2}\sum_{i=1}^{N}(\mathbf{U}_i - \mathbf{T}_U\pi_i)^T(\mathbf{U}_i - \mathbf{T}_U\pi_i)(\pi_i^{-1} - 1)) \tag{4.7.10}$$

by DCT. Further, it follows from DCT that under $\mathbf{P}^*$,

$$\begin{aligned} &\mathbf{m}(\mathbb{U}_n(t_1), \ldots, \mathbb{U}_n(t_k))^T \xrightarrow{\mathcal{L}} N(0, \mathbf{m}\Gamma_3\mathbf{m}^T) \text{ for any } \mathbf{m} \neq 0, \text{ and hence} \\ &(\mathbb{U}_n(t_1), \ldots, \mathbb{U}_n(t_k)) \xrightarrow{\mathcal{L}} N(0, \Gamma_3) \end{aligned} \tag{4.7.11}$$

as $\nu \to \infty$. Relative compactness and weak convergence of finite dimensional distributions of $\{\mathbb{U}_n : \nu \geq 1\}$ imply that $\mathbb{U}_n \xrightarrow{\mathcal{L}} \mathbb{U}$ as $\nu \to \infty$ in $(\tilde{D}[0,1], \tilde{\mathcal{D}})$ with respect to the sup norm metric, for $Q(s,\omega)$ under $\mathbf{P}^*$, where $\mathbb{U}$ has mean 0 and covariance kernel

$$\begin{aligned} &\lim_{\nu\to\infty} E_{\mathbf{P}}\Bigg(nN^{-2}\sum_{i=1}^{N}(\mathbb{1}_{[U_i \leq t_1]} - F_{u,N}(t_1) - R(t_1)\pi_i)\times \\ &(\mathbb{1}_{[U_i \leq t_2]} - F_{u,N}(t_2) - R(t_2)\pi_i)(\pi_i^{-1} - 1)\Bigg), \end{aligned} \tag{4.7.12}$$

with $R(t)$=$\sum_{i=1}^{N}(\mathbb{1}_{[U_i \leq t]} - F_{u,N}(t))(1-\pi_i)/\sum_{i=1}^{N}\pi_i(1-\pi_i)$. Moreover, it follows from Theorem 2.3.2 in [79] that $\mathbb{U}$ has almost sure continuous paths. Next, note that $\sum_{i=1}^{N}\pi_i(1 - \pi_i) \to \infty$ as $\nu \to \infty$ under $Q(s,\omega)$ *a.s.* [**P**] since $Q(s,\omega)$ satisfies Assumption 4.2.2-*(ii)*, and Assumption 4.2.1 holds. Then, it can be shown using Theorem 6.1 in [40] that under $Q(s,\omega)$,

$var(\sum_{i\in s}(N\pi_i)^{-1}) \to 0$ as $\nu \to \infty$ *a.s.* [**P**]. Consequently, $\sum_{i\in s}(N\pi_i)^{-1} \xrightarrow{p} 1$ as $\nu \to \infty$ under $\mathbf{P}^*$. Then, under $\mathbf{P}^*$, $\mathbb{Z}_n$=$\mathbb{U}_n/\sum_{i\in s}(N\pi_i)^{-1} \xrightarrow{\mathcal{L}} \mathbb{Z} \stackrel{\mathcal{L}}{=} \mathbb{U}$ as $\nu \to \infty$ in $(\tilde{D}[0,1], \tilde{\mathcal{D}})$ with respect to the sup norm metric, for $Q(s,\omega)$. This further implies that under $\mathbf{P}^*$, $\mathbb{Z}_n \xrightarrow{\mathcal{L}} \mathbb{U}$ as $\nu \to \infty$ in $(\tilde{D}[0,1], \tilde{\mathcal{D}})$ with respect to the Skorohod metric, for $Q(s,\omega)$.

Now, it follows from Donsker theorem that under **P**, $\mathbb{W}_N \xrightarrow{\mathcal{L}} \mathbb{W}$ as $\nu \to \infty$ in $(\tilde{D}[0,1], \tilde{\mathcal{D}})$ with respect to the Skorohod metric, where $\mathbb{W}$ is the standard Brownian bridge in $\tilde{D}[0,1]$ and has almost sure continuous paths. Hence, under $\mathbf{P}^*$, both $\mathbb{Z}_n$ and $\mathbb{W}_N$ are tight in $(\tilde{D}[0,1], \tilde{\mathcal{D}})$ with respect to the Skorohod metric by Theorem 5.2 in [6]. Then, it follows from Lemma B.2 in [8] that under $\mathbf{P}^*$, $\mathbb{H}_n$=$\mathbb{Z}_n + \sqrt{n/N}\mathbb{W}_N$ is tight in $(\tilde{D}[0,1], \tilde{\mathcal{D}})$ with respect to the Skorohod metric, for $d(i,s)$=$(N\pi_i)^{-1}$ and $Q(s,\omega)$ since Assumption 4.2.1 holds. It also follows from (iii) of Theorem 5.1 in [69] that

$$\mathbf{m}\big(\mathbb{Z}_n(t_1) + \sqrt{n/N}\mathbb{W}_N(t_1), \ldots, \mathbb{Z}_n(t_k) + \sqrt{n/N}\mathbb{W}_N(t_k)\big)^T \xrightarrow{\mathcal{L}} N(0, \mathbf{m}(\Gamma_3 + \lambda\Gamma_4)\mathbf{m}^T) \tag{4.7.13}$$

as $\nu \to \infty$ under $\mathbf{P}^*$ for $k \geq 1$ and $m \neq 0$ because $\mathbf{m}(\mathbb{Z}_n(t_1), \ldots, \mathbb{Z}_n(t_k))^T \xrightarrow{\mathcal{L}} N(0, \mathbf{m}\Gamma_3\mathbf{m}^T)$ as $\nu \to \infty$ under $Q(s,\omega)$ *a.s.* [**P**], and $\sqrt{n/N}\mathbf{m}(\mathbb{W}_N(t_1), \ldots, \mathbb{W}_N(t_k))^T \xrightarrow{\mathcal{L}} N(0, \lambda\mathbf{m}\Gamma_4\mathbf{m}^T)$ as $\nu \to \infty$ under **P**. Here, Γ_4 is a $k \times k$ matrix such that

$$((\Gamma_4))_{ij} = t_i \wedge t_j - t_i t_j \text{ for } 1 \leq i < j \leq k. \tag{4.7.14}$$

Therefore, under $\mathbf{P}^*$, $\mathbb{H}_n \xrightarrow{\mathcal{L}} \mathbb{H}$ in $(\tilde{D}[0,1], \tilde{\mathcal{D}})$ with respect to the Skorohod metric, for $d(i,s)$=$(N\pi_i)^{-1}$ and $Q(s,\omega)$, where $\mathbb{H}$ is a mean 0 Gaussian process with covariance kernel

$$\lim_{\nu\to\infty} E_{\mathbf{P}}\bigg(nN^{-2}\sum_{i=1}^{N}(\mathbb{1}_{[U_i \leq t_1]} - F_{u,N}(t_1) - R(t_1)\pi_i)\times (\mathbb{1}_{[U_i \leq t_2]} - F_{u,N}(t_2) - R(t_2)\pi_i)(\pi_i^{-1} - 1)\bigg) + \lambda(t_1 \wedge t_2 - t_1 t_2) \text{ for } t_1, t_2 \in [0,1]. \tag{4.7.15}$$

We can choose independent random functions, $\mathbb{H}_1, \mathbb{H}_2 \in \tilde{D}[0,1]$ defined on some probability space such that $\mathbb{H}_1 \stackrel{\mathcal{L}}{=} \mathbb{U}$ and $\mathbb{H}_2 \stackrel{\mathcal{L}}{=} \mathbb{W}$. Since $\mathbb{U}$ and $\mathbb{W}$ have almost sure continuous paths, $\mathbb{H}_1$ and $\mathbb{H}_2$ have almost sure continuous paths. Hence, $\mathbb{H}_1 + \sqrt{\lambda}\mathbb{H}_2$ has almost sure continuous paths. Next, note that $\mathbb{H}_1$ and $\mathbb{H}_2$ are mean 0 Gaussian processes because $\mathbb{U}$ and $\mathbb{W}$ are mean 0 Gaussian processes. Thus $\mathbb{H}_1 + \sqrt{\lambda}\mathbb{H}_2$ is a mean 0 Gaussian process. Also, note that the covariance kernel of $\mathbb{H}$ is the sum of covariance kernels of $\mathbb{U}$ and $\sqrt{\lambda}\mathbb{W}$. Thus the covariance kernel of $\mathbb{H}_1 + \sqrt{\lambda}\mathbb{H}_2$ is the same as that of $\mathbb{H}$. Therefore, $\mathbb{H}_1 + \sqrt{\lambda}\mathbb{H}_2 \stackrel{\mathcal{L}}{=} \mathbb{H}$. Hence, $\mathbb{H}$ has almost sure continuous paths. Then, under $\mathbf{P}^*$, $\mathbb{H}_n \xrightarrow{\mathcal{L}} \mathbb{H}$ in $(\tilde{D}[0,1], \tilde{\mathcal{D}})$ with respect to the sup norm metric, for $d(i,s)$=$(N\pi_i)^{-1}$ and $Q(s,\omega)$ by Skorohod representation theorem.

Finally, we shall show that the conclusion of Proposition 4.2.1 holds for the high entropy sampling design $P(s,\omega)$, which satisfies Assumption 4.2.2. Note that for $d(i,s)$=$(N\pi_i)^{-1}$, $E_{\mathbf{P}^*}(f(\mathbb{H}_n))$= $E_{\mathbf{P}}(\sum_{s\in\mathcal{S}} f(\mathbb{H}_n)Q(s,\omega)) \rightarrow \int f dP_{\mathbb{H}}$ as $\nu \rightarrow \infty$ given any bounded continuous (with respect to the sup norm metric) $\tilde{\mathcal{D}}$-measurable function f, where $P_{\mathbb{H}}$ is the probability distribution corresponding to $\mathbb{H}$. Then, it follows from Lemmas 2 and 3 in [4] that

$$\begin{aligned}&\left|\sum_{s\in\mathcal{S}} f(\mathbb{H}_n)(P(s,\omega) - Q(s,\omega))\right| \leq K_2 \sum_{s\in\mathcal{S}} |P(s,\omega) - Q(s,\omega)|\\&\leq K_2(2D(P||Q))^{1/2} \leq K_2(2D(P||R))^{1/2},\end{aligned} \tag{4.7.16}$$

for some constant $K_2 > 0$, where $R(s,\omega)$ is a rejective sampling design such that $D(P||R) \rightarrow 0$ as $\nu \rightarrow \infty$ *a.s.* [**P**]. This implies that $E_{\mathbf{P}}(\sum_{s\in\mathcal{S}} f(\mathbb{H}_n)P(s,\omega)) \rightarrow \int f dP_{\mathbb{H}}$ as $\nu \rightarrow \infty$ for $d(i,s)$=$(N\pi_i)^{-1}$ by DCT, and hence, the conclusion of Proposition 4.2.1 holds for the high entropy sampling design $P(s,\omega)$. □

Proof of Theorem 4.2.1. Recall $\mathbb{H}_n$ and $\mathbb{W}_n$ from (4.7.2) in the proof of Proposition 4.2.1, and suppose that $0 \leq t_1, \ldots, t_k \leq 1$ for some $k \geq 1$. Then, for $d(i,s)$=$(N\pi_i)^{-1}$, we have

$$\begin{aligned}&\mathbf{m}_1\big(\mathbb{H}_n(t_1), \ldots, \mathbb{H}_n(t_k)\big)^T + \sqrt{n/N}\mathbf{m}_2\big(\mathbb{W}_N(t_1), \ldots, \mathbb{W}_N(t_k)\big)^T\\&= \mathbf{m}_1\big(\mathbb{H}_n(t_1) - \sqrt{n/N}\mathbb{W}_N(t_1), \ldots, \mathbb{H}_n(t_k) - \sqrt{n/N}\mathbb{W}_N(t_k)\big)^T +\\&\sqrt{n/N}(\mathbf{m}_1 + \mathbf{m}_2)\big(\mathbb{W}_N(t_1), \ldots, \mathbb{W}_N(t_k)\big)^T = \mathbf{m}_1\big(\mathbb{Z}_n(t_1), \ldots, \mathbb{Z}_n(t_1)\big)^T +\\&\sqrt{n/N}(\mathbf{m}_1 + \mathbf{m}_2)\big(\mathbb{W}_N(t_1), \ldots, \mathbb{W}_N(t_k)\big)^T\end{aligned} \tag{4.7.17}$$

given any $\mathbf{m}_1, \mathbf{m}_2 \in \mathbb{R}^k$ and $\mathbf{m}_1, \mathbf{m}_2 \neq 0$, where $\mathbb{Z}_n$ is as in (4.7.2). Further, suppose that $P(s,\omega)$ denotes a high entropy sampling design satisfying Assumption 4.2.2. Then, it can be shown in the same way as the derivation of the result in (4.7.13) that under $\mathbf{P}^*$,

$$\begin{aligned}&\mathbf{m}_1\big(\mathbb{Z}_n(t_1), \ldots, \mathbb{Z}_n(t_1)\big)^T + \sqrt{n/N}(\mathbf{m}_1 + \mathbf{m}_2)\big(\mathbb{W}_N(t_1), \ldots, \mathbb{W}_N(t_k)\big)^T \xrightarrow{\mathcal{L}}\\&N\big(0, \mathbf{m}_1\Gamma_3\mathbf{m}_1^T + \lambda(\mathbf{m}_1 + \mathbf{m}_2)\Gamma_4(\mathbf{m}_1 + \mathbf{m}_2)^T\big)\end{aligned} \tag{4.7.18}$$

for $P(s,\omega)$. Here, Γ_3 is as in (4.7.10), and Γ_4 as in (4.7.14). Thus in view of (4.7.17) and (4.7.18), we have

$$\begin{aligned}&\big(\mathbb{H}_n(t_1),\ldots,\mathbb{H}_n(t_k),\sqrt{n/N}\mathbb{W}_N(t_1),\ldots,\sqrt{n/N}\mathbb{W}_N(t_k)\big)^T\\&\xrightarrow{\mathcal{L}} N_{2k}(0,\Gamma_5),\ \text{for } d(i,s)=(N\pi_i)^{-1} \text{ and } P(s,\omega) \text{ under } \mathbf{P}^*,\ \text{where}\end{aligned} \tag{4.7.19}$$

$$\Gamma_5=\begin{bmatrix}\Gamma_3+\lambda\Gamma_4 & \lambda\Gamma_4\\ \lambda\Gamma_4 & \lambda\Gamma_4\end{bmatrix}.$$

The result stated in (4.7.19) implies weak convergence of finite dimensional distributions of the process $(\mathbb{H}_n,\sqrt{n/N}\mathbb{W}_N)$ for $d(i,s)$=$(N\pi_i)^{-1}$. Recall from the 3^{rd} paragraph in the proof of Proposition 4.2.1 that under **P**,

$$\mathbb{W}_N=\left\{\sqrt{N}(F_{u,N}(t)-t): t\in[0,1]\right\}\xrightarrow{\mathcal{L}}\mathbb{W} \tag{4.7.20}$$

as $\nu\to\infty$ in $(\tilde{D}[0,1],\tilde{\mathcal{D}})$ with respect to the Skorohod metric, where $\mathbb{W}$ is the standard Brownian bridge in $\tilde{D}[0,1]$ and has almost sure continuous paths. Then, $(\mathbb{H}_n,\sqrt{n/N}\mathbb{W}_N)$ is tight in $(\tilde{D}[0,1]\times\tilde{D}[0,1],\tilde{\mathcal{D}}\times\tilde{\mathcal{D}})$ with respect to the Skorohod metric, for $d(i,s)$=$(N\pi_i)^{-1}$ and $P(s,\omega)$ because both $\mathbb{H}_n$ and $\sqrt{n/N}\mathbb{W}_N$ are tight in $(\tilde{D}[0,1],\tilde{\mathcal{D}})$ with respect to the Skorohod metric, for $d(i,s)$=$(N\pi_i)^{-1}$ and $P(s,\omega)$ in view of (4.7.20) and Proposition 4.2.1. Therefore, under $\mathbf{P}^*$,

$$(\mathbb{H}_n,\sqrt{n/N}\mathbb{W}_N)\xrightarrow{\mathcal{L}}\mathbb{V}=(\mathbb{V}_1,\mathbb{V}_2) \tag{4.7.21}$$

as $\nu\to\infty$ in $(\tilde{D}[0,1]\times\tilde{D}[0,1],\tilde{\mathcal{D}}\times\tilde{\mathcal{D}})$ with respect to the Skorohod metric, for $d(i,s)$=$(N\pi_i)^{-1}$ and $P(s,\omega)$, where $\mathbb{V}$ is a mean 0 Gaussian process in $\tilde{D}[0,1]\times\tilde{D}[0,1]$ with almost sure continuous paths. The covariance kernel of $\mathbb{V}$ is obtained from Γ_5 above. Next, recall from the paragraph preceding Assumption 4.2.1 that F_y denotes the superpopulation distribution function of y. Then, by (4.7.21), continuous mapping theorem and Skorohod representation theorem, we have

$$(\mathbb{H}_n\circ F_y,\mathbb{W}_N\circ F_y)\xrightarrow{\mathcal{L}}(\mathbb{V}_1\circ F_y,\mathbb{V}_2\circ F_y)\text{ as }\nu\to\infty \tag{4.7.22}$$

in $(\tilde{D}(\mathbb{R})\times\tilde{D}(\mathbb{R}),\tilde{\mathcal{D}}_{\mathbb{R}}\times\tilde{\mathcal{D}}_{\mathbb{R}})$ with respect to the sup norm metric, for $d(i,s)$=$(N\pi_i)^{-1}$ and $P(s,\omega)$. Here, $\tilde{D}(\mathbb{R})$ denotes the class of all bounded right continuous functions defined on $\mathbb{R}$ with finite left limits, and $\tilde{\mathcal{D}}_{\mathbb{R}}$ denotes the σ-field on $\tilde{D}(\mathbb{R})$ generated by the open balls (ball σ-field) with respect to the sup norm metric. Note that $(\mathbb{V}_1\circ F_y,\mathbb{V}_2\circ F_y)$ has almost sure continuous paths because F_y is continuous by Assumption 4.2.3. Let us now consider the quantile map

$$\phi(F) = F^{-1} = Q \text{ for any distribution function F,} \tag{4.7.23}$$

where $F^{-1}(p)$=$Q(p)$=$\inf\{t \in \mathbb{R} : F(t) \geq p\}$ for any $0 < p < 1$. Now, suppose that $\tilde{D}$ denotes the set of distribution functions on $\mathbb{R}$ restricted to $[Q_y(\alpha)-\epsilon, Q_y(\beta)+\epsilon]$ for some $0 < \alpha < \beta < 1$ and $\epsilon > 0$, where Q_y is the superpopulation quantile function of y. Then, it can be shown in the same way as the proof of Lemma 3.9.23–*(i)* in [84] that ϕ: $\tilde{D} \subset \tilde{D}[Q_y(\alpha) - \epsilon, Q_y(\beta) + \epsilon] \to D[\alpha, \beta]$ is Hadamard differentiable at F_y tangentially to $C[Q_y(\alpha) - \epsilon, Q_y(\beta) + \epsilon]$. Note that

$$\begin{aligned} &\mathbb{H}_n \circ F_y = \{\sqrt{n}(\hat{F}_y(t) - F_y(t)) : t \in \mathbb{R}\} \text{ and} \\ &\sqrt{n/N}\,\mathbb{W}_N \circ F_y = \{\sqrt{n}(F_{y,N}(t) - F_y(t)) : t \in \mathbb{R}\}, \end{aligned} \tag{4.7.24}$$

where $\hat{F}_y(t)$=$\sum_{i\in s} d(i,s)\mathbb{1}_{[Y_i \leq t]}/\sum_{i \in s} d(i,s)$ and $F_{y,N}(t)$=$\sum_{i=1}^{N} \mathbb{1}_{[Y_i \leq t]}/N$. This is because F_y is continuous by Assumption 4.2.3. Then by (4.7.22), (4.7.24), functional delta method (see Theorem 3.9.4 in [84]) and Hadamard differentiability of ϕ, we have

$$\begin{aligned} &\big(\{\sqrt{n}(\hat{Q}_y(p) - Q_y(p)) : p \in [\alpha, \beta]\}, \{\sqrt{n}(Q_{y,N}(p) - Q_y(p)) : p \in [\alpha, \beta]\}\big) \xrightarrow{\mathcal{L}} \\ &(-\tilde{\mathbb{V}}_1, -\tilde{\mathbb{V}}_2)/f_y \circ Q_y \end{aligned} \tag{4.7.25}$$

as $\nu \to \infty$ in $(D[\alpha,\beta]\times D[\alpha,\beta], \mathcal{D}\times\mathcal{D})$ with respect to the sup norm metric, for $d(i,s)$=$(N\pi_i)^{-1}$ and $P(s,\omega)$. Here, f_y is the superpopulation density function of y, $(\tilde{\mathbb{V}}_1, \tilde{\mathbb{V}}_2)$ is a mean 0 Gaussian process in $D[\alpha,\beta] \times D[\alpha,\beta]$, and $(\tilde{\mathbb{V}}_1, \tilde{\mathbb{V}}_2) \stackrel{\mathcal{L}}{=} (\mathbb{V}_1, \mathbb{V}_2)$. Then, by continuous mapping theorem, we have

$$\{\sqrt{n}(\hat{Q}_y(p) - Q_{y,N}(p)) : p \in [\alpha, \beta]\} \xrightarrow{\mathcal{L}} -(\tilde{\mathbb{V}}_1 - \tilde{\mathbb{V}}_2)/f_y \circ Q_y = \mathbb{Q} \text{ (say)} \tag{4.7.26}$$

as $\nu \to \infty$ in $(D[\alpha,\beta], \mathcal{D})$ with respect to the sup norm metric, for $d(i,s)$=$(N\pi_i)^{-1}$ and $P(s,\omega)$. The covariance kernel of $\mathbb{Q}$ is obtained from the matrix

$$\begin{bmatrix} I_k & -I_k \end{bmatrix} \Gamma_5 \begin{bmatrix} I_k \\ -I_k \end{bmatrix} = \Gamma_3.$$

Here, I_k is the $k \times k$ identity matrix.

We shall next show the weak convergence of the quantile processes constructed based on $\hat{Q}_{y,RA}(p)$, $\hat{Q}_{y,DI}(p)$ and $\hat{Q}_{y,REG}(p)$ in $(D[\alpha,\beta], \mathcal{D})$ with respect to the sup norm metric, for $d(i,s)$=$(N\pi_i)^{-1}$ and $P(s,\omega)$. Recall $\hat{Q}_x$ and $Q_{x,N}$ from Section 4.1, and Q_x from the paragraph preceding Assumption 4.2.1. Note that

$$\begin{aligned}\sqrt{n}\big(\hat{Q}_{y,RA}(p) - Q_{y,N}(p)\big) = \sqrt{n}\big(\hat{Q}_y(p) - Q_y(p)\big) - \sqrt{n}\big(Q_{y,N}(p) - Q_y(p)\big) + \\ \big(\hat{Q}_y(p)/\hat{Q}_x(p)\big)\big\{\sqrt{n}\big(Q_{x,N}(p) - Q_x(p)\big) - \sqrt{n}\big(\hat{Q}_x(p) - Q_x(p)\big)\big\}.\end{aligned} \tag{4.7.27}$$

First, it can be shown in the same way as the derivation of the results in (4.7.22) and (4.7.25) that under $\mathbf{P}^*$, $\big(\{\sqrt{n}(\hat{F}_y(t) - F_y(t)) : t \in \mathbb{R}\}, \{\sqrt{n}(F_{y,N}(t) - F_y(t)) : t \in \mathbb{R}\}, \{\sqrt{n}(\hat{F}_x(t) - F_x(t)) : t \in \mathbb{R}\}, \{\sqrt{n}(F_{x,N}(t) - F_x(t)) : t \in \mathbb{R}\}\big)$ converges weakly to some mean 0 Gaussian process with almost sure continuous paths as $\nu \to \infty$, and hence $\big(\{\sqrt{n}(\hat{Q}_y(p) - Q_y(p)) : p \in [\alpha, \beta]\}, \{\sqrt{n}(Q_{y,N}(p) - Q_y(p)) : p \in [\alpha, \beta]\}, \{\sqrt{n}(\hat{Q}_x(p) - Q_x(p)) : p \in [\alpha, \beta]\}, \{\sqrt{n}(Q_{x,N}(p) - Q_x(p)) : p \in [\alpha, \beta]\}\big)$ converges weakly to some mean 0 Gaussian process with almost sure continuous paths as $\nu \to \infty$. Then, we have

$$\sup_{p \in [\alpha, \beta]} |\hat{Q}_y(p)/\hat{Q}_x(p) - Q_y(p)/Q_x(p)| \xrightarrow{p} 0 \tag{4.7.28}$$

as $\nu \to \infty$ under $\mathbf{P}^*$. Further, it can be shown in the same way as the derivation of the result in (4.7.26) that under $\mathbf{P}^*$,

$$\begin{aligned}\big\{\sqrt{n}(\hat{Q}_y(p) - Q_{y,N}(p)) + \big(Q_y(p)/Q_x(p)\big) \times \\ \sqrt{n}\big(Q_{x,N}(p) - \hat{Q}_x(p)\big) : p \in [\alpha, \beta]\big\} \xrightarrow{\mathcal{L}} \mathbb{Q} \text{ as } \nu \to \infty\end{aligned} \tag{4.7.29}$$

in $(D[\alpha, \beta], \mathcal{D})$ with respect to the sup norm metric, for $d(i, s)$=$(N\pi_i)^{-1}$ and $P(s, \omega)$. Here, $\mathbb{Q}$ is a mean 0 Gaussian process in $\tilde{D}[\alpha, \beta]$ with almost sure continuous paths. Therefore, in view of (4.7.27)–(4.7.29),

$$\big\{\sqrt{n}(\hat{Q}_{y,RA}(p) - Q_{y,N}(p)) : p \in [\alpha, \beta]\big\} \xrightarrow{\mathcal{L}} \mathbb{Q} \text{ as } \nu \to \infty \tag{4.7.30}$$

in $(D[\alpha, \beta], \mathcal{D})$ with respect to the sup norm metric, for $d(i, s)$=$(N\pi_i)^{-1}$ and $P(s, \omega)$ under $\mathbf{P}^*$. The covariance kernel of $\mathbb{Q}$ is obtained from the asymptotic covariance kernel of $\big(\{\sqrt{n}(\hat{F}_y(t) - F_y(t)) : t \in \mathbb{R}\}, \{\sqrt{n}(F_{y,N}(t) - F_y(t)) : t \in \mathbb{R}\}, \{\sqrt{n}(\hat{F}_x(t) - F_x(t)) : t \in \mathbb{R}\}, \{\sqrt{n}(F_{x,N}(t) - F_x(t)) : t \in \mathbb{R}\}\big)$. Next, note that

$$\begin{aligned}\sqrt{n}\big(\hat{Q}_{y,DI}(p) - Q_{y,N}(p)\big) = \sqrt{n}\big(\hat{Q}_y(p) - Q_y(p)\big) - \sqrt{n}\big(Q_{y,N}(p) - Q_y(p)\big) + \\ \left(\sum_{i \in s} \pi_i^{-1} Y_i \Big/ \sum_{i \in s} \pi_i^{-1} X_i\right)\big\{\sqrt{n}\big(Q_{x,N}(p) - Q_x(p)\big) - \sqrt{n}\big(\hat{Q}_x(p) - Q_x(p)\big)\big\}\end{aligned} \tag{4.7.31}$$

and

$$\sqrt{n}\big(\hat{Q}_{y,REG}(p)-Q_{y,N}(p)\big)=\sqrt{n}\big(\hat{Q}_y(p)-Q_y(p)\big)-\sqrt{n}\big(Q_{y,N}(p)-Q_y(p)\big)+ \\ \left(\sum_{i\in s}\pi_i^{-1}X_iY_i\Big/\sum_{i\in s}\pi_i^{-1}X_i^2\right)\big\{\sqrt{n}\big(Q_{x,N}(p)-Q_x(p)\big)-\sqrt{n}\big(\hat{Q}_x(p)-Q_x(p)\big)\big\}. \tag{4.7.32}$$

It can be shown using Theorem 6.1 in [40] and similar arguments in the last paragraph of the proof of Proposition 4.2.1 that under $P(s,\omega)$,

$$\sum_{i\in s}(N\pi_i)^{-1}\mathbf{W}_i-\sum_{i=1}^{N}\mathbf{W}_i/N\xrightarrow{p}0 \text{ as } \nu\to\infty \textit{ a.s. } [\mathbf{P}] \tag{4.7.33}$$

because $E_{\mathbf{P}}||\mathbf{W}_i||^2<\infty$. Here, $\mathbf{W}_i$=(X_i,Y_i,X_iY_i,X_i^2). Since $\sum_{i=1}^{N}\mathbf{W}_i/N\to E_{\mathbf{P}}(\mathbf{W}_i)$ as $\nu\to\infty$ *a.s.* [**P**] by SLLN, we have

$$\sum_{i\in s}\pi_i^{-1}Y_i\Big/\sum_{i\in s}\pi_i^{-1}X_i\xrightarrow{p}E_{\mathbf{P}}(Y_i)/E_{\mathbf{P}}(X_i) \text{ and} \\ \sum_{i\in s}\pi_i^{-1}X_iY_i\Big/\sum_{i\in s}\pi_i^{-1}X_i^2\xrightarrow{p}E_{\mathbf{P}}(X_iY_i)/E_{\mathbf{P}}(X_i)^2 \tag{4.7.34}$$

as $\nu\to\infty$ for $P(s,\omega)$ under $\mathbf{P}^*$. Therefore, using (4.7.31), (4.7.32) and similar arguments as in the case of $\{\sqrt{n}(\hat{Q}_{y,RA}(p)-Q_{y,N}(p)):p\in[\alpha,\beta]\}$, we can say that under $\mathbf{P}^*$, $\{\sqrt{n}(\hat{Q}_{y,DI}(p)-Q_{y,N}(p)):p\in[\alpha,\beta]\}$ and $\{\sqrt{n}(\hat{Q}_{y,REG}(p)-Q_{y,N}(p)):p\in[\alpha,\beta]\}$ converge weakly to a mean 0 Gaussian process with almost sure continuous paths in $(D[\alpha,\beta],\mathcal{D})$ with respect to the sup norm metric, for $d(i,s)$=$(N\pi_i)^{-1}$ and $P(s,\omega)$. □

Before we give the proof of Proposition 4.2.2, recall $\{U_i\}_{i=1}^{N}$ from (4.2.2) and $F_{u,N}(t)$ from (4.7.1). Define $\tilde{\mathbb{U}}_n(t)$=$\sqrt{n}\sum_{i\in s}(NX_i)^{-1}G_i(\mathbb{1}_{[U_i\le t]}-F_{u,N}(t))$ for $0\le t\le 1$ and $\tilde{\mathbb{B}}_n(t_1,t_2)$=$\tilde{\mathbb{U}}_n(t_2)-\tilde{\mathbb{U}}_n(t_1)$ for $0\le t_1<t_2\le 1$.

Proof of Proposition 4.2.2. Using Lemmas 4.8.3 and 4.8.4 in Section 4.8, it can be shown in the same way as in the first two paragraphs of the proof of Proposition 3.1 that under $\mathbf{P}^*$, $\tilde{\mathbb{U}}_n\xrightarrow{\mathcal{L}}\tilde{\mathbb{U}}$ as $\nu\to\infty$ in $(\tilde{D}[0,1],\tilde{\mathcal{D}})$ with respect to the sup norm metric, for RHC sampling design, where $\tilde{\mathbb{U}}$ is a mean 0 Gaussian process in $\tilde{D}[0,1]$ with almost sure continuous paths. Moreover, the covariance kernel of $\tilde{\mathbb{U}}$ is

$$\lim_{\nu\to\infty}E_{\mathbf{P}}\left(n\gamma(\overline{X}/N)\sum_{i=1}^{N}(\mathbb{1}_{[U_i\le t_1]}-F_{u,N}(t_1))(\mathbb{1}_{[U_i\le t_2]}-F_{u,N}(t_2))X_i^{-1}\right). \tag{4.7.35}$$

It can be shown that under RHC sampling design,

$$var\Big(\sum_{i\in s}(NX_i)^{-1}G_i\Big) = \gamma\sum_{i=1}^{N}(X_i-\overline{X})^2/NX_i\overline{X} = \gamma\Big(\overline{X}\sum_{i=1}^{N}X_i/N-1\Big) \to 0 \quad (4.7.36)$$

as $\nu\to\infty$ *a.s.* [**P**] since $n\gamma\to c>0$ by Lemma 2.7.5 in Section 2.7 of Chapter 2, and Assumptions 4.2.4 and 4.2.6 hold. Consequently, under **P***, $\sum_{i\in s}(NX_i)^{-1}G_i \xrightarrow{p} 1$ as $\nu\to\infty$. Therefore, under **P***,

$$\tilde{\mathbb{Z}}_n = \tilde{\mathbb{U}}_n\Big/\sum_{i\in s}(NX_i)^{-1}G_i \xrightarrow{\mathcal{L}} \tilde{\mathbb{Z}} \overset{\mathcal{L}}{=} \tilde{\mathbb{U}} \quad (4.7.37)$$

as $\nu\to\infty$ in $(\tilde{D}[0,1],\tilde{\mathcal{D}})$ with respect to the sup norm metric, for RHC sampling design. Next, note that

$$\mathbb{H}_n = \big\{\sqrt{n}(\hat{F}_u(t)-t) : t\in[0,1]\big\} = \tilde{\mathbb{Z}}_n + \sqrt{n/N}\,\mathbb{W}_N, \quad (4.7.38)$$

for $d(i,s)$=$(NX_i)^{-1}G_i$, where $\mathbb{W}_N$=$\{\sqrt{N}(F_{u,N}(t)-t) : t\in[0,1]\}$. Also, note that under **P**, $\mathbb{W}_N \xrightarrow{\mathcal{L}} \mathbb{W}$ as $\nu\to\infty$ in $(\tilde{D}[0,1],\tilde{\mathcal{D}})$ with respect to the Skorohod metric by Donsker theorem, where $\mathbb{W}$ is the standard Brownian bridge. Therefore, using the same arguments as in the 3^{rd} paragraph of the proof of Proposition 3.1, we can show that under **P***, $\mathbb{H}_n \xrightarrow{\mathcal{L}} \mathbb{H}$ as $\nu\to\infty$ in $(\tilde{D}[0,1],\tilde{\mathcal{D}})$ with respect to the sup norm metric, for $d(i,s)$=$(NX_i)^{-1}G_i$ and RHC sampling design, where $\mathbb{H}$ is a mean 0 Gaussian process with covariance kernel

$$\lim_{\nu\to\infty} E_{\mathbf{P}}\big(n\gamma(\overline{X}/N)\sum_{i=1}^{N}(\mathbb{1}_{[U_i\le t_1]}-F_{u,N}(t_1))(\mathbb{1}_{[U_i\le t_2]}-F_{u,N}(t_2))X_i^{-1}\big)+ \\ \lambda(t_1\wedge t_2 - t_1t_2), \quad (4.7.39)$$

for $t_1,t_2\in[0,1]$. Also, $\mathbb{H}$ has almost sure continuous paths. It can be shown using Lemma 4.8.3, Assumption 4.2.4 and DCT that the expression in (4.7.39) becomes

$$cE_{\mathbf{P}}(X_i)E_{\mathbf{P}}\Big(\big(\mathbb{1}_{[U_i\le t_1]}-\mathbf{P}(U_i\le t_1)\big)\big(\mathbb{1}_{[U_i\le t_2]}-\mathbf{P}(U_i\le t_2)\big)X_i^{-1}\Big)+ \\ \lambda(t_1\wedge t_2 - t_1t_2), \quad (4.7.40)$$

where c=$\lim_{\nu\to\infty} n\gamma$. □

Proof of Theorem 4.2.2. The proof follows in view of Proposition 4.2.2 in the same way as the proof of Theorem 4.2.1 follows in view of Proposition 4.2.1. □

Proof of Proposition 4.3.1. Let us denote the stratified multistage cluster sampling design by $P(s, \omega)$.

(i) Recall $F_{y,H}$ from the paragraph preceding Assumption 4.3.5, and consider $\{U_i\}_{i=1}^N$ as in (4.2.2) with $F_{y,H}$ replacing F_y. Also, recall $F_{u,N}(t)$ and $\mathbb{U}_n(t)$ from (4.7.1). Note that $F_{u,N}(t) \to t$ as $\nu \to \infty$ *a.s.* [**P**] for any $t \in [0,1]$ by Assumption 4.3.3 and SLLN. Therefore, using Lemmas 4.8.6 and 4.8.7 in Section 4.8, one can show in the same way as in the first two paragraphs of the proof of Proposition 4.2.1 that under $\mathbf{P}^*$,

$$\mathbb{U}_n \xrightarrow{\mathcal{L}} \mathbb{U} \text{ as } \nu \to \infty \tag{4.7.41}$$

in $(\tilde{D}[0,1], \tilde{\mathcal{D}})$ with respect to the sup norm metric, for $P(s, \omega)$. Here, $\mathbb{U}$ is a mean 0 Gaussian process in $\tilde{D}[0,1]$ with covariance kernel

$$\begin{aligned} K_1(t_1, t_2) = \lambda \sum_{h=1}^{H} \Lambda_h(\Lambda_h/\lambda\lambda_h - 1) E_{\mathbf{P}}\big(\mathbb{1}_{[Y'_{hjl} \le \tilde{Q}_{y,H}(t_1)]} - \mathbf{P}(Y'_{hjl} \le \tilde{Q}_{y,H}(t_1))\big) \times \\ \big(\mathbb{1}_{[Y'_{hjl} \le \tilde{Q}_{y,H}(t_2)]} - \mathbf{P}(Y'_{hjl} \le \tilde{Q}_{y,H}(t_2))\big) \end{aligned} \tag{4.7.42}$$

for $t_1, t_2 \in [0,1]$. Here, $\tilde{Q}_{y,H}(p)$=$\inf\{t \in \mathbb{R} : \tilde{F}_{y,H}(t) \ge p\}$, $\tilde{F}_{y,H}(t)$=$\sum_{h=1}^{H} \Lambda_h F_{y,h}(t)$, and λ_h's and Λ_h's are as in Assumption 4.3.1. Moreover, $\mathbb{U}$ has almost sure continuous paths. Next, it can be shown using Assumption 4.3.1 that $var(\sum_{i\in s}(N\pi_i)^{-1})$= $o(1)$, and hence $\sum_{i \in s}(N\pi_i)^{-1} \xrightarrow{p} 1$ as $\nu \to \infty$ under $P(s,\omega)$ for any given $\omega \in \Omega$. Here, π_i=$m_h r_h / M_h N_{hj}$ when the i^{th} population unit belongs to the j^{th} cluster of stratum h. Therefore, it follows from DCT that under $\mathbf{P}^*$, $\sum_{i\in s}(N\pi_i)^{-1} \xrightarrow{p} 1$, and hence under $\mathbf{P}^*$,

$$\mathbb{Z}_n = \mathbb{U}_n \Big/ \sum_{i \in s}(N\pi_i)^{-1} \xrightarrow{\mathcal{L}} \mathbb{Z} \overset{\mathcal{L}}{=} \mathbb{U} \tag{4.7.43}$$

as $\nu \to \infty$ in $(\tilde{D}[0,1], \tilde{\mathcal{D}})$ with respect to the sup norm metric, for the sampling design $P(s, \omega)$.

Next, recall $\mathbb{W}_N$ from the 1^{st} paragraph in the proof of Proposition 4.2.1. Then, using assumptions Assumptions 4.3.1 and 4.3.3, and Lemma 4.8.8 in Section 4.8, it can be shown that

$$\begin{aligned} cov_{\mathbf{P}}(\mathbb{W}_N(t_1), \mathbb{W}_N(t_2)) = \sum_{h=1}^{H}(N_h/N) E_{\mathbf{P}}\bigg(\mathbb{1}_{[Y'_{hjl} \le Q_{y,H}(t_1)]} - \\ \mathbf{P}(Y'_{hjl} \le Q_{y,H}(t_1))\bigg)\bigg(\mathbb{1}_{[Y'_{hjl} \le Q_{y,H}(t_2)]} - \mathbf{P}(Y'_{hjl} \le Q_{y,H}(t_2))\bigg) \to \end{aligned} \tag{4.7.44}$$

$$\sum_{h=1}^{H} \Lambda_h E_{\mathbf{P}}\Big(\mathbb{1}_{[Y'_{hjl} \leq \tilde{Q}_{y,H}(t_1)]} - \mathbf{P}(Y'_{hjl} \leq \tilde{Q}_{y,H}(t_1))\Big)\Big(\mathbb{1}_{[Y'_{hjl} \leq \tilde{Q}_{y,H}(t_2)]} -$$
$$\mathbf{P}(Y'_{hjl} \leq \tilde{Q}_{y,H}(t_2))\Big) = K_2(t_1, t_2) \text{ (say)}$$

as $\nu \to \infty$ for any $t_1, t_2 \in [0,1]$. Then, under **P**, $\mathbb{W}_N \xrightarrow{\mathcal{L}} \mathbb{W}$ as $\nu \to \infty$ in $(\tilde{D}[0,1], \tilde{\mathcal{D}})$ with respect to the Skorohod metric by (4.7.44) above and Theorem 3.3.1 in [79] (see p. 109), where $\mathbb{W}$ is a mean 0 Gaussian process in $\tilde{D}[0,1]$ with covariance kernel $K_2(t_1,t_2)$. Also, $\mathbb{W}$ has almost sure continuous paths. Therefore, using similar arguments as in the proof of Proposition 4.2.1, we can say that under $\mathbf{P}^*$,

$$\mathbb{H}_n = \mathbb{Z}_n + \sqrt{n/N}\,\mathbb{W}_N = \{\sqrt{n}(\hat{F}_u(t) - t) : t \in [0,1]\} \xrightarrow{\mathcal{L}} \mathbb{H} \tag{4.7.45}$$

as $\nu \to \infty$ in $(\tilde{D}[0,1], \tilde{\mathcal{D}})$ with respect to the sup norm metric, for $d(i,s)$=$(N\pi_i)^{-1}$ and $P(s,\omega)$, where $\mathbb{H}$ is a mean 0 Gaussian process in $\tilde{D}[0,1]$ with covariance kernel

$$K_1(t_1,t_2) + \lambda K_2(t_1,t_2). \tag{4.7.46}$$

Moreover, $\mathbb{H}$ has almost sure continuous paths. This completes the proof of *(i)*.

(ii) Using Hoeffding's inequality (see [76]), and Assumptions 4.2.1, 4.3.3 and 4.3.4, it can be shown that $F_{u,N}(t) \to t$ as $\nu \to \infty$ *a.s.* [**P**] for any $t \in [0,1]$. Therefore, using Lemmas 4.8.6 and 4.8.7, and the Assumption 4.3.4, one can show in the same way as in *(i)* that under $\mathbf{P}^*$,

$$\mathbb{Z}_n \xrightarrow{\mathcal{L}} \mathbb{U} \text{ as } \nu \to \infty \tag{4.7.47}$$

in $(\tilde{D}[0,1], \tilde{\mathcal{D}})$ with respect to the sup norm metric, for $P(s,\omega)$, where $\mathbb{U}$ is a mean 0 Gaussian process in $\tilde{D}[0,1]$ with covariance kernel

$$K_1(t_1,t_2) = \lim_{\nu\to\infty} \lambda \sum_{h=1}^{H} N_h(N_h - n_h) E_{\mathbf{P}}\big(\mathbb{1}_{[Y'_{hjl} \leq Q_{y,H}(t_1)]} -$$
$$\mathbf{P}(Y'_{hjl} \leq Q_{y,H}(t_1))\big)\big(\mathbb{1}_{[Y'_{hjl} \leq Q_{y,H}(t_2)]} - \mathbf{P}(Y'_{hjl} \leq Q_{y,H}(t_2))\big)/n_h N, \tag{4.7.48}$$

for $t_1, t_2 \in [0,1]$. Moreover, $\mathbb{U}$ has almost sure continuous paths. Next, given any $t_1, t_2 \in [0,1]$,

$$cov_{\mathbf{P}}(\mathbb{W}_N(t_1), \mathbb{W}_N(t_2)) = \sum_{h=1}^{H}(N_h/N)E_{\mathbf{P}}\big(\mathbb{1}_{[Y'_{hjl}\leq Q_{y,H}(t_1)]}- \mathbf{P}(Y'_{hjl} \leq Q_{y,H}(t_1))\big)\big(\mathbb{1}_{[Y'_{hjl}\leq Q_{y,H}(t_2)]} - \mathbf{P}(Y'_{hjl} \leq Q_{y,H}(t_2))\big) \to K_2(t_1, t_2) \tag{4.7.49}$$

as $\nu \to \infty$ for some covariance kernel $K_2(t_1, t_2)$ by Assumption 4.3.5. Then, under **P**,

$$\mathbb{W}_N \xrightarrow{\mathcal{L}} \mathbb{W} \text{ as } \nu \to \infty \tag{4.7.50}$$

in $(\tilde{D}[0, 1], \tilde{\mathcal{D}})$ with respect to the Skorohod metric by Theorem 3.3.1 in [79] (see p. 109), where $\mathbb{W}$ is a mean 0 Gaussian process in $\tilde{D}[0, 1]$ with covariance kernel $K_2(t_1, t_2)$. Also, $\mathbb{W}$ has almost sure continuous paths. Therefore, using similar arguments as in the proof of Proposition 4.2.1, we can say that under $\mathbf{P}^*$,

$$\mathbb{H}_n = \mathbb{Z}_n + \sqrt{n/N}\mathbb{W}_N = \{\sqrt{n}(\hat{F}_u(t) - t) : t \in [0, 1]\} \xrightarrow{\mathcal{L}} \mathbb{H} \text{ as } \nu \to \infty \tag{4.7.51}$$

in $(\tilde{D}[0, 1], \tilde{\mathcal{D}})$ with respect to the sup norm metric, for $d(i, s)$=$(N\pi_i)^{-1}$ and $P(s, \omega)$, where $\mathbb{H}$ is a mean 0 Gaussian process in $\tilde{D}[0, 1]$ with almost sure continuous paths and p.d. covariance kernel

$$K_1(t_1, t_2) + \lambda K_2(t_1, t_2). \tag{4.7.52}$$

This completes the proof of *(ii)*. □

Proof of Theorem 4.3.1. The proof follows in view of Proposition 4.3.1 in the same way as the proof of Theorem 4.2.1 follows in view of Proposition 4.2.1. □

Proof of Theorem 4.4.1. By conclusions of Theorems 4.2.1 and 4.2.2, and continuous mapping theorem, we have

$$\int_{\alpha}^{\beta} \sqrt{n}(G(p) - Q_{y,N}(p))J(p)dp \xrightarrow{\mathcal{L}} \int_{\alpha}^{\beta} \mathbb{Q}(p)J(p)dp \text{ as } \nu \to \infty \tag{4.7.53}$$

for high entropy and RHC sampling deigns under $\mathbf{P}^*$. Note that $\mathbb{Q}(p)J(p)$ is Riemann integrable on $[\alpha, \beta]$ implying Z=$\int_{\alpha}^{\beta} \mathbb{Q}(p)J(p)dp$=$\lim_{m\to\infty} m^{-1} \sum_{i=0}^{m-1} \mathbb{Q}\big(\alpha + i(\beta - \alpha)/m\big)J\big(\alpha + i(\beta - \alpha)/m\big)$ under the aforementioned sampling designs. By DCT, we have

$$E(\exp(itZ)) = \lim_{m\to\infty} \exp\Big\{ -m^{-2}\sum_{i=0}^{m-1}\sum_{j=0}^{m-1} K\big(\alpha + i(\beta-\alpha)/m,$$
$$\alpha + j(\beta-\alpha)/m\big)J\big(\alpha+i(\beta-\alpha)/m\big)J\big(\alpha+j(\beta-\alpha)/m\big)(t^2/2)\Big\} \quad (4.7.54)$$

since $\mathbb{Q}$ is a mean 0 Gaussian process in $D[\alpha,\beta]$ with covariance kernel $K(p_1,p_2)$. Note that $K(p_1,p_2)$ in the case of any high entropy sampling design (see (4.2.3)) is continuous on $[\alpha,\beta]\times[\alpha,\beta]$ by the assumption of this theorem, whereas $K(p_1,p_2)$ in the case of RHC sampling design (see (4.2.5)) is continuous on $[\alpha,\beta]\times[\alpha,\beta]$ by Assumption 4.2.3. Then, $E(\exp(itZ))$=exp $\big(-t^2\int_\alpha^\beta\int_\alpha^\beta K(p_1,p_2)J(p_1)J(p_2)dp_1dp_2/2\big)$ under the above sampling designs since $K(p_1,p_2)$ is continuous on $[\alpha,\beta]\times[\alpha,\beta]$, and hence Riemann integrable on $[\alpha,\beta]\times[\alpha,\beta]$. Therefore,

$$\int_\alpha^\beta \mathbb{Q}(p)J(p)dp \sim N(0,\sigma_1^2), \text{ where } \sigma_1^2 = \int_\alpha^\beta\int_\alpha^\beta K(p_1,p_2)J(p_1)J(p_2)dp_1dp_2. \quad (4.7.55)$$

Hence, under $\mathbf{P}^*$, $\int_\alpha^\beta \sqrt{n}(G(p)-Q_{y,N}(p))J(p)dp \xrightarrow{\mathcal{L}} N(0,\sigma_1^2)$ as $\nu\to\infty$ for high entropy and RHC sampling deigns.

Next, for any $k\geq 1$ and $p_1,\ldots,p_k\in[\alpha,\beta]$, we have

$$\sqrt{n}(f(G(p_1),\ldots,G(p_k)) - f(Q_{y,N}(p_1),\ldots,Q_{y,N}(p_k))) = a_N\sqrt{n}T_n + \sqrt{n}\epsilon(T_n) \quad (4.7.56)$$

by delta method, where a_N=$\nabla f(Q_{y,N}\,(p_1),\ldots,Q_{y,N}(p_k))$, T_n=$G(p_k)-Q_{y,N}(p_1),\ldots,G(p_k)-Q_{y,N}(p_k)$, and $\epsilon(T_n)\to 0$ as $T_n\to 0$. It follows from conclusions of Theorems 4.2.1 and 4.2.2 that under $\mathbf{P}^*$

$$\sqrt{n}T_n \xrightarrow{\mathcal{L}} N_k(0,\Delta) \text{ as } \nu\to\infty \quad (4.7.57)$$

for high entropy and RHC sampling deigns, where Δ is a $k\times k$ matrix such that $((\Delta))_{ij}$=$K(p_i,p_j)$ for $1\leq i,j\leq k$. It can be shown that $Q_{y,N}(p)\to Q_y(p)$ as $\nu\to\infty$ *a.s.* [**P**] for any $p\in(0,1)$, when $\{(Y_i,X_i) : 1\leq i\leq N\}$ are i.i.d. Thus $a_N\to a$ as $\nu\to\infty$ *a.s.* [**P**] for some a. Consequently, under $\mathbf{P}^*$, $\sqrt{n}(f(G(p_1),\ldots,G(p_k))-f(Q_{y,N}(p_1),\ldots,Q_{y,N}(p_k)))\xrightarrow{\mathcal{L}} N(0,\sigma_2^2)$ as $\nu\to\infty$ for the aforesaid sampling designs, where σ_2^2=$a\Delta a^T$. This completes the proofs of *(i)* and *(ii)*. □

Proof of Theorem 4.4.2. It can be shown using Assumptions 4.2.1, 4.3.1 and 4.3.3, and Lemma 4.8.8 in Section 4.8 that asymptotic covariance kernels of the quantile processes considered in this chapter under stratified multistage cluster sampling design with SRSWOR (see (4.3.1)) are continuous on $[\alpha,\beta]\times[\alpha,\beta]$, when H is fixed as $\nu\rightarrow\infty$. Moreover, by the assumption of this theorem, asymptotic covariance kernels of the aforementioned quantile processes are continuous on $[\alpha,\beta]\times[\alpha,\beta]$, when $H\rightarrow\infty$ as $\nu\rightarrow\infty$. Then, the asymptotic normality of $\int_{\alpha}^{\beta}\sqrt{n}(G(p)-Q_{y,N}(p))J(p)dp$ for the above sampling design under $\mathbf{P}^*$ can be shown using similar arguments as in the 1^{st} paragraph of the proof of Theorem 4.4.1.

Next, if H is fixed as $\nu\rightarrow\infty$, then it can be shown using A6 that $Q_{y,N}(p)\rightarrow\tilde{Q}_{y,H}(p)$ as $\nu\rightarrow\infty$ *a.s.* [$\mathbf{P}$] for any $p\in(0,1)$, where $\tilde{Q}_{y,H}(p)$=$\{t\in\mathbb{R}:\tilde{F}_{y,H}(t)\geq p\}$, $\tilde{F}_{y,H}(t)$=$\sum_{h=1}^{H}\Lambda_h F_{y,h}(t)$, and Λ_h's are as in Assumption 4.3.1. Further, if $H\rightarrow\infty$ as $\nu\rightarrow\infty$, then it can be shown using Assumption 4.3.6 that $Q_{y,N}(p)\rightarrow\tilde{Q}_y(p)$ as $\nu\rightarrow\infty$ *a.s.* [$\mathbf{P}$] for any $p\in(0,1)$, where $\tilde{Q}_y(p)$=$\{t\in\mathbb{R}:\tilde{F}_y(t)\geq p\}$, and $\tilde{F}_y$ is as in Assumption 4.3.6. Thus $a_N\rightarrow a$ as $\nu\rightarrow\infty$ *a.s.* [$\mathbf{P}$] for some a, where a_N is as in the 2^{nd} paragraph of the proof of Theorem 4.4.1. Then, given any $k\geq 1$ and $p_1,\ldots,p_k\in[\alpha,\beta]$, the the asymptotic normality of $\sqrt{n}(f(G(p_1),\ldots,G(p_k))-f(Q_{y,N}(p_1),\ldots,Q_{y,N}(p_k)))$ for the above sampling design under $\mathbf{P}^*$ can be shown using similar arguments as in the 2^{nd} paragraph of the proof of Theorem 4.4.1. This completes the proofs of *(i)* and *(ii)*. □

Proof of Theorem 4.4.3. *(i)* We shall prove this theorem using (4.8.6) in Lemma 4.8.5 in Section 4.8. Fix $\epsilon>0$, and suppose that

$$B_\epsilon(s,\omega)=\{p_1,p_2\in[\alpha,\beta]:|\hat{K}(p_1,p_2)-K(p_1,p_2)|\leq\epsilon\}\text{ for }s\in\mathcal{S}\text{ and }\omega\in\Omega. \quad (4.7.58)$$

Then, we have

$$\begin{aligned}&\int_{\alpha}^{\beta}\int_{\alpha}^{\beta}|(\hat{K}(p_1,p_2)-K(p_1,p_2))J(p_1)J(p_2)|dp_1dp_2\leq K\Big(\iint_{B_\epsilon(s,\omega)}|\hat{K}(p_1,p_2)-\\&K(p_1,p_2)|dp_1dp_2+\iint_{(B_\epsilon(s,\omega))^c}|\hat{K}(p_1,p_2)-K(p_1,p_2)|dp_1dp_2\Big)\\&\leq K\big(\epsilon(\beta-\alpha)^2+\iint_{(B_\epsilon(s,\omega))^c}|\hat{K}(p_1,p_2)-K(p_1,p_2)|dp_1dp_2\big)\end{aligned} \quad (4.7.59)$$

for some constant $K>0$ since J is continuous on $[\alpha,\beta]$. Now, let W_n=$\sup_{p_1,p_2\in[\alpha,\beta]}|\hat{K}(p_1,p_2)$ $-K(p_1,p_2)|$. Then,

$$\iint_{(B_\epsilon(s,\omega))^c}|\hat{K}(p_1,p_2)-K(p_1,p_2)|dp_1dp_2\leq \quad (4.7.60)$$

$$W_n \int_\alpha^\beta \int_\alpha^\beta \mathbb{1}_{[(B_\epsilon(s,\omega))^c]}(p_1,p_2)dp_1dp_2.$$

Further, under a high entropy sampling design,

$$\begin{aligned} &E_{\mathbf{P}^*}\Big(\int_\alpha^\beta \int_\alpha^\beta \mathbb{1}_{[(B_\epsilon(s,\omega))^c]}(p_1,p_2)dp_1dp_2\Big) = \int_\alpha^\beta \int_\alpha^\beta \mathbf{P}^*\big(|\hat{K}(p_1,p_2) - K(p_1,p_2)| \\ &> \epsilon\big)dp_1dp_2 \to 0 \text{ as } \nu \to \infty \text{ by DCT since } \hat{K}(p_1,p_2) \xrightarrow{p} K(p_1,p_2) \text{ as } \nu \to \infty \end{aligned} \tag{4.7.61}$$

for any $p_1, p_2 \in [\alpha, \beta]$ under $\mathbf{P}^*$ by (4.8.6) in Lemma 4.8.5 in Section 4.8. Therefore, under $\mathbf{P}^*$,

$$\begin{aligned} &\int_\alpha^\beta \int_\alpha^\beta \mathbb{1}_{[(B_\epsilon(s,\omega))^c]}(p_1,p_2)dp_1dp_2 \xrightarrow{p} 0, \text{ and} \\ &\iint_{(B_\epsilon(s,\omega))^c} |\hat{K}(p_1,p_2) - K(p_1,p_2)|dp_1dp_2 \xrightarrow{p} 0 \text{ as } \nu \to \infty \end{aligned} \tag{4.7.62}$$

for a high entropy sampling design because W_n=$O_p(1)$ as $\nu \to \infty$ by (4.8.6) in Lemma 4.8.5. Hence, $\int_\alpha^\beta \int_\alpha^\beta |(\hat{K}(p_1,p_2) - K(p_1,p_2))\ J(p_1)J(p_2)|dp_1dp_2 \xrightarrow{p} 0$ as $\nu \to \infty$ under $\mathbf{P}^*$. This completes the proof of the first part of *(i)*. The proof of the other part of *(i)* follows in a straight forward way. Also, the proof of *(ii)* follows exactly the same way as the proof of *(i)*. □

Proof of Theorem 4.4.4. The proof follows exactly the same way as the proof of Theorem 4.4.3 in view of Lemma 4.8.9 in Section 4.8. □

Proof of Theorem 4.5.1. *(i)* Suppose that δ_1^2, δ_2^2, δ_3^2 and δ_4^2 are the asymptotic variances of the estimators of $\int_\alpha^\beta Q_{y,N}(p)J(p)dp$ based on $\hat{Q}_y(p)$, $\hat{Q}_{y,RA}(p)$, $\hat{Q}_{y,DI}(p)$ and $\hat{Q}_{y,REG}(p)$, respectively, under $P(s,\omega)$. Here, $P(s,\omega)$ denotes one of SRSWOR, RHC and any HEπPS sampling designs. It follows from Lemma 4.8.10 in Section 4.8 that Assumption 4.2.2 holds under SRSWOR and any HEπPS sampling designs by the assumptions of Theorem 4.5.1. Then, in view of Theorem 4.4.1, we have

$$\delta_i^2 = \int_\alpha^\beta \int_\alpha^\beta K_i(p_1,p_2)J(p_1)J(p_2)dp_1dp_2 \text{ for } 1 \le i \le 4 \tag{4.7.63}$$

where $K_i(p_1,p_2)$'s are as in the paragraph preceding Theorem 4.5.1. Therefore, the conclusion of *(i)* in Theorem 4.5.1 holds in a straightforward way.

(ii) The proof follows exactly the same way as the proof of *(i)*. □

Proof of Theorem 4.5.2. *(i)* Suppose that η_1^2, η_2^2 and η_3^2 are the asymptotic variances of the estimators of $\int_\alpha^\beta Q_{y,N}(p)J(p)dp$ based on $G(p)$ under SRSWOR, RHC and any HEπPS sampling designs, respectively. Here, $G(p)$ denotes one of $\hat{Q}_y(p)$, $\hat{Q}_{y,RA}(p)$, $\hat{Q}_{y,DI}(p)$ and $\hat{Q}_{y,REG}(p)$. Then, in view of Theorem 4.4.1, we have

$$\eta_i^2 = \int_\alpha^\beta \int_\alpha^\beta K_i^*(p_1, p_2)J(p_1)J(p_2)dp_1dp_2 \text{ for } 1 \leq i \leq 3 \tag{4.7.64}$$

where $K_i^*(p_1, p_2)$'s are as in the paragraph preceding Theorem 4.5.2. Therefore, the conclusion of *(i)* in Theorem 4.5.2 holds in a straightforward way.

(ii) The proof follows exactly the same way as the proof of *(i)*. □

Proof of Theorem 4.5.3. It follows from (4.7.25) in the proof of Theorem 4.2.1 that under $\mathbf{P}^*$

$$\{\sqrt{n}(\hat{Q}_y(p) - Q_y(p)) :\in [\alpha, \beta]\} \xrightarrow{\mathcal{L}} -\tilde{\mathbb{V}}_1/f_y \circ Q_y \tag{4.7.65}$$

as $\nu \to \infty$ in $(D[\alpha, \beta], \mathcal{D})$ with respect to the sup norm metric, for $d(i, s)$=$(N\pi_i)^{-1}$ and SRSWOR. Here, Q_y and f_y are superpopulation quantile and density functions of y, respectively, and $\tilde{\mathbb{V}}_1$ is a mean 0 Gaussian process in $D[\alpha, \beta]$ with covariance kernel

$$\begin{aligned} K(p_1, p_2) &= \lim_{\nu\to\infty}(1 - n/N)E_{\mathbf{P}}\bigg(\sum_{i=1}^{N}(\mathbb{1}_{[Y_i \leq Q_y(p_1)]} - F_{y,N}(Q_y(p_1))\times \\ &(\mathbb{1}_{[Y_i \leq Q_y(p_2)]} - F_{y,N}(Q_y(p_2))/N\bigg) + \lambda(p_1 \wedge p_2 - p_1p_2) \\ &= p_1 \wedge p_2 - p_1p_2 \text{ for } p_1, p_2 \in [\alpha, \beta]. \end{aligned} \tag{4.7.66}$$

The result in (4.7.65) implies that under $\mathbf{P}^*$

$$\sqrt{n}(\hat{Q}_y(0.5) - Q_y(0.5)) \xrightarrow{\mathcal{L}} N(0, \sigma_1^2) \text{ as } \nu \to \infty \tag{4.7.67}$$

for $d(i, s)$=$(N\pi_i)^{-1}$ and SRSWOR, where σ_1^2=$1/4f_y^2(Q_y(0.5))$. Next, it can be shown using Theorems 1 and 3 in [74] that under SRSWOR,

$$\sqrt{n}(\overline{y} - \overline{Y}) \xrightarrow{\mathcal{L}} N(0, \sigma_2^2) \text{ and } \sqrt{n}(\hat{\overline{Y}}_{GREG} - \overline{Y}) \xrightarrow{\mathcal{L}} N(0, \sigma_3^2) \tag{4.7.68}$$

as $\nu \to \infty$ *a.s.* [$\mathbf{P}$], where σ_2^2=$(1 - \lambda)\sigma_y^2$, σ_3^2=$(1 - \lambda)\sigma_y^2(1 - \rho_{xy}^2)$, σ_y^2 is the superpopulation variance of y, and ρ_{xy} is the superpopulation correlation coefficient between x and y. Further, it can be shown in the same way as the proof of the result in (4.7.13) that under $\mathbf{P}^*$,

$$\sqrt{n}(\overline{y} - E_{\mathbf{P}}(Y_i)) \xrightarrow{\mathcal{L}} N(0, \sigma_2^2 + \lambda\sigma_y^2) \text{ and } \sqrt{n}(\hat{\overline{Y}}_{GREG} - E_{\mathbf{P}}(Y_i)) \xrightarrow{\mathcal{L}} N(0, \sigma_3^2 + \lambda\sigma_y^2) \quad (4.7.69)$$

as $\nu \to \infty$. Therefore, the conclusion of Theorem 4.5.3 holds in a straightforward way in view of (4.7.67) and (4.7.69). □

4.8. Proofs of additional results required to prove the main results

Let us fix $k \geq 1$ and $p_1, \ldots, p_k \in (0,1)$, and recall $\mathbf{V}_1, \ldots, \mathbf{V}_N$ from the 3^{rd} paragraph in Section 4.2. Define $\hat{\overline{\mathbf{V}}}_1$=$\sum_{i\in s}(N\pi_i)^{-1}\mathbf{V}_i$. Suppose that $P(s,\omega)$ denotes a high entropy sampling design satisfying Assumption 4.2.2, and $Q(s,\omega)$ denotes a rejective sampling design having inclusion probabilities equal to those of $P(s,\omega)$. Recall from the paragraph preceding the proof of Proposition 4.2.1 that such a rejective sampling design always exists. Now, we state the following lemma.

Lemma 4.8.1. *Fix* $\boldsymbol{m} \in \mathbb{R}^{2k}$ *such that* $\boldsymbol{m} \neq 0$*. Suppose that Assumption 4.2.1 holds. Then, under* $Q(s,\omega)$ *as well as* $P(s,\omega)$*, we have*

$$\sqrt{n}\boldsymbol{m}\hat{\overline{\boldsymbol{V}}}_1^T \xrightarrow{\mathcal{L}} N(0, \boldsymbol{m}\Gamma\boldsymbol{m}^T) \text{ as } \nu \to \infty \text{ a.s. } [\boldsymbol{P}],$$

where Γ *is as mentioned in Assumption 4.2.2-(ii).*

Proof. The proof follows exactly the same way as the derivation of the result, which appears in the proof of Lemma 2.7.2 in Section 2.7 of Chapter 2, that $\sqrt{n}\mathbf{m}_1(\hat{\overline{\mathbf{V}}}_1 - \overline{\mathbf{V}})^T \xrightarrow{\mathcal{L}} N(0, \mathbf{m}_1\Gamma_1\mathbf{m}_1^T)$ as $\nu \to \infty$ under each of SRSWOR, LMS and any HEπPS sampling designs for any $\mathbf{m}_1 \in \mathbb{R}^p$, $\mathbf{m}_1 \neq 0$ and Γ_1=$\lim_{\nu\to\infty}\Sigma_1$. □

Next, recall $\{U_i\}_{i=1}^N$ from (4.2.2) in Section 4.2, $F_{u,N}(t)$ and $\mathbb{U}_n(t)$ from (4.7.1) in Section 4.7, and $B_{u,N}(t_1,t_2)$ and $\mathbb{B}_n(t_1,t_2)$ from (4.7.3) in Section 4.7. Now, we state the following lemma.

Lemma 4.8.2. *Suppose that Assumption 4.2.1 holds. Then, there exist constants* $L_1, L_2 > 0$ *such that under* $Q(s,\omega)$*,*

$$E\big[\big(\mathbb{B}_n(t_1,t_2)\big)^2\big(\mathbb{B}_n(t_2,t_3)\big)^2\big] \leq L_1\big(B_{u,N}(t_1,t_3)\big)^2 \text{ a.s. } [\boldsymbol{P}]$$

for any $0 \leq t_1 < t_2 < t_3 \leq 1$ *and* $\nu \geq 1$, *and*

$$\overline{\lim}_{\nu\to\infty} E\big(\mathbb{B}_n(t_1,t_2)\big)^4 \leq L_2\big(t_2-t_1\big)^2 \text{ a.s. } [\boldsymbol{P}]$$

for any $0 \leq t_1 < t_2 \leq 1$.

Proof. Suppose that for i=$1,\ldots,N$, ξ_i=1, when the i^{th} population unit is included in the sample, and ξ_i=0 otherwise. Further, suppose that $S_{k,N}$=$\{(i_1,\ldots,i_k) : i_1,\ldots,i_k \in \{1,2,\ldots,N\}$ and i_1, $\ldots, i_k$ are all distinct$\}$ for k=$2,3,4$. Recall from the proof of the preceding Lemma that under $Q(s,\omega)$, $\sum_{i=1}^N \pi_i(1-\pi_i)/n$ is bounded away from 0 as $\nu \to \infty$ *a.s.* [**P**]. Then, it follows from the proof of Corollary 5.1 in [7] that there exists a constant $K_1 > 0$ such that for all $\nu \geq 1$

$$\begin{aligned}
&\max_{(i_1,i_2)\in S_{2,N}} \big|E\big((\xi_{i_1}-\pi_{i_1})(\xi_{i_2}-\pi_{i_2})\big)\big| < K_1 n/N^2,\\
&\max_{(i_1,i_2,i_3)\in S_{3,N}} \big|E\big((\xi_{i_1}-\pi_{i_1})(\xi_{i_2}-\pi_{i_2})(\xi_{i_3}-\pi_{i_3})\big)\big| < K_1 n^2/N^3, \text{ and} \qquad (4.8.1)\\
&\max_{(i_1,i_2,i_3,i_4)\in S_{4,N}} \big|E\big((\xi_{i_1}-\pi_{i_1})(\xi_{i_2}-\pi_{i_2})(\xi_{i_3}-\pi_{i_3})(\xi_{i_4}-\pi_{i_4})\big)\big| < K_1 n^2/N^4
\end{aligned}$$

under $Q(s,\omega)$ *a.s.* [**P**]. Now, let

$$\begin{aligned}
&B_i = \mathbb{1}_{[t_1<U_i\leq t_2]} - B_{u,N}(t_1,t_2), C_i = \mathbb{1}_{[t_2<U_i\leq t_3]} - B_{u,N}(t_2,t_3),\\
&\alpha_i = B_i(\xi_i/\pi_i - 1) \text{ and } \beta_i = C_i(\xi_i/\pi_i - 1)
\end{aligned}$$

for given any i=$1,\ldots,N$ and $0 \leq t_1 < t_2 < t_3 \leq 1$. Then, we have

$$\begin{aligned}
&E\big[\big(\mathbb{B}_n(t_1,t_2)\big)^2\big(\mathbb{B}_n(t_2,t_3)\big)^2\big] = (n^2/N^4)E\Big[\sum_{i=1}^N \alpha_i^2\beta_i^2 + \sum_{(i_1,i_2)\in S_{2,N}} \alpha_{i_1}\alpha_{i_2}\beta_{i_1}\beta_{i_2}+\\
&\sum_{(i_1,i_2)\in S_{2,N}} \alpha_{i_1}^2\beta_{i_2}^2 + \sum_{(i_1,i_2)\in S_{2,N}} \alpha_{i_1}^2\beta_{i_1}\beta_{i_2} + \sum_{(i_1,i_2)\in S_{2,N}} \alpha_{i_1}\alpha_{i_2}\beta_{i_2}^2 + \sum_{(i_1,i_2,i_3)\in S_{3,N}} \alpha_{i_1}^2\beta_{i_2}\beta_{i_3}+\\
&\sum_{(i_1,i_2,i_3)\in S_{3,N}} \alpha_{i_1}\alpha_{i_2}\beta_{i_3}^2 + \sum_{(i_1,i_2,i_3,i_4)\in S_{4,N}} \alpha_{i_1}\alpha_{i_2}\beta_{i_3}\beta_{i_4}\Big].
\end{aligned}$$

Note that $Q(s,\omega)$ satisfies Assumption 4.2.2–*(ii)* because $P(s,\omega)$ satisfies Assumption 4.2.2–*(ii)*, and $P(s,\omega)$ and $Q(s,\omega)$ have the same inclusion probabilities. Then, we have

$$(n^2/N^4)E\Big[\sum_{i=1}^N \alpha_i^2\beta_i^2\Big] = (n^2/N^4)\sum_{i=1}^N E(\xi_i-\pi_i)^4 B_i^2 C_i^2/\pi_i^4 \leq \qquad (4.8.2)$$

$$(K_2/N)\sum_{i=1}^{N} B_i^2 C_i^2 \le K_3\big(B_{u,N}(t_1,t_3)\big)^2$$

a.s. [**P**] for all $\nu \ge 1$ and some constants $K_2, K_3 > 0$ since Assumption 4.2.1 holds, and $\mathbb{1}_{[t_1<U_i\le t_2]}\mathbb{1}_{[t_2<U_i\le t_3]}$ =0 for any $0 \le t_1 < t_2 < t_3 \le 1$. Next, suppose that $\{\pi_{i_1 i_2} : 1 \le i_1 < i_2 \le N\}$ are second order inclusion probabilities of $Q(s,\omega)$. Then, we note that

$$\begin{aligned}
&(n^2/N^4)E\Big[\sum_{(i_1,i_2)\in S_{2,N}} \alpha_{i_1}\alpha_{i_2}\beta_{i_1}\beta_{i_2}\Big] = (n^2/N^4)\times \\
&\sum_{(i_1,i_2)\in S_{2,N}} E\Big((\xi_{i_1}-\pi_{i_1})^2(\xi_{i_2}-\pi_{i_2})^2\Big)B_{i_1}B_{i_2}C_{i_1}C_{i_2}/\pi_{i_1}^2\pi_{i_2}^2 \le (K_4/n^2)\times \\
&\sum_{(i_1,i_2)\in S_{2,N}} (|\pi_{i_1 i_2}-\pi_{i_1}\pi_{i_2}|+\pi_{i_1}\pi_{i_2})|B_{i_1}C_{i_1}||B_{i_2}C_{i_2}| \le (K_5/N^2)\times \\
&\sum_{(i_1,i_2)\in S_{2,N}} |B_{i_1}C_{i_1}||B_{i_2}C_{i_2}| \le K_6(B_{u,N}(t_1,t_3))^2
\end{aligned} \tag{4.8.3}$$

a.s. [**P**] for all $\nu \ge 1$ and some constants $K_4, K_5, K_6 > 0$ since Assumption 4.2.2-*(ii)* holds, $E((\xi_{i_1}-\pi_{i_1})^2(\xi_{i_2}-\pi_{i_2})^2)$=$(\pi_{i_1 i_2}-\pi_{i_1}\pi_{i_2})(1-2\pi_{i_1})(1-2\pi_{i_2})+\pi_{i_1}\pi_{i_2}(1-\pi_{i_1})(1-\pi_{i_2})$ for $(i_1,i_2)\in S_{2,N}$, and $\max_{(i_1,i_2)\in S_{2,N}}\big|E\big((\xi_{i_1}-\pi_{i_1})(\xi_{i_2}-\pi_{i_2})\big)\big|$=$\max_{(i_1,i_2)\in S_{2,N}}|\pi_{i_1 i_2}-\pi_{i_1}\pi_{i_2}| < K_1 n/N^2$ *a.s.* [**P**] by (4.8.1). An inequality similar to (4.8.3) holds for $(n^2/N^4)E$ $[\sum_{(i_1,i_2)\in S_{2,N}}\alpha_{i_1}^2\beta_{i_2}^2]$. Since, $\big|E\big((\xi_{i_1}-\pi_{i_1})^3(\xi_{i_2}-\pi_{i_2})\big)\big| \le 7|\pi_{i_1 i_2}-\pi_{i_1}\pi_{i_2}|$, inequalities similar to (4.8.3) also hold for $(n^2/N^4)E\,[\sum_{(i_1,i_2)\in S_{2,N}}\alpha_{i_1}^2\beta_{i_1}\beta_{i_2}]$ and $(n^2/N^4)E[\sum_{(i_1,i_2)\in S_{2,N}}\alpha_{i_1}\alpha_{i_2}\beta_{i_2}^2]$. Note that

$$\begin{aligned}
&E\big((\xi_{i_1}-\pi_{i_1})^2(\xi_{i_2}-\pi_{i_2})(\xi_{i_3}-\pi_{i_3})\big) = (1-2\pi_{i_1})E\big((\xi_{i_1}-\pi_{i_1})(\xi_{i_2}-\pi_{i_2})\times \\
&(\xi_{i_3}-\pi_{i_3})\big) + \pi_{i_1}(1-\pi_{i_1})E\big((\xi_{i_2}-\pi_{i_2})(\xi_{i_3}-\pi_{i_3})\big) \text{ for } (i_1,i_2,i_3)\in S_{3,N}.
\end{aligned}$$

Also, note that

$$\begin{aligned}
&\max_{(i_1,i_2,i_3)\in S_{3,N}}\big|E\big((\xi_{i_1}-\pi_{i_1})(\xi_{i_2}-\pi_{i_2})(\xi_{i_3}-\pi_{i_3})\big)\big| < K_1 n^2/N^3 \text{ and} \\
&\max_{(i_1,i_2,i_3,i_4)\in S_{4,N}}\big|E\big((\xi_{i_1}-\pi_{i_1})(\xi_{i_2}-\pi_{i_2})(\xi_{i_3}-\pi_{i_3})(\xi_{i_4}-\pi_{i_4})\big)\big| < K_1 n^2/N^4 \textit{ a.s. } [\mathbf{P}]
\end{aligned}$$

by (4.8.1). Therefore, it can be shown in the same way as in (4.8.2) and (4.8.3) that under $Q(s,\omega)$,

$$(n^2/N^4)E\Big[\sum_{(i_1,i_2,i_3)\in S_{3,n}} \alpha_{i_1}^2\beta_{i_2}\beta_{i_3}\Big] \le K_7\big(B_{u,N}(t_1,t_3)\big)^2,$$

$$(n^2/N^4)E\Big[\sum_{(i_1,i_2,i_3)\in S_{3,N}} \alpha_{i_1}\alpha_{i_2}\beta_{i_3}^2\Big] \leq K_7\big(B_{u,N}(t_1,t_3)\big)^2 \text{ and}$$
$$(n^2/N^4)E\Big[\sum_{(i_1,i_2,i_3,i_4)\in S_{4,N}} \alpha_{i_1}\alpha_{i_2}\beta_{i_3}\beta_{i_4}\Big] \leq K_7\big(B_{u,N}(t_1,t_3)\big)^2 \textit{ a.s. } [\mathbf{P}]$$

for all $\nu \geq 1$ and some constant $K_7 > 0$. Hence, there exists a constant $K_8 > 0$ such that under $Q(s,\omega)$, $E\big[\big(\mathbb{B}_n(t_1,t_2)\big)^2\,\big(\mathbb{B}_n(t_2,t_3)\big)^2\big] \leq K_8\big(B_{u,N}(t_1,t_3)\big)^2$ *a.s.* [**P**] for any $\nu \geq 1$ and $0 \leq t_1 < t_2 < t_3 \leq 1$.

Next, one can shown that

$$E\big(\mathbb{B}_n(t_1,t_2)\big)^4 = (n^2/N^4)E\big[\sum_{i=1}^{N}\alpha_i^4 + 2\sum_{(i_1,i_2)\in S_{2,N}}\alpha_{i_1}^2\alpha_{i_2}^2 +$$
$$2\sum_{(i_1,i_2)\in S_{2,N}}\alpha_{i_1}^3\alpha_{i_2} + 2\sum_{(i_1,i_2,i_3)\in S_{3,N}}\alpha_{i_1}^2\alpha_{i_2}\alpha_{i_3} + \sum_{(i_1,i_2,i_3,i_4)\in S_{4,N}}\alpha_{i_1}\alpha_{i_2}\alpha_{i_3}\alpha_{i_4}\big].$$

It can also be shown in the same way as in (4.8.2) and (4.8.3) that under $Q(s,\omega)$,

$$(n^2/N^4)E\big[\sum_{i=1}^{N}\alpha_i^4\big] = O(1/n) \text{ as } \nu\to\infty \textit{ a.s. } [\mathbf{P}], \text{ and}$$
$$(n^2/N^4)E\big[2\sum_{(i_1,i_2)\in S_{2,N}}\alpha_{i_1}^2\alpha_{i_2}^2 + 2\sum_{(i_1,i_2)\in S_{2,N}}\alpha_{i_1}^3\alpha_{i_2} + 2\sum_{(i_1,i_2,i_3)\in S_{3,N}}\alpha_{i_1}^2\alpha_{i_2}\alpha_{i_3} +$$
$$\sum_{(i_1,i_2,i_3,i_4)\in S_{4,N}}\alpha_{i_1}\alpha_{i_2}\alpha_{i_3}\alpha_{i_4}\big] \leq K_9\big(B_{u,N}(t_1,t_2)\big)^2 \text{ given any } \nu \geq 1 \textit{ a.s. } [\mathbf{P}]$$

for some constant $K_9 > 0$. Therefore, under $Q(s,\omega)$, $\overline{\lim}_{\nu\to\infty} E\big(\mathbb{B}_n(t_1,t_2)\big)^4 \leq K_9\big(t_2-t_1)\big)^2$ *a.s.* [**P**] because $B_{u,N}(t_1,t_2) \to (t_2-t_1)$ *a.s.* [**P**] by SLLN. Hence, the result follows. □

Next, fix $k \geq 1$ and $p_1,\ldots,p_k \in (0,1)$ and define $\hat{\overline{\mathbf{V}}}_2$=$\sum_{i\in s}(NX_i)^{-1}G_i\mathbf{V}_i$, where $\mathbf{V}_i$'s are as in the 3^{rd} paragraph of Section 4.2 and G_i's are as in the 1^{st} paragraph of Section 4.1. Also, recall γ from the paragraph preceding the statement of Theorem 4.2.2 in Section 4.2.

Lemma 4.8.3. *Fix* $\boldsymbol{m} \in \mathbb{R}^{2k}$ *such that* $\boldsymbol{m} \neq 0$. *Suppose that* $E_{\boldsymbol{P}}(X_i)^{-1} < \infty$, *and Assumptions 4.2.1 and 4.2.4–4.2.6 hold. Then, under RHC sampling design, we have*

$$\sqrt{n}\boldsymbol{m}\hat{\overline{\boldsymbol{V}}}_2^T \xrightarrow{\mathcal{L}} N(0,\boldsymbol{m}\Gamma_6\boldsymbol{m}^T) \textit{ as } \nu\to\infty \textit{ a.s. } [\boldsymbol{P}],$$

where Γ_6=$cE_{\boldsymbol{P}}(X_i)E_{\boldsymbol{P}}[(\boldsymbol{R}_i - E_{\boldsymbol{P}}(\boldsymbol{R}_i))^T(\boldsymbol{R}_i - E_{\boldsymbol{P}}(\boldsymbol{R}_i))/X_i]$, *and* c=$\lim_{\nu\to\infty} n\gamma$.

Note that Γ_6 is p.d. by Assumption 4.2.5. Also, note that $\lim_{\nu\to\infty} n\gamma$ exists by Lemma 2.7.5 in Section 2.7 of Chapter 2.

Proof. The proof follows exactly the same way as the derivation of the result, which appears in the proof of Lemma 2.7.2 in Section 2.7 of Chapter 2, that $\sqrt{n}\mathbf{m}_1(\hat{\overline{\mathbf{V}}}_2 - \overline{\mathbf{V}})^T \xrightarrow{\mathcal{L}} N(0, \mathbf{m}_1\Gamma_2\mathbf{m}_1^T)$ as $\nu \to \infty$ under RHC sampling design for any $\mathbf{m}_1 \in \mathbb{R}^p$, $\mathbf{m}_1 \neq 0$ and Γ_2=$\lim_{\nu\to\infty} \Sigma_2$. □

Before we state the next result, recall $\{U_i\}_{i=1}^N$ from (4.2.2) in Section 4.2, and $F_{u,N}(t)$ from (4.7.1) and $B_{u,N}(t_1, t_2)$ from (4.7.3) in Section 4.7. Define $\tilde{\mathbb{U}}_n(t)$=$\sqrt{n}\sum_{i\in s}(NX_i)^{-1}G_i(\mathbb{1}_{[U_i\le t]} - F_{u,N}(t))$ for $0 \le t \le 1$ and $\tilde{\mathbb{B}}_n(t_1, t_2)$=$\tilde{\mathbb{U}}_n(t_2) - \tilde{\mathbb{U}}_n(t_1)$ for $0 \le t_1 < t_2 \le 1$.

Lemma 4.8.4. *Suppose that Assumptions 4.2.4 and 4.2.6 hold. Then, there exist constants* $L_1, L_2 > 0$ *such that under RHC sampling design,*

$$E\big[\big(\tilde{\mathbb{B}}_n(t_1, t_2)\big)^2\big(\tilde{\mathbb{B}}_n(t_2, t_3)\big)^2\big] \le L_1\big(B_{u,N}(t_1, t_3)\big)^2 \text{ a.s. } [\boldsymbol{P}]$$

for any $0 \le t_1 < t_2 < t_3 \le 1$ *and* $\nu \ge 1$*, and*

$$\overline{\lim}_{\nu\to\infty} E\big(\tilde{\mathbb{B}}_n(t_1, t_2)\big)^4 \le L_2\big(t_2 - t_1\big)^2 \text{ a.s. } [\boldsymbol{P}]$$

for any $0 \le t_1 < t_2 \le 1$.

Proof. Recall from Section 4.2 that RHC sampling design is implemented in two steps. In the first step, the entire population is randomly divided into n groups, say $\mathcal{P}_1, \ldots, \mathcal{P}_n$ of sizes $\tilde{N}_1 \cdots, \tilde{N}_n$ respectively. Then, in the second step, a unit is selected from each group independently. For each r=$1, \ldots, n$, the q^{th} unit from $\mathcal{P}_r$ is selected with probability X'_{qr}/Q_r, where X'_{qr} is the x value of the q^{th} unit in $\mathcal{P}_r$ and Q_r=$\sum_{q=1}^{\tilde{N}_r} X'_{qr}$. Let E_1 and E_2 denote design expectations with respect to the 1^{st} and the 2^{nd} steps, respectively. Suppose that (y_r, x_r) is the value of (y, x) corresponding to the r^{th} unit in the sample for r=$1, \ldots, n$. Further, suppose that z_r=$F_y(y_r)$ for r=$1, \ldots, n$, where F_y is the superpopulation distribution function of y. Define

$$\alpha_r = Q_r(\mathbb{1}_{[t_1<z_r\le t_2]} - B_{u,N}(t_1, t_2))/x_r \text{ and } \beta_r = Q_r(\mathbb{1}_{[t_2<z_r\le t_3]} - B_{u,N}(t_2, t_3))/x_r$$

for $0 \le t_1 < t_2 < t_3 \le 1$ and r=$1, \ldots, n$. Note that $\tilde{\mathbb{U}}_n(t)$=$\sqrt{n}\sum_{i\in s}(NX_i)^{-1}G_i(\mathbb{1}_{[U_i\le t]} - F_{u,N}(t))$=$\sqrt{n}\sum_{r=1}^n Q_r(\mathbb{1}_{[z_r\le t]} - F_{u,N}(t))/Nx_r$. Then, we have

$$E\big[\big(\tilde{\mathbb{B}}_n(t_1,t_2)\big)^2\big(\tilde{\mathbb{B}}_n(t_2,t_3)\big)^2\big] = (n^2/N^4)E_1E_2\Bigg[\sum_{r=1}^{n}\alpha_r^2\beta_r^2 + \sum_{(r_1,r_2)\in S_{2,n}}\alpha_{r_1}\alpha_{r_2}\beta_{r_1}\beta_{r_2}+$$
$$\sum_{(r_1,r_2)\in S_{2,n}}\alpha_{r_1}^2\beta_{r_2}^2 + \sum_{(r_1,r_2)\in S_{2,n}}\alpha_{r_1}^2\beta_{r_1}\beta_{r_2} + \sum_{(r_1,r_2)\in S_{2,n}}\alpha_{r_1}\alpha_{r_2}\beta_{r_2}^2 + \sum_{(r_1,r_2,r_3)\in S_{3,n}}\alpha_{r_1}^2\beta_{r_2}\beta_{r_3}$$
$$+ \sum_{(r_1,r_2,r_3)\in S_{3,n}}\alpha_{r_1}\alpha_{r_2}\beta_{r_3}^2 + \sum_{(r_1,r_2,r_3,r_4)\in S_{4,n}}\alpha_{r_1}\alpha_{r_2}\beta_{r_3}\beta_{r_4}\Bigg],$$

where $S_{k,n}$=$\{(r_1,\ldots,r_k)\ :\ r_1,\ldots,r_k\ \in\ \{1,2,\ldots,n\}$ and $r_1,\ldots,r_k$ are all distinct$\}$ for k= $2,3,4$. Suppose that for i=$1,\ldots,N$,

$$\xi_{ir} = \begin{cases} 1, \text{ when the } i^{th} \text{ population unit is selected in the } r^{th} \text{ group } \mathcal{P}_r, \text{ and} \\ 0, \text{ otherwise.} \end{cases}$$

Note that by Assumption 4.2.4, $\max_{1\leq i\leq N} X_i/\min_{1\leq i\leq N} X_i \leq K_1$ *a.s.* [**P**] for all $\nu \geq 1$ and some constant $K_1 > 0$. Also, note that $n\max_{1\leq r\leq n}\tilde{N}_r/N \leq 2$ for all $\nu \geq 1$ because $\{\tilde{N}_r\}_{r=1}^{n}$ are as in page 484 of [66]. Recall B_i and C_i from the proof of Lemma 4.8.2. Then, we have

$$(n^2/N^4)E_1\Bigg[\sum_{r=1}^{n}E_2(\alpha_r^2\beta_r^2)\Bigg] = (n^2/N^4)E_1\Bigg[\sum_{r=1}^{n}\Bigg(\sum_{i=1}^{N}B_i^2C_i^2\xi_{ir}/X_i^3\Bigg)Q_r^3\Bigg] \leq$$
$$(K_1)^3(n^2/N^4)E_1\Bigg[\sum_{r=1}^{n}\Bigg(\sum_{i=1}^{N}B_i^2C_i^2\xi_{ir}\Bigg)\tilde{N}_r^3\Bigg] \leq (K_2/N)\Bigg[\sum_{i=1}^{N}B_i^2C_i^2E_1\Bigg(\sum_{r=1}^{n}\xi_{ir}\Bigg)\Bigg] \quad (4.8.4)$$
$$= (K_2/N)\Bigg[\sum_{i=1}^{N}B_i^2C_i^2\Bigg] \leq K_3\big(B_{u,N}(t_1,t_3)\big)^2$$

a.s. [**P**] for all $\nu \geq 1$ and some constants $K_2, K_3 > 0$ since $\sum_{r=1}^{n}\xi_{ir}$=1 for any $1 \leq i \leq N$. Next, recall $S_{2,N}$ from the proof of Lemma 4.8.2 and note that

$$(n^2/N^4)E_1\Bigg[\sum_{(r_1,r_2)\in S_{2,n}}E_2(\alpha_{r_1}\alpha_{r_2}\beta_{r_1}\beta_{r_2})\Bigg] = (n^2/N^4)\times$$
$$E_1\Bigg[\sum_{(r_1,r_2)\in S_{2,n}}E_2(\alpha_{r_1}\beta_{r_1})E_2(\alpha_{r_2}\beta_{r_2})\Bigg] = (n^2/N^4)E_1\Bigg[\sum_{(r_1,r_2)\in S_{2,n}}$$
$$\Bigg(\sum_{(i_1,i_2)\in S_{2,N}}B_{i_1}C_{i_1}B_{i_2}C_{i_2}\xi_{i_1r_1}\xi_{i_2r_2}/X_{i_1}X_{i_2}\Bigg)Q_{r_1}Q_{r_2}\Bigg] \leq (K_1)^2(n^2/N^4)\times \quad (4.8.5)$$
$$E_1\Bigg[\sum_{(r_1,r_2)\in S_{2,n}}\Bigg(\sum_{(i_1,i_2)\in S_{2,N}}|B_{i_1}C_{i_1}||B_{i_2}C_{i_2}|\xi_{i_1r_1}\xi_{i_2r_2}\Bigg)N_{r_1}N_{r_2}\Bigg] \leq$$

$$K_4 \sum_{i_1=1}^{N} |B_{i_1} C_{i_1}| \sum_{i_2=1}^{N} |B_{i_2} C_{i_2}|/N(N-1) \leq K_5 \big(B_{u,N}(t_1, t_3)\big)^2$$

a.s. [**P**] for all $\nu \geq 1$ and some constants $K_4, K_5 > 0$ since units are selected from $\mathcal{P}_{r_1}$ and $\mathcal{P}_{r_2}$ independently, $\{\tilde{N}_r\}_{r=1}^{n}$ are as in page 484 of [66], and $E_1(\xi_{i_1 r_1}\xi_{i_2 r_2})$=$N_{r_1}N_{r_2}/N(N-1)$ for any $(r_1, r_2) \in S_{2,n}$ and $(i_1, i_2) \in S_{2,N}$. It can be shown that an inequality similar to (4.8.5) holds for each of $(n^2/N^4)E_1E_2[\sum_{(r_1,r_2)\in S_{2,n}} \alpha_{r_1}^2 \beta_{r_2}^2]$, $(n^2/N^4)E_1E_2[\sum_{(r_1,r_2)\in S_{2,n}} \alpha_{r_1}^2 \beta_{r_1}\beta_{r_2}]$ and $(n^2/N^4)E_1E_2[\sum_{(r_1,r_2)\in S_{2,n}} \alpha_{r_1}\alpha_{r_2}\beta_{r_2}^2]$. Note that

$$E_1(\xi_{i_1 r_1}\xi_{i_2 r_2}\xi_{i_3 r_3}) = N_{r_1}N_{r_2}N_{r_3}/(N(N-1)(N-2))$$

for $(r_1, r_2, r_3) \in S_{3,n}$ and $(i_1, i_2, i_3) \in S_{3,N}$, and $\sum_{(r_1,r_2,r_3)\in S_{3,n}} N_{r_1}N_{r_2}N_{r_3}/N(N-1)(N-2)$ is bounded. Also, note that

$$E_1(\xi_{i_1 r_1}\xi_{i_2 r_2}\xi_{i_3 r_3}\xi_{i_4 r_4}) = (N_{r_1}N_{r_2}N_{r_3}N_{r_4})/N(N-1)(N-2)(N-3)$$

for $(r_1, r_2, r_3, r_4) \in S_{4,n}$ and $(i_1, i_2, i_3, i_4) \in S_{4,N}$, and $\sum_{(r_1,r_2,r_3,r_4)\in S_{4,n}} N_{r_1}N_{r_2}N_{r_3}N_{r_4}/N(N-1)(N-2)(N-3)$ is bounded. Then, it can be shown in the same way as in (4.8.4) and (4.8.5) above that

$$(n^2/N^4)E_1E_2\Big[\sum_{(r_1,r_2,r_3)\in S_{3,n}} \alpha_{r_1}^2 \beta_{r_2}\beta_{r_3}\Big] \leq K_6\big(B_{u,N}(t_1, t_3)\big)^2,$$

$$(n^2/N^4)E_1E_2\Big[\sum_{(r_1,r_2,r_3)\in S_{3,n}} \alpha_{r_1}\alpha_{r_2}\beta_{r_3}^2\Big] \leq K_6\big(B_{u,N}(t_1, t_3)\big)^2 \text{ and}$$

$$(n^2/N^4)E_1E_2\Big[\sum_{(r_1,r_2,r_3,r_4)\in S_{4,n}} \alpha_{r_1}\alpha_{r_2}\beta_{r_3}\beta_{r_4}\Big] \leq K_6\big(B_{u,N}(t_1, t_3)\big)^2 \; a.s. \; [\mathbf{P}]$$

for all $\nu \geq 1$ and some constant $K_6 > 0$. Thus

$$E\big[\big(\tilde{\mathbb{B}}_n(t_1, t_2)\big)^2\big(\tilde{\mathbb{B}}_n(t_2, t_3)\big)^2\big] \leq K_7\big(B_{u,N}(t_1, t_3)\big)^2 a.s. [\mathbf{P}]$$

for all $\nu \geq 1$ and some constant $K_7 > 0$.

Next, note that

$$E\big(\tilde{\mathbb{B}}_n(t_1, t_2)\big)^4 = (n^2/N^4)E_1E_2\Big[\sum_{r=1}^{n} \alpha_r^4 + 2\sum_{(r_1,r_2)\in S_{2,n}} \alpha_{r_1}^2\alpha_{r_2}^2 +$$

$$2\sum_{(r_1,r_2)\in S_{2,n}}\alpha_{r_1}^3\alpha_{r_2}+2\sum_{(r_1,r_2,r_3)\in S_{3,n}}\alpha_{r_1}^2\alpha_{r_2}\alpha_{r_3}+\sum_{(r_1,r_2,r_3,r_4)\in S_{4,n}}\alpha_{r_1}\alpha_{r_2}\alpha_{r_3}\alpha_{r_4}\Bigg].$$

It can be shown in the same way as in (4.8.4) and (4.8.5) above that

$$(n^2/N^4)E_1E_2\left[\sum_{r=1}^{n}\alpha_r^4\right]=O(1/n)\text{ as }\nu\to\infty\textit{ a.s. }[\mathbf{P}],\text{ and }(n^2/N^4)\times$$
$$E_1E_2\Bigg[2\sum_{(r_1,r_2)\in S_{2,n}}\alpha_{r_1}^2\alpha_{r_2}^2+2\sum_{(r_1,r_2)\in S_{2,n}}\alpha_{r_1}^3\alpha_{r_2}+2\sum_{(r_1,r_2,r_3)\in S_{3,n}}\alpha_{r_1}^2\alpha_{r_2}\alpha_{r_3}+$$
$$\sum_{(r_1,r_2,r_3,r_4)\in S_{4,n}}\alpha_{r_1}\alpha_{r_2}\alpha_{r_3}\alpha_{r_4}\Bigg]\le K_8\big(B_{u,N}(t_1,t_2)\big)^2\text{ given any }\nu\ge1\textit{ a.s. }[\mathbf{P}]$$

for some constant $K_8>0$. Therefore, $\overline{\lim}_{\nu\to\infty}E\big(\tilde{\mathbb{B}}_n(t_1,t_2)\big)^4\le K_8\big(t_2-t_1\big)^2$ *a.s.* [**P**] since $B_{u,N}(t_1,t_2)\to(t_2-t_1)$ as $\nu\to\infty$ *a.s.* [**P**] by SLLN. □

Next, we state the following lemma, which is required to prove Theorem 4.4.3.

Lemma 4.8.5. *(i) Fix* $0<\alpha<\beta<1$*. Suppose that the assumptions of Theorem 4.2.1 hold,* $K(p_1,p_2)$ *is as in* (4.2.3) *in Section 4.2, and* $\hat{K}(p_1,p_2)$ *is as in* (4.4.4) *in Section 4.4.1. Then, under* ***P*****,*

$$\sup_{p_1,p_2\in[\alpha,\beta]}|\hat{K}(p_1,p_2)-K(p_1,p_2)|=O_p(1)\textit{ and }\hat{K}(p_1,p_2)\xrightarrow{p}K(p_1,p_2)\textit{ as }\nu\to\infty\qquad(4.8.6)$$

for any $p_1,p_2\in[\alpha,\beta]$ *and high entropy sampling design satisfying Assumption 4.2.2.*
(ii) Further, if the assumptions of Theorem 4.2.2 hold, $K(p_1,p_2)$ *is as in* (4.2.5) *in Section 4.2, and* $\hat{K}(p_1,p_2)$ *is as in* (4.4.5) *in Section 4.4.1. Then, the above results hold under RHC sampling design.*

Proof. *(i)* Let us first consider a high entropy sampling design $P(s,\omega)$ satisfying Assumption 4.2.2, and a rejective sampling design $Q(s,\omega)$ having inclusion probabilities equal to those of $P(s,\omega)$. Since, $K(p_1,p_2)$ in (4.2.3) in Section 4.2 and $\hat{K}(p_1,p_2)$ in (4.4.4) in Section 4.4.1 depend on $P(s,\omega)$ only through its inclusion probabilities, and $P(s,\omega)$ and $Q(s,\omega)$ have equal inclusion probabilities, it is enough to show that the results in (4.8.6) hold for $Q(s,\omega)$. The results in (4.8.6) holds for $P(s,\omega)$ in the same way as the conclusion of Proposition 4.2.1 holds for $P(s,\omega)$ in Section 4.7. We shall first show that under $\mathbf{P}^*$,

$$\sup_{p_1,p_2\in[\alpha,\beta]}|\hat{K}(p_1,p_2)-K(p_1,p_2)|=O_p(1)\text{ as }\nu\to\infty\text{ for }Q(s,\omega).$$

It can be shown in the same way as the derivation of the result in (4.7.25) in Section 4.7 that under $\mathbf{P}^*$, $\{\sqrt{n}(\hat{Q}_y(p)-Q_y(p)) : p \in [\alpha/2,(1+\beta)/2]\}$ converges weakly to a mean 0 Gaussian process as $\nu \to \infty$ in $(D[\alpha/2,(1+\beta)/2],\mathcal{D})$ with respect to the sup norm metric, for $Q(s,\omega)$. Consequently,

$$\sup_{p\in[\alpha/2,(1+\beta)/2]} |\sqrt{n}(\hat{Q}_y(p)-Q_y(p))| = O_p(1) \tag{4.8.7}$$

as $\nu \to \infty$ under $\mathbf{P}^*$ by continuous mapping theorem. Then, under $\mathbf{P}^*$, we have

$$\sup_{p\in[\alpha,\beta]} |\sqrt{n}(\hat{Q}_y(p+1/\sqrt{n})-\hat{Q}_y(p-1/\sqrt{n}))/2| = O_p(1) \text{ as } \nu\to\infty \text{ for } Q(s,\omega)$$

since $\alpha - 1/\sqrt{n} \geq \alpha/2$ and $\beta + 1/\sqrt{n} \leq (1+\beta)/2$ for all sufficiently large ν, and $f_y \circ Q_y$ is bounded away from 0 on $[\alpha/2,(1+\beta)/2]$ by Assumption 4.2.3. Here, we recall from Table 4.5 in Section 4.4.1 that $\sqrt{n}(\hat{Q}_y(p+1/\sqrt{n})-\hat{Q}_y(p-1/\sqrt{n}))/2$ is the estimator of $1/f_y(Q_y(p))$. Similarly, under $\mathbf{P}^*$,

$$\sup_{p\in[\alpha,\beta]} |\sqrt{n}(\hat{Q}_x(p+1/\sqrt{n})-\hat{Q}_x(p-1/\sqrt{n}))/2| = O_p(1) \text{ as } \nu\to\infty \text{ for } Q(s,\omega).$$

It further follows from (4.7.28) and (4.7.34) in the proof of Theorem 4.2.1 in Section 4.7 that under $\mathbf{P}^*$,

$$\sup_{p\in[\alpha,\beta]} |\hat{Q}_y(p)/\hat{Q}_x(p)-Q_y(p)/Q_x(p)| \xrightarrow{p} 0, \sum_{i\in s}\pi_i^{-1}Y_i \Big/ \sum_{i\in s}\pi_i^{-1}X_i \xrightarrow{p} E_{\mathbf{P}}(Y_i)\Big/E_{\mathbf{P}}(X_i)$$
$$\text{and } \sum_{i\in s}\pi_i^{-1}X_iY_i \Big/ \sum_{i\in s}\pi_i^{-1}X_i^2 \xrightarrow{p} E_{\mathbf{P}}(X_iY_i)\Big/E_{\mathbf{P}}(X_i^2) \text{ as } \nu\to\infty \text{ for } Q(s,\omega).$$

Similarly, it can be shown that under $\mathbf{P}^*$,

$$\sum_{i\in s}(1-\pi_i)\Big/\sum_{i=1}^{N}\pi_i(1-\pi_i) \xrightarrow{p} 1 \text{ as } \nu\to\infty \text{ for } Q(s,\omega).$$

Consequently, under $\mathbf{P}^*$,

$$\sup_{p_1,p_2\in[\alpha,\beta]} |\hat{K}(p_1,p_2)-K(p_1,p_2)| = O_p(1) \text{ as } \nu\to\infty \text{ for } Q(s,\omega).$$

This completes the proof of the first result in (4.8.6) for $Q(s,\omega)$.

Next, if we establish that under $\mathbf{P}^*$,

$$\hat{K}(p_1,p_2)-\tilde{K}(p_1,p_2)\xrightarrow{p}0 \text{ and } \tilde{K}(p_1,p_2)\xrightarrow{p}K(p_1,p_2)$$

as $\nu\to\infty$ for $Q(s,\omega)$ and any $p_1,p_2\in[\alpha,\beta]$, then the result

$$\hat{K}(p_1,p_2)\xrightarrow{p}K(p_1,p_2) \text{ as } \nu\to\infty \text{ for } Q(s,\omega) \text{ and any } p_1,p_2\in[\alpha,\beta] \text{ under } \mathbf{P}^*$$

will follow. Here,

$$\tilde{K}(p_1,p_2)=(n/N^2)\sum_{i\in s}(\zeta_i(p_1)-\overline{\zeta}(p_1)-S(p_1)\pi_i)(\zeta_i(p_2)-\overline{\zeta}(p_2)-S(p_2)\pi_i)(\pi_i^{-1}-1)\pi_i^{-1}.$$

Note that

$$\tilde{K}(p_1,p_2)-(n/N^2)\sum_{i=1}^{N}(\zeta_i(p_1)-\overline{\zeta}(p_1)-S(p_1)\pi_i)(\zeta_i(p_2)-\overline{\zeta}(p_2)-S(p_2)\pi_i)(\pi_i^{-1}-1)$$
$$\xrightarrow{p}0 \text{ as } \nu\to\infty \text{ for any } p_1,p_2\in[\alpha,\beta] \text{ under } \mathbf{P}^*$$

in the same way as the derivation of the result $\sum_{i\in s}(N\pi_i)^{-1}\xrightarrow{p}1$ for $Q(s,\omega)$ under $\mathbf{P}^*$ in the proof of Proposition 4.2.1 (see the last few lines in 2^{nd} paragraph of the proof of Proposition 4.2.1 in Section 4.7). Also, note that $(n/N^2)\sum_{i=1}^{N}(\zeta_i(p_1)-\overline{\zeta}(p_1)-S(p_1)\pi_i)(\zeta_i(p_2)-\overline{\zeta}(p_2)-S(p_2)\pi_i)(\pi_i^{-1}-1)$ has a deterministic limit *a.s.* [$\mathbf{P}$] for any $p_1,p_2\in[\alpha,\beta]$ in view of Assumption 4.2.2-*(i)*. Further,

$$E_{\mathbf{P}}\Bigg(\lim_{\nu\to\infty}(n/N^2)\sum_{i=1}^{N}(\zeta_i(p_1)-\overline{\zeta}(p_1)-S(p_1)\pi_i)(\zeta_i(p_2)-\overline{\zeta}(p_2)-S(p_2)\pi_i)(\pi_i^{-1}-1)\Bigg)$$
$$=K(p_1,p_2) \text{ for any } p_1,p_2\in[\alpha,\beta]$$

in view of Assumption 4.2.2-*(ii)* and DCT. Therefore, as $\nu\to\infty$,

$$(n/N^2)\sum_{i=1}^{N}(\zeta_i(p_1)-\overline{\zeta}(p_1)-S(p_1)\pi_i)(\zeta_i(p_2)-\overline{\zeta}(p_2)-S(p_2)\pi_i)(\pi_i^{-1}-1)\to K(p_1,p_2)$$

a.s. [$\mathbf{P}$], and hence $\tilde{K}(p_1,p_2)\xrightarrow{p}K(p_1,p_2)$ under $\mathbf{P}^*$ for any $p_1,p_2\in[\alpha,\beta]$.

Next, let us fix $\nu\geq1$, $t>0$, $\delta>0$ and $p\in[\alpha,\beta]$. Then, we have

$$\Bigg\{\sqrt{n}|\hat{Q}_y(p)-Q_y(p)|\leq t \text{ and } \sum_{i\in s}(\mathbb{1}_{[Y_i\leq Q_y(p)+t/\sqrt{n}]}-\mathbb{1}_{[Y_i\leq Q_y(p)-t/\sqrt{n}]})/N\pi_i \qquad (4.8.8)$$

$$\leq \delta\Big\} \subseteq \Big\{|\sum_{i\in s} \mathbb{1}_{[Y_i \leq \hat{Q}_y(p)]}/N\pi_i - \sum_{i\in s} \mathbb{1}_{[Y_i \leq Q_y(p)]}/N\pi_i| \leq \delta\Big\}.$$

Further, one can show that under **P***,

$$\sum_{i\in s}(\mathbb{1}_{[Y_i\leq Q_y(p)+t/\sqrt{n}]} - \mathbb{1}_{[Y_i\leq Q_y(p)-t/\sqrt{n}]})\Big/ N\pi_i -$$
$$F_{y,N}(Q_y(p)+t/\sqrt{n}) + F_{y,N}(Q_y(p)-t/\sqrt{n}) \xrightarrow{p} 0 \text{ as } \nu\to\infty$$

in the same way as the derivation of the result $\sum_{i\in s}(N\pi_i)^{-1} \xrightarrow{p} 1$ for $Q(s,\omega)$ under **P*** in the proof of Proposition 4.2.1. Moreover, under **P**, $F_{y,N}(Q_y(p)+t/\sqrt{n}) - F_{y,N}(Q_y(p)-t/\sqrt{n}) \xrightarrow{p} 0$ as $\nu\to\infty$ by Chebyshev's inequality and Assumption 4.2.3. Thus as $\nu\to\infty$

$$\sum_{i\in s}(\mathbb{1}_{[Y_i\leq Q_y(p)+t/\sqrt{n}]} - \mathbb{1}_{[Y_i\leq Q_y(p)-t/\sqrt{n}]})/N\pi_i \xrightarrow{p} 0 \text{ under } \mathbf{P}^*. \tag{4.8.9}$$

Moreover, it follows from (4.8.7) above that as $\nu\to\infty$,

$$\sqrt{n}|\hat{Q}_y(p) - Q_y(p)| = O_p(1) \text{ under } \mathbf{P}^*. \tag{4.8.10}$$

Therefore, using (4.8.8), (4.8.9) and (4.8.10) above, one can show that

$$\sum_{i\in s}\mathbb{1}_{[Y_i\leq \hat{Q}_y(p)]}\Big/ N\pi_i - \sum_{i\in s}\mathbb{1}_{[Y_i\leq Q_y(p)]}\Big/ N\pi_i \xrightarrow{p} 0 \text{ as } \nu\to\infty \text{ under } \mathbf{P}^*.$$

Now, suppose that p_n=$p + c/\sqrt{n}$ for $c\in\mathbb{R}$. Then, we have

$$Q_y(p_n) = Q_y(p) + (c/\sqrt{n})(1/f_y(Q_y(\epsilon_n)))$$

by Taylor expansion, where $\epsilon_n\to p$ as $\nu\to\infty$. Thus one can show that as $\nu\to\infty$,

$$\sqrt{n}(\hat{F}_y(Q_y(p_n)) - \hat{F}_y(Q_y(p)) - F_y(Q_y(p_n)) + p) \xrightarrow{p} 0 \text{ under } \mathbf{P}^*$$

in the same way as the derivation of the result $\sum_{i\in s}(N\pi_i)^{-1} \xrightarrow{p} 1$ for $Q(s,\omega)$ under **P*** in the proof of Proposition 4.2.1. Further, it can be shown that

$$\hat{Q}_y(p) - Q_y(p) = (p - \hat{F}_y(Q_y(p)))/f_y(Q_y(p)) + o_p(1/\sqrt{n}) \text{ as } \nu\to\infty \text{ under } \mathbf{P}^*.$$

Similarly, we have

$$\hat{Q}_y(p_n) - Q_y(p_n) = (p_n - \hat{F}_y(Q_y(p_n)))/f_y(Q_y(p_n)) + o_p(1/\sqrt{n}) \text{ as } \nu \to \infty \text{ under } \mathbf{P}^*.$$

Therefore,

$$\sqrt{n}(\hat{Q}_y(p+1/\sqrt{n}) - \hat{Q}_y(p-1/\sqrt{n}))/2 \xrightarrow{p} 1/f_y(Q_y(p)) \text{ as } \nu \to \infty \text{ under } \mathbf{P}^*.$$

Similarly,

$$\sum_{i\in s} \mathbb{1}_{[X_i \le \hat{Q}_x(p)]}/N\pi_i - \sum_{i\in s} \mathbb{1}_{[X_i \le Q_x(p)]}/N\pi_i \xrightarrow{p} 0 \text{ and}$$
$$\sqrt{n}(\hat{Q}_x(p+1/\sqrt{n}) - \hat{Q}_x(p-1/\sqrt{n}))/2 \xrightarrow{p} 1/f_x(Q_x(p)) \text{ as } \nu \to \infty \text{ under } \mathbf{P}^*.$$

Hence, under $\mathbf{P}^*$, $\hat{K}(p_1,p_2) - \tilde{K}(p_1,p_2) \xrightarrow{p} 0$ as $\nu \to \infty$ for $Q(s,\omega)$ and any $p_1,p_2 \in [\alpha,\beta]$. This completes the proof of *(i)*. The proof of *(ii)* follows exactly the same way as the proof of *(i)*. □

Next, suppose that $P(s,\omega)$ denotes the stratified multistage cluster sampling design with SRSWOR mentioned in Section 4.3. Fix $k \ge 1$ and $p_1,\ldots,p_k \in (0,1)$. Recall $\mathbf{R}'_{hjl}$ from the paragraph preceding Assumption 4.3.5. Define

$$\mathbf{V}'_{hjl} = \mathbf{R}'_{hjl} - \bar{\mathbf{R}}' \text{ and } \hat{\bar{\mathbf{V}}}_3 = \sum_{h=1}^{H}\sum_{j\in s_h}\sum_{l\in s_{hj}} M_h N_{hj}\mathbf{V}'_{hjl}/m_h r_h N$$

for h=$1,\ldots,H$, j=$1,\ldots,M_h$ and l=$1,\ldots,N_{hj}$, where $\bar{\mathbf{R}}'$=$\sum_{h=1}^{H}\sum_{j=1}^{M_h}\sum_{l=1}^{N_{hj}}\mathbf{R}'_{hjl}/N$. Now, we state the following lemma.

Lemma 4.8.6. *(i) Fix $\boldsymbol{m} \in \mathbb{R}^{2k}$ such that $\boldsymbol{m} \ne 0$. Suppose that H is fixed as $\nu \to \infty$, and Assumptions 4.2.1, 4.3.1 and 4.3.3 hold. Then, under $P(s,\omega)$,*

$$\sqrt{n}\boldsymbol{m}\hat{\bar{\boldsymbol{V}}}_3^T \xrightarrow{\mathcal{L}} N(0,\lambda\boldsymbol{m}\Gamma_7\boldsymbol{m}^T) \text{ as } \nu \to \infty \text{ a.s. } [\boldsymbol{P}]$$

for some p.d. matrix Γ_7, where λ is as in Assumption 4.2.1.
(ii) Further, if $H \to \infty$ as $\nu \to \infty$, and Assumptions 4.2.1 and 4.3.3–4.3.5 hold, then the same result holds.

Proof. Note that

$$\sqrt{n}\mathbf{m}\hat{\bar{\mathbf{V}}}_3^T = \sqrt{n}\sum_{h=1}^{H}\sum_{j\in s_h}\sum_{l\in s_{hj}} M_h N_{hj}\mathbf{V}'_{hjl}\mathbf{m}^T/m_h r_h N = \sum_{h=1}^{H}\mathcal{T}_h \text{ (say)}.$$

(i) We shall first show that $\mathcal{T}_h$=$\sqrt{n}\sum_{j\in s_h}\sum_{l\in s_{hj}} M_h N_{hj}\mathbf{V}'_{hjl}\mathbf{m}^T/m_h r_h N$ is asymptotically normal under two stage cluster sampling design with SRSWOR for each h=$1,\ldots,H$. Then, the asymptotic normality of $\sum_{h=1}^{H}\mathcal{T}_h$ follows from the independence of $\{\mathcal{T}_h\}_{h=1}^{H}$. For establishing the asymptotic normality of $\mathcal{T}_h$, we shall use Theorem 2.1 in [62].

Let Θ_h=$\sum_{j\in s_h}\sum_{l=1}^{N_{hj}}\mathbf{V}'_{hjl}\mathbf{m}^T/\sqrt{m_h}$ for h=$1,\ldots,H$. Note that $\Theta_h/\sqrt{m_h}$ is the HT estimator of $\sum_{j=1}^{M_h}\sum_{l=1}^{N_{hj}}\mathbf{V}'_{hjl}\mathbf{m}^T/M_h$ under SRSWOR. Also, note that Assumption 4.2.2-*(ii)* holds trivially under SRSWOR. It follows from Assumptions 4.2.1 and 4.3.1 that $\sum_{j=1}^{M_h}|\sum_{l=1}^{N_{hj}}\mathbf{V}'_{hjl}\mathbf{m}^T|^{2+\delta}/M_h$=$O(1)$ as $\nu\to\infty$ for any $0<\delta\leq 2$ and $\omega\in\Omega$.

Now, it can be shown that $var(\Theta_h)$=$\sigma^2_{h,1}-\sigma^2_{h,2}+\sigma^2_{h,3}$. Here,

$$\sigma^2_{h,1} = (1-f_h)\sum_{j=1}^{M_h} N_{hj}^2\big((\bar{\mathbf{R}}'_{hj}-\bar{\mathbf{R}}')\mathbf{m}^T\big)^2/(M_h-1),$$

$$\sigma^2_{h,2} = 2(1-f_h)N_h\big((\bar{\mathbf{R}}'_h-\bar{\mathbf{R}}')\mathbf{m}^T\big)\sum_{j=1}^{M_h} N_{hj}\big((\bar{\mathbf{R}}'_{hj}-\bar{\mathbf{R}}')\mathbf{m}^T\big)/M_h(M_h-1)$$

$$\text{and } \sigma^2_{h,3} = (1-f_h)N_h^2\big((\bar{\mathbf{R}}'_h-\bar{\mathbf{R}}')\mathbf{m}^T\big)^2/M_h(M_h-1)$$

with f_h=m_h/M_h, $\bar{\mathbf{R}}'_{hj}$=$\sum_{l=1}^{N_{hj}}\mathbf{R}'_{hjl}/N_{hj}$ and $\bar{\mathbf{R}}'_h$=$\sum_{j=1}^{M_h}\sum_{l=1}^{N_{hj}}\mathbf{R}'_{hjl}/N_h$. Next, we note that

$$\sigma^2_{h,1} = (1-f_h)\Bigg(\sum_{j=1}^{M_h} N_{hj}^2\big(\bar{\mathbf{R}}'_{hj}\mathbf{m}^T\big)^2 - 2\big(\bar{\mathbf{R}}'\mathbf{m}^T\big)\sum_{j=1}^{M_h} N_{hj}^2\big(\bar{\mathbf{R}}'_{hj}\mathbf{m}^T\big)+ \tilde{N}_h(\bar{\mathbf{R}}'\mathbf{m}^T)^2\Bigg)\Bigg/(M_h-1), \quad (4.8.11)$$

where $\tilde{N}_h$=$\sum_{j=1}^{M_h}N_{hj}^2$. Let us consider the first term on the right hand side of (4.8.11). Using Assumptions 4.2.1 and 4.3.1, and Hoeffding's inequality, it can be shown that

$$(1-f_h)\sum_{j=1}^{M_h} N_{hj}^2\big((\bar{\mathbf{R}}'_{hj}\mathbf{m}^T)^2 - E_{\mathbf{P}}(\bar{\mathbf{R}}'_{hj}\mathbf{m}^T)^2\big)/(M_h-1)\to 0 \text{ as } \nu\to\infty \textit{ a.s. } [\mathbf{P}].$$

Further, we have

$$(1-f_h)\sum_{j=1}^{M_h} N_{hj}^2 E_{\mathbf{P}}(\bar{\mathbf{R}}'_{hj}\mathbf{m}^T)^2/(M_h-1) = (1-f_h)(N_h\tilde{\sigma}_h^2+\tilde{N}_h\mu_h^2)/(M_h-1),$$

where $\tilde{\sigma}_h^2$=$E_{\mathbf{P}}\big[\big(\mathbf{R}'_{hjl}-E_{\mathbf{P}}(\mathbf{R}'_{hjl})\big)\mathbf{m}^T\big]^2$=$\mathbf{m}\Gamma_h\mathbf{m}^T$ (recall Γ_h from the paragraph preceding Assumption 4.3.5) and μ_h=$E_{\mathbf{P}}$ $(\mathbf{R}'_{hjl}\mathbf{m}^T)$. Thus

$$(1-f_h)\sum_{j=1}^{M_h} N_{hj}^2 (\bar{\mathbf{R}}'_{hj}\mathbf{m}^T)^2/(M_h-1) = (1-f_h)(N_h\tilde{\sigma}_h^2+\tilde{N}_h\mu_h^2)/(M_h-1)+o(1) \quad (4.8.12)$$

as $\nu\to\infty$ *a.s.* [**P**]. Using similar arguments, we can say that

$$\sigma_{h,1}^2 = (1-f_h)(N_h\tilde{\sigma}_h^2+\tilde{N}_h(\mu_h-\tilde{\mu})^2)/(M_h-1)+o(1),$$
$$\sigma_{h,2}^2 = 2(1-f_h)N_h^2(\mu_h-\tilde{\mu})^2/M_h(M_h-1)+o(1) \text{ and}$$
$$\sigma_{h,3}^2 = (1-f_h)N_h^2(\mu_h-\tilde{\mu})^2/M_h(M_h-1)+o(1) \text{ as } \nu\to\infty \textit{ a.s. } [\mathbf{P}],$$

where $\tilde{\mu}$=$\sum_{h=1}^H \Lambda_h\mu_h$ (recall Λ_h's from Assumption 4.3.1). Then, we have

$$var(\Theta_h) = (1-f_h)N_h\tilde{\sigma}_h^2/(M_h-1)+o(1) \quad (4.8.13)$$

as $\nu\to\infty$ *a.s.* [**P**] by Assumption 4.3.1.

Next, recall $F_{y,H}(t)$ and $F_{x,H}(t)$ from the paragraph preceding Assumption 4.3.5. It can be shown that

$$\sup_{t\in\mathbb{R}}|F_{y,H}(t)-\tilde{F}_{y,H}(t)|\to 0 \text{ and } \sup_{t\in\mathbb{R}}|F_{x,H}(t)-\tilde{F}_{x,H}(t)|\to 0 \text{ as } \nu\to\infty \quad (4.8.14)$$

by Assumption 4.3.1, where $\tilde{F}_{y,H}(t)$=$\sum_{h=1}^H \Lambda_h F_{y,h}(t)$ and $\tilde{F}_{x,H}(t)$=$\sum_{h=1}^H \Lambda_h F_{x,h}(t)$. Then, it follows from Lemma 4.8.8 that

$$Q_{y,H}(p_r)\to\tilde{Q}_{y,H}(p_r) \text{ as } \nu\to\infty \text{ for any } r=1,\ldots,k, \quad (4.8.15)$$

where $\tilde{Q}_{y,H}(p)$=inf$\{t\in\mathbb{R}:\tilde{F}_{y,H}(t)\geq p\}$. Similarly,

$$Q_{x,H}(p_r)\to\tilde{Q}_{x,H}(p_r) \text{ as } \nu\to\infty \text{ for any } r=1,\ldots,k, \quad (4.8.16)$$

where $\tilde{Q}_{x,H}(p)$=inf$\{t\in\mathbb{R}:\tilde{F}_{x,H}(t)\geq p\}$. Let

$$\tilde{\mathbf{R}}_{hjl} = \left(\mathbb{1}_{[Y'_{hjl} \le \tilde{Q}_{y,H}(p_1)]}, \ldots, \mathbb{1}_{[Y'_{hjl} \le \tilde{Q}_{y,H}(p_k)]}, \mathbb{1}_{[X'_{hjl} \le \tilde{Q}_{x,H}(p_1)]}, \ldots, \mathbb{1}_{[X'_{hjl} \le \tilde{Q}_{x,H}(p_k)]}\right),$$

where (Y'_{hjl}, X'_{hjl}) is as in the second paragraph of Section 4.3. Then,

$$\tilde{\sigma}_h^2 = \mathbf{m}\Gamma_h\mathbf{m}^T \to mE_{\mathbf{P}}\big(\tilde{\mathbf{R}}_{hjl} - E_{\mathbf{P}}(\tilde{\mathbf{R}}_{hjl})\big)^T\big(\tilde{\mathbf{R}}_{hjl} - E_{\mathbf{P}}(\tilde{\mathbf{R}}_{hjl})\big)\mathbf{m}^T$$

as $\nu \to \infty$ for any h=$1, \ldots, H$ in view of Assumption 4.3.3. Moreover, $E_{\mathbf{P}}\big(\tilde{\mathbf{R}}_{hjl} - E_{\mathbf{P}}(\tilde{\mathbf{R}}_{hjl})\big)^T \times \big(\tilde{\mathbf{R}}_{hjl} - E_{\mathbf{P}}(\tilde{\mathbf{R}}_{hjl})\big)$ is a p.d. matrix because Assumption 4.3.2 holds. Therefore,

$$\underline{\lim}_{\nu\to\infty}((M_h - 1)/M_h)var(\Theta_h) > 0 \ a.s.\ [\mathbf{P}]$$

by (4.8.13) above and Assumption 4.3.1. Hence, one can show that

$$(\Theta_h - E(\Theta_h))/\sqrt{var(\Theta_h)} \xrightarrow{\mathcal{L}} N(0,1) \text{ as } \nu \to \infty \text{ under SRSWOR } \ a.s.\ [\mathbf{P}]$$

in the same way as the derivation of the result, which appears in the proof of Lemma 2.7.2 in Section 2.7 of Chapter 2, that $\sqrt{n}\mathbf{m}_1(\hat{\overline{\mathbf{V}}}_1 - \overline{\mathbf{V}})^T \xrightarrow{\mathcal{L}} N(0, \mathbf{m}_1\Gamma_1\mathbf{m}_1^T)$ as $\nu \to \infty$ under SRSWOR for any $\mathbf{m}_1 \in \mathbb{R}^p$, $\mathbf{m}_1 \neq 0$ and Γ_1=$\lim_{\nu\to\infty}\Sigma_1$. Thus the condition C1 of Theorem 2.1 in [62] holds *a.s.* [**P**].

Next, suppose that $\overline{\mathbf{V}}'_{hj}$=$\sum_{l=1}^{N_{hj}} \mathbf{V}'_{hjl}/N_{hj}$. Note that for any h=$1, \ldots, H$, $\sum_{l=1}^{N_{hj}}\big((\mathbf{V}'_{hjl} - \overline{\mathbf{V}}'_{hj})\mathbf{m}^T\big)^2/N_{hj}$ are independent bounded random variables for $1 \le j \le M_h$. Then, by Assumptions 4.2.1 and 4.3.1, and Hoeffding's inequality, we have

$$\sum_{j=1}^{M_h}(N_{hj}^2/r_h)(1/m_h)\left[\sum_{l=1}^{N_{hj}}\big((\mathbf{V}'_{hjl} - \overline{\mathbf{V}}'_{hj})\mathbf{m}^T\big)^2/N_{hj}\right] = (1/r_h m_h)\sum_{j=1}^{M_h}N_{hj}(N_{hj}-1)\tilde{\sigma}_h^2 + o(1)$$

as $\nu \to \infty$ *a.s.* [**P**]. Thus

$$\underline{\lim}_{\nu\to\infty}\Big[\sum_{j=1}^{M_h}(N_{hj}^2/r_h)(1/m_h)\Big\{\sum_{l=1}^{N_{hj}}\big((\mathbf{V}'_{hjl} - \overline{\mathbf{V}}'_{hj})\mathbf{m}^T\big)^2/N_{hj}\Big\}\Big]^2 > 0$$

a.s. [**P**]. Further, in view of Assumption 4.3.1, we have

$$\left[\sum_{j=1}^{M_h}(N_{hj}^4/r_h^2)(1/M_h)\left\{\sum_{l=1}^{N_{hj}}\big((\mathbf{V}'_{hjl} - \overline{\mathbf{V}}'_{hj})\mathbf{m}^T\big)^2/N_{hj}\right\}^2\right] \le K$$

for all sufficiently large ν and some constant $K > 0$ *a.s.* [**P**]. Therefore,

$$\lim_{\nu\to\infty}\Bigg[\sum_{j=1}^{M_h}(N_{hj}^4/r_h^2)(M_h/m_h)^3\Big\{\sum_{l=1}^{N_{hj}}\big((\mathbf{V}'_{hjl}-\overline{\mathbf{V}}'_{hj})\mathbf{m}^T\big)^2/N_{hj}\Big\}^2\Bigg]\Big/$$
$$\Bigg[\sum_{j=1}^{M_h}(N_{hj}^2/r_h)(M_h/m_h)\Big\{\sum_{l=1}^{N_{hj}}\Big((\mathbf{V}'_{hjl}-\overline{\mathbf{V}}'_{hj})\mathbf{m}^T\Big)^2/N_{hj}\Big\}\Bigg]^2=0$$

a.s. [**P**] by Assumption 4.3.1. Thus the condition C2 of Theorem 2.1 in [62] holds *a.s.* [**P**] by Assumption 4.3.1 and Proposition 4.1 in [62].

The condition C3 of Theorem 2.1 in [62] holds for any $\omega \in \Omega$ by (b) of Proposition 2.3 in [62] since SRSWOR is used to select samples from clusters in the 1^{st} stage and from population units of the selected clusters in the 2^{nd} stage. Therefore, the conditions C1, C2 and C3 of Theorem 2.1 in [62] hold *a.s.* [**P**]. Hence, by Theorem 2.1 in [62], we have

$$(\sqrt{n_h}/N_h)(N/\sqrt{n})(\mathcal{T}_h-E(\mathcal{T}_h))/(var((\sqrt{n_h}/N_h)(N/\sqrt{n})\mathcal{T}_h))^{1/2}\xrightarrow{\mathcal{L}}N(0,1) \qquad (4.8.17)$$

as $\nu \to \infty$ under two stage cluster sampling design with SRSWOR *a.s.* [**P**] for any h=$1,\ldots,H$. Now,

$$var((\sqrt{n_h}/N_h)(N/\sqrt{n})\mathcal{T}_h)=\sum_{j=1}^{M_h}\tilde{c}_{hj}\big((\bar{\mathbf{R}}'_{hj}-\bar{\mathbf{R}}')\mathbf{m}^T\big)^2-\tilde{c}_h\big((\bar{\mathbf{R}}'_h-N_h\bar{\mathbf{R}}'/M_h)\mathbf{m}^T\big)^2+$$
$$\sum_{j=1}^{M_h}\tilde{d}_{hj}\sum_{l=1}^{N_{hj}}\big((\mathbf{R}'_{hjl}-\bar{\mathbf{R}}'_{hj})\mathbf{m}^T\big)^2,$$

where

$$\tilde{c}_{hj}=(N/N_h)^2(n_h/n)c_{hj},\tilde{d}_{hj}=(N/N_h)^2(n_h/n)d_{hj},\text{ and }\tilde{c}_h=(N/N_h)^2(n_h/n)c_h.$$

Here,

$$c_{hj}=c_hN_{hj}^2/M_h,d_{hj}=nM_h(1-f_{hj})N_{hj}^2/m_hr_h(N_{hj}-1)N^2,$$
$$c_h=nM_h^3(1-f_h)/m_h(M_h-1)N^2,f_h=m_h/M_h,\text{ and }f_{hj}=r_h/N_{hj}.$$

It can be shown using Hoeffding's inequality that

$$var((\sqrt{n_h}/N_h)(N/\sqrt{n})\mathcal{T}_h)=(1-n_h/N_h)\tilde{\sigma}_h^2+o(1)\text{ as }\nu\to\infty\text{ a.s. }[\mathbf{P}].$$

Therefore, using (4.8.17) above and Assumption 4.3.1, it can be shown that

$$\sum_{h=1}^{H}\mathcal{T}_h=\sum_{h=1}^{H}(\mathcal{T}_h-E(\mathcal{T}_h))\xrightarrow{\mathcal{L}}N(0,\Delta^2)\text{ as }\nu\to\infty\ a.s.\ [\mathbf{P}].$$

Here,

$$\Delta^2=\lim_{\nu\to\infty}\sum_{h=1}^{H}nN_h(N_h-n_h)\tilde{\sigma}_h^2/n_hN^2=\lim_{\nu\to\infty}\sum_{h=1}^{H}nN_h(N_h-n_h)\mathbf{m}\Gamma_h\mathbf{m}^T/n_hN^2$$
$$=\lambda\sum_{h=1}^{H}\Lambda_h(\Lambda_h/\lambda\lambda_h-1)E_{\mathbf{P}}\big(\tilde{\mathbf{R}}_{hjl}\mathbf{m}^T-E_{\mathbf{P}}(\tilde{\mathbf{R}}_{hjl}\mathbf{m}^T)\big)^2=\lambda\mathbf{m}\Gamma_7\mathbf{m}^T>0$$

with Γ_7=$\sum_{h=1}^{H}\Lambda_h(\Lambda_h/\lambda\lambda_h-1)E_{\mathbf{P}}\big(\tilde{\mathbf{R}}_{hjl}-E_{\mathbf{P}}(\tilde{\mathbf{R}}_{hjl})\big)^T\ \big(\tilde{\mathbf{R}}_{hjl}-E_{\mathbf{P}}(\tilde{\mathbf{R}}_{hjl})\big)$. This completes the proof of *(i)*.

(ii) Since, population units are sampled independently across the strata in $P(s,\omega)$, asymptotic normality of $\sum_{h=1}^{H}\mathcal{T}_h$ under $P(s,\omega)$ follows by applying Lyapunov's central limit theorem (CLT) to independent random variables $\{\mathcal{T}_h\}_{h=1}^{H}$. Note that for any $\delta>0$, we have

$$|\mathcal{T}_h|^{2+\delta}\leq\epsilon(\nu)(m_h/\sqrt{n})^{2+\delta}$$

by Assumption 4.3.4, where $\epsilon(\nu)$ does not depend on s and ω, and $\epsilon(\nu)$=$O(1)$ as $\nu\to\infty$. Therefore, under $P(s,\omega)$,

$$\sum_{h=1}^{H}E|\mathcal{T}_h|^{2+\delta}\leq\epsilon(\nu)(H/n^{1+\delta/2})\sum_{h=1}^{H}M_h^{2+\delta}/H=O(n^{-\delta/2})$$

as $\nu\to\infty$ for any $0<\delta\leq 2$ and $\omega\in\Omega$. Hence, under $P(s,\omega)$, $\sum_{h=1}^{H}E|\mathcal{T}_h-E(\mathcal{T}_h)|^{2+\delta}\to 0$ as $\nu\to\infty$ for any $0<\delta\leq 2$ and $\omega\in\Omega$.

Next, we have

$$\sum_{h=1}^{H}var(\mathcal{T}_h)=\sum_{h=1}^{H}\sum_{j=1}^{M_h}c_{hj}\big((\bar{\mathbf{R}}'_{hj}-\bar{\mathbf{R}}')\mathbf{m}^T\big)^2-\sum_{h=1}^{H}c_h\big((\bar{\mathbf{R}}'_h-N_h\bar{\mathbf{R}}'/M_h)\mathbf{m}^T\big)^2$$
$$+\sum_{h=1}^{H}\sum_{j=1}^{M_h}d_{hj}\sum_{l=1}^{N_{hj}}\big((\mathbf{R}'_{hjl}-\bar{\mathbf{R}}'_{hj})\mathbf{m}^T\big)^2=\Delta_1^2-\Delta_2^2+\Delta_3^2\text{ (say).}$$

Now, it can be shown using Assumptions 4.2.1 and 4.3.4, and Hoeffding's inequality that

$$\Delta_1^2 - \Delta_2^2 + \Delta_3^2 = \sum_{h=1}^{H} c_h(\tilde{N}_h - N_h^2/M_h)(\mu_h - \mu^*)^2/M_h + \sum_{h=1}^{H} nM_h\tilde{N}_h\tilde{\sigma}_h^2/$$
$$m_h r_h N^2 - \sum_{h=1}^{H} nN_h\tilde{\sigma}_h^2/N^2 + o(1) \text{ as } \nu \to \infty \textit{ a.s. } [\mathbf{P}], \tag{4.8.18}$$

in the same way as the derivation of the result in (4.8.12). Here, μ^*=$\sum_{h=1}^{H} N_h\mu_h/N$. The first term on the right hand side of (4.8.18) converges to 0 as $\nu \to \infty$ by Assumption 4.2.1 and Assumption 4.3.4. Moreover, we have

$$\sum_{h=1}^{H} nM_h\tilde{N}_h\tilde{\sigma}_h^2/m_h r_h N^2 - \sum_{h=1}^{H} nN_h\tilde{\sigma}_h^2/N^2$$
$$= (n/N^2)\sum_{h=1}^{H} M_h(\tilde{N}_h - N_h^2/M_h)\tilde{\sigma}_h^2/m_h r_h + (n/N^2)\sum_{h=1}^{H} N_h(N_h - n_h)\tilde{\sigma}_h^2/n_h. \tag{4.8.19}$$

The first term on the right hand side of (4.8.19) converges to 0 and

$$(n/N^2)\sum_{h=1}^{H} N_h(N_h - n_h)\tilde{\sigma}_h^2/n_h = \lambda\sum_{h=1}^{H} N_h(N_h - n_h)\tilde{\sigma}_h^2/n_h N + o(1) \text{ as } \nu \to \infty$$

by Assumption 4.3.4. Therefore,

$$\Delta_1^2 - \Delta_2^2 + \Delta_3^2 = \lambda\sum_{h=1}^{H} N_h(N_h - n_h)\tilde{\sigma}_h^2/n_h N + o(1),$$

and hence

$$\sum_{h=1}^{H} var(\mathcal{T}_h) = \Delta_1^2 - \Delta_2^2 + \Delta_3^2 \to \lambda\mathbf{m}\Gamma_1\mathbf{m}^T > 0$$

as $\nu \to \infty$ *a.s.* [**P**] for some p.d. matrix Γ_1 in view of Assumption 4.3.5. Here, Γ_1 is as in Assumption 4.3.5. Thus the Lyapunov's condition $\sum_{h=1}^{H} E|\mathcal{T}_h - E(\mathcal{T}_h)|^{2+\delta}/(\sum_{h=1}^{H} var(\mathcal{T}_h))^{1+\delta/2} \to 0$ as $\nu \to \infty$ for some $\delta > 0$, holds under $P(s,\omega)$ *a.s.* [**P**]. Consequently, $\sum_{h=1}^{H} \mathcal{T}_h \xrightarrow{\mathcal{L}} N(0, \lambda\mathbf{m}\Gamma_6\mathbf{m}^T)$ as $\nu \to \infty$ *a.s.* [**P**] with Γ_7=Γ_1. This completes the proof of *(ii)*. □

Next, consider $\{U_i\}_{i=1}^N$ as in (4.2.2) in Section 4.2 with $F_{y,H}$ replacing F_y. Also, consider $B_{u,N}(t_1,t_2)$ and $\mathbb{B}_n(t_1,t_2)$ as in (4.7.3) in Section 4.7. Now, we state the following lemma.

Lemma 4.8.7. *(i) Suppose that H is fixed as $\nu \to \infty$, and Assumptions 4.2.1, 4.3.1 and 4.3.3 hold. Then, under $P(s,\omega)$,*

$$E\big[\big(\mathbb{B}_n(t_1,t_2)\big)^2\big(\mathbb{B}_n(t_2,t_3)\big)^2\big] \leq L_1\big(B_{u,N}(t_1,t_3)\big)^2 \text{ a.s. } [\boldsymbol{P}]$$

for any $0 \leq t_1 < t_2 < t_3 \leq 1$, $\nu \geq 1$ *and some constant* $L_1 > 0$, *and*

$$\overline{\lim}_{\nu\to\infty} E\big(\mathbb{B}_n(t_1,t_2)\big)^4 \leq L_2(t_2-t_1)^2 \text{ a.s. } [\boldsymbol{P}]$$

for any $0 \leq t_1 < t_2 \leq 1$ *and some constant* $L_2 > 0$.

(ii) Further, if $H \to \infty$ *as* $\nu \to \infty$*, and Assumptions 4.2.1, 4.3.3 and 4.3.4 hold, then the same results hold.*

Proof. Recall Y'_{hjl} from the 2^{nd} paragraph in Section 4.3. Let us define U'_{hjl}=$F_{y,H}(Y'_{hjl})$ for any given h=$1,\ldots,H$, j=$1,\ldots,M_h$ and l=$1,\ldots,N_{hj}$. Consider $F_{u,N}(t)$ and $\mathbb{U}_n(t)$ as in (4.7.1) in Section 4.7. Recall from Section 4.3 that given any h, j and l, Y'_{hjl}=Y_i for some $i \in \{1,\ldots,N\}$. Also, recall from Section 4.3 that under $P(s,\omega)$, the inclusion probability of the i^{th} population unit is π_i=$m_h r_h/M_h N_{hj}$ if it belongs to the j^{th} cluster of the h^{th} stratum. Then, we have $\mathbb{U}_n(t)$=$\sqrt{n}\sum_{h=1}^{H}\sum_{j\in s_h}\sum_{l\in s_{hj}} M_h N_{hj}\big(\mathbb{1}_{[U'_{hjl}\leq t]} - F_{u,N}(t)\big)/m_h r_h N$.

Now, suppose that for h=$1,\ldots,H$, j=$1,\ldots,M_h$ and l=$1,\ldots,N_{hj}$,

$$\xi_{hjl} = \begin{cases} 1, & \text{if the } l^{th} \text{ unit of the } j^{th} \text{ cluster in the } h^{th} \text{ stratum is selected in the sample, and} \\ 0, & \text{otherwise.} \end{cases}$$

Then, we have

$$\mathbb{Z}_n(t) = (\sqrt{n}/N)\sum_{h=1}^{H}\sum_{j=1}^{M_h}\sum_{l=1}^{N_{hj}}\big((M_h N_{hj}\xi_{hjl}/m_h r_h) - 1\big)\big(\mathbb{1}_{[Z'_{hjl}\leq t]} - F_{z,N}(t)\big).$$

Further, suppose that

$$\tilde{\alpha}_h = \sum_{j=1}^{M_h}\sum_{l=1}^{N_{hj}}\big((M_h N_{hj}\xi_{hjl}/m_h r_h) - 1\big)\bar{A}_{hjl} \text{ and}$$
$$\tilde{\beta}_h = \sum_{j=1}^{M_h}\sum_{l=1}^{N_{hj}}\big((M_h N_{hj}\xi_{hjl}/m_h r_h) - 1\big)\bar{B}_{hjl}$$

for h=$1,\ldots,H$ and $0 \leq t_1 < t_2 < t_3 \leq 1$, where $\bar{A}_{hjl}$=$\mathbb{1}_{[t_1<Z'_{hjl}\leq t_2]} - B_{z,N}(t_1,t_2)$ and

$\bar{B}_{hjl}$=$\mathbb{1}_{[t_2<Z'_{hjl}\leq t_3]} - B_{z,N}(t_2,t_3)$. Now, let us define $S_{k,H}$={$(h_1,\ldots,h_k)$: $h_1,\ldots,h_k \in \{1,2,\ldots,H\}$ and $h_1,\ldots,h_k$ are all distinct} for k=2, 3, 4. Then, we have

$$E\big[\big(\mathbb{B}_n(t_1,t_2)\big)^2\big(\mathbb{B}_n(t_2,t_3)\big)^2\big] = (n^2/N^4)E\Big[\sum_{h=1}^{H}\tilde{\alpha}_h^2\tilde{\beta}_h^2+ \sum_{(h_1,h_2)\in S_{2,H}}\tilde{\alpha}_{h_1}^2\tilde{\beta}_{h_2}^2+\sum_{(h_1,h_2)\in S_{2,H}}\tilde{\alpha}_{h_1}^2\tilde{\beta}_{h_1}\tilde{\beta}_{h_2}+\sum_{(h_1,h_2)\in S_{2,H}}\tilde{\alpha}_{h_1}\tilde{\alpha}_{h_2}\tilde{\beta}_{h_2}^2+ \sum_{(h_1,h_2)\in S_{2,H}}\tilde{\alpha}_{h_1}\tilde{\alpha}_{h_2}\tilde{\beta}_{h_1}\tilde{\beta}_{h_2}+\sum_{(h_1,h_2,h_3)\in S_{3,H}}\tilde{\alpha}_{h_1}^2\tilde{\beta}_{h_2}\tilde{\beta}_{h_3}+ \sum_{(h_1,h_2,h_3)\in S_{3,H}}\tilde{\alpha}_{h_1}\tilde{\alpha}_{h_2}\tilde{\beta}_{h_3}^2+\sum_{(h_1,h_2,h_3,h_4)\in S_{4,H}}\tilde{\alpha}_{h_1}\tilde{\alpha}_{h_2}\tilde{\beta}_{h_3}\tilde{\beta}_{h_4}\Big]. \tag{4.8.20}$$

(i) Suppose that $\overline{\alpha}_{hjl}$=$\big((M_hN_{hj}\xi_{hjl}/m_hr_h)-1\big)\bar{A}_{hjl}$, $\overline{\beta}_{hjl}$=$\big((M_hN_{hj}\xi_{hjl}/m_hr_h)-1\big)\bar{B}_{hjl}$, α^*_{hj}=$\sum_{l=1}^{N_{hj}}\overline{\alpha}_{hjl}$ and β^*_{hj}=$\sum_{l=1}^{N_{hj}}\overline{\beta}_{hjl}$ for h=1, …, H, j=1, …, M_h, l=1, …, N_{hj} and $0\leq t_1<t_2<t_3\leq 1$. Then, we have $\tilde{\alpha}_h$=$\sum_{j=1}^{M_h}\alpha^*_{hj}$ and $\tilde{\beta}_h$=$\sum_{j=1}^{M_h}\beta^*_{hj}$. Now, let us consider the first term on right hand side of (4.8.20). Further, suppose that $S_{k,h}$={$(j_1,\ldots,j_k)$: $j_1,\ldots,j_k\in s_h$ and $j_1,\ldots,j_k$ are all distinct}, k=2, 3, 4, $1\leq h\leq H$. Then, we have

$$(n^2/N^4)\sum_{h=1}^{H}E(\tilde{\alpha}_h^2\tilde{\beta}_h^2) = (n^2/N^4)\sum_{h=1}^{H}E\Big[\sum_{j=1}^{M_h}(\alpha^*_{hj}\beta^*_{hj})^2+\sum_{(j_1,j_2)\in S_{2,h}}(\alpha^*_{hj_1}\beta^*_{hj_2})^2 +\sum_{(j_1,j_2)\in S_{2,h}}(\alpha^*_{hj_1})^2\beta^*_{hj_1}\beta^*_{hj_2}+\sum_{(j_1,j_2)\in S_{2,h}}\alpha^*_{hj_1}\alpha^*_{hj_2}(\beta^*_{hj_2})^2+ \sum_{(j_1,j_2)\in S_{2,h}}\alpha^*_{hj_1}\alpha^*_{hj_2}\beta^*_{hj_1}\beta^*_{hj_2}+\sum_{(j_1,j_2,j_3)\in S_{3,h}}(\alpha^*_{hj_1})^2\beta^*_{hj_2}\beta^*_{hj_3}+ \sum_{(j_1,j_2,j_3)\in S_{3,h}}\alpha^*_{hj_1}\alpha^*_{hj_2}(\beta^*_{hj_3})^2+\sum_{(j_1,j_2,j_3,j_4)\in S_{4,h}}\alpha^*_{hj_1}\alpha^*_{hj_2}\beta^*_{hj_3}\beta^*_{hj_4}\Big]. \tag{4.8.21}$$

Next, consider the first term on the right hand side of (4.8.21). Suppose that $S_{k,hj}$={$(l_1,\ldots,l_k)$: $l_1,\ldots,l_k\in\{1,\ldots,N_{hj}\}$ and $l_1,\ldots,l_k$ are all distinct}, k=2, 3, 4, j=1, …, M_h and $1\leq h\leq H$. Then, we have

$$(n^2/N^4)\sum_{h=1}^{H}E\Big[\sum_{j=1}^{M_h}(\alpha^*_{hj}\beta^*_{hj})^2\Big] = (n^2/N^4)\sum_{h=1}^{H}E\Big[\sum_{j=1}^{M_h}\Big(\sum_{l=1}^{N_{hj}}(\overline{\alpha}_{hjl}\overline{\beta}_{hjl})^2 +\sum_{(l_1,l_2)\in S_{2,hj}}(\overline{\alpha}_{hjl_1}\overline{\beta}_{hjl_2})^2+\sum_{(l_1,l_2)\in S_{2,hj}}(\overline{\alpha}_{hjl_1})^2\overline{\beta}_{hjl_1}\overline{\beta}_{hjl_2} +\sum_{(l_1,l_2)\in S_{2,hj}}\overline{\alpha}_{hjl_1}\overline{\alpha}_{hjl_2}(\overline{\beta}_{hjl_2})^2+\sum_{(l_1,l_2)\in S_{2,hj}}\overline{\alpha}_{hjl_1}\overline{\alpha}_{hjl_2}\overline{\beta}_{hjl_1}\overline{\beta}_{hj_2}+ \tag{4.8.22}$$

$$\sum_{(l_1,l_2,l_3)\in S_{3,hj}} (\overline{\alpha}_{hjl_1})^2\overline{\beta}_{hjl_2}\overline{\beta}_{hjl_3} + \sum_{(l_1,l_2,l_3)\in S_{3,hj}} \overline{\alpha}_{hjl_1}\overline{\alpha}_{hjl_2}(\overline{\beta}_{hjl_3})^2$$
$$+ \sum_{(l_1,l_2,l_3,l_4)\in S_{4,hj}} \overline{\alpha}_{hjl_1}\overline{\alpha}_{hjl_2}\overline{\beta}_{hjl_3}\overline{\beta}_{hjl_4}\Bigg)\Bigg].$$

Now, consider the first term on the right hand side of (4.8.22). Note that $N/n=O(1)$ and $N/n=O(1)$ and $\max_{1\leq h\leq H, 1\leq j\leq M_h}(nM_hN_{hj}/r_hm_hN)=O(1)$ as $\nu \to \infty$ by Assumptions 4.2.1 and 4.3.1. Then, we have

$$(n^2/N^4)\sum_{h=1}^{H}\sum_{j=1}^{M_h}\sum_{l=1}^{N_{hj}} E(\overline{\alpha}_{hjl}\overline{\beta}_{hjl})^2 = (n^2/N^4)\sum_{h=1}^{H}\sum_{j=1}^{M_h}\sum_{l=1}^{N_{hj}} E\big((M_hN_{hj}\xi_{hjl}/m_hr_h)$$
$$-1\big)^4\bar{A}_{hjl}^2\bar{B}_{hjl}^2 \leq (K_1/N^2)\sum_{h=1}^{H}\sum_{j=1}^{M_h}\sum_{l=1}^{N_{hj}}\Big(\mathbb{1}_{[t_1<Z'_{hjl}\leq t_2]} + B_{z,N}(t_1,t_2)\Big)\times \tag{4.8.23}$$
$$\Big(\mathbb{1}_{[t_2<Z'_{hjl}\leq t_3]} + B_{z,N}(t_2,t_3)\Big) \leq K_2\big(B_{z,N}(t_1,t_3)\big)^2$$

a.s. [**P**] for all $\nu \geq 1$ and some constants $K_1, K_2 > 0$. Inequalities similar to (4.8.23) can be shown to hold for the other terms on the right hand side of (4.8.22). Thus

$$(n^2/N^4)\sum_{h=1}^{H}\sum_{j=1}^{M_h} E(\alpha^*_{hj}\beta^*_{hj})^2 \leq K_3\big(B_{z,N}(t_1,t_3)\big)^2 \tag{4.8.24}$$

a.s. [**P**] for any $0 \leq t_1 < t_2 < t_3 \leq 1$, $\nu \geq 1$ and some constant $K_3 > 0$. Inequalities similar to (4.8.24) can also be shown to hold for the other terms on the right hand side of (4.8.21). Therefore,

$$(n^2/N^4)\sum_{h=1}^{H} E(\tilde{\alpha}_h^2\tilde{\beta}_h^2) \leq K_4\big(B_{z,N}(t_1,t_3)\big)^2 \tag{4.8.25}$$

a.s. [**P**] for any $0 \leq t_1 < t_2 < t_3 \leq 1$, $\nu \geq 1$ and some constant $K_4 > 0$. Furthermore, inequalities similar to (4.8.25) can be shown to hold for the other terms on the right hand side of (4.8.20). Consequently, $E\big[\big(\mathbb{B}_n(t_1,t_2)\big)^2\big(\mathbb{B}_n(t_2,t_3)\big)^2\big] \leq K_5\big(B_{z,N}(t_1,t_3)\big)^2$ *a.s.* [**P**] for any $0 \leq t_1 < t_2 < t_3 \leq 1$, $\nu \geq 1$ and some constant $K_5 > 0$. Moreover, it can be shown in the same way that $\overline{\lim}_{\nu\to\infty}E\big(\mathbb{B}_n(u,t)\big)^4 \leq K_6(t-u)^2$ *a.s.* [**P**] for any $0 \leq u < t \leq 1$ and some constant $K_6 > 0$ because $B_{z,N}(u,t) \to (t-u)$ as $\nu \to \infty$ *a.s.* [**P**] by Assumption 4.3.3 and SLLN. This completes the proof of *(i)*.

(ii) It follows from Assumptions 4.2.1 and 4.3.4 that $N/n=O(1)$ and $\max_{1\leq h\leq H, 1\leq j\leq M_h}$ $(nM_hN_{hj}/r_hm_hN)=O(1)$ as $\nu \to \infty$. Then the proof of the result in *(ii)* follows the same way as the proof of the result in *(i)*. □

Next, recall λ_h's from Assumption 4.3.1, $F_{y,H}$ and $Q_{y,H}$ from the paragraph preceding Assumption 4.3.5 and $\tilde{F}_y$ from Assumption 4.3.6. Let us define $\tilde{Q}_y(p)$=inf$\{t \in \mathbb{R} : \tilde{F}_y(t) \geq p\}$ for $0 < p < 1$. Also, recall $\tilde{F}_{y,H}$ and $\tilde{Q}_{y,H}$ from the paragraph containing (4.8.14)–(4.8.16) in the proof of *(i)* in Lemma 4.8.6. Then, we state the following lemma.

Lemma 4.8.8. *(i) Suppose that H is fixed as $\nu \to \infty$, and Assumptions 4.3.1 and 4.3.3 hold. Then, for any* $0 < \alpha < \beta < 1$*,*

$$\sup_{p\in[\alpha,\beta]} |Q_{y,H}(p) - \tilde{Q}_{y,H}(p)| \to 0 \text{ as } \nu \to \infty.$$

(ii) Further, suppose that $H \to \infty$ as $\nu \to \infty$, and Assumptions 4.3.3, 4.3.4 and 4.3.6 hold. Then, for any $0 < \alpha < \beta < 1$*,*

$$\sup_{p\in[\alpha,\beta]} |Q_{y,H}(p) - \tilde{Q}_y(p)| \to 0 \text{ as } \nu \to \infty.$$

Proof. *(i)* Note that the inverse of $F_{y,H}|_{\mathcal{C}_y}$, say $F_{y,H}^{-1} : (0,1) \to \mathcal{C}_y$, exists and is differentiable by Assumption 4.3.3, and $F_{y,H}^{-1}(p)$=$Q_{y,H}(p)$ for any $0 < p < 1$. Also, note that the inverse of $\tilde{F}_{y,H}|_{\mathcal{C}_y}$, say $\tilde{F}_{y,H}^{-1} : (0,1) \to \mathcal{C}_y$, exists and is differentiable, and $\tilde{F}_{y,H}^{-1}(p)$=$\tilde{Q}_{y,H}(p)$ for any $0 < p < 1$. Clearly, $\tilde{Q}_{y,H}$ is uniformly continuous on $[\alpha/2, (1+\beta)/2]$. Then, given any $\epsilon > 0$ there exists a $\delta > 0$ such that

$$|\tilde{Q}_{y,H}(p_1) - \tilde{Q}_{y,H}(p_2)| \leq \epsilon, \text{ whenever } |p_1 - p_2| \leq \delta \text{ and } p_1, p_2 \in [\alpha/2, (1+\beta)/2].$$

Now, it follows that

$$\sup_{p\in[\alpha,\beta]} |p - \tilde{F}_{y,H}(Q_{y,H}(p))| = \sup_{p\in[\alpha,\beta]} |F_{y,H}(Q_{y,H}(p)) - \tilde{F}_{y,H}(Q_{y,H}(p))| \to 0$$

as $\nu \to \infty$. This further implies that

$$\sup_{p\in[\alpha,\beta]} |p - \tilde{F}_{y,H}(Q_{y,H}(p))| \leq \min\{\alpha/2, (1-\beta)/2, \delta\}$$

for all sufficiently large ν. Therefore,

$$\alpha/2 \leq \tilde{F}_{y,H}(Q_{y,H}(p)) \leq (1+\beta)/2 \text{ for all } p \in [\alpha, \beta]$$

and all sufficiently large ν. Hence,

$$\sup_{p\in[\alpha,\beta]} |Q_{y,H}(p) - \tilde{Q}_{y,H}(p)| = \sup_{p\in[\alpha,\beta]} |\tilde{Q}_{y,H}(\tilde{F}_{y,H}(Q_{y,H}(p))) - \tilde{Q}_{y,H}(p)| \leq \epsilon$$

for all sufficiently large ν. This completes the proof of *(i)*. The proof of *(ii)* follows exactly the same way as the proof of *(i)*. □

Next, we state the following lemma, which is required to prove Theorem 4.4.4.

Lemma 4.8.9. *Fix* $0 < \alpha < \beta < 1$. *Suppose that the assumptions of Theorem 4.3.1 hold,* $K(p_1, p_2)$ *is as in* (4.3.1) *in Section 4.4, and* $\hat{K}(p_1, p_2)$ *is as in* (4.4.7) *in Section 4.4.1. Then, the results in* (4.8.6) *of Lemma 4.8.5 hold under stratified multistage cluster sampling design with SRSWOR.*

Proof. The proof follows exactly the same way as the proof of *(i)* in Lemma 4.8.5 for the cases, when H is fixed as $\nu \to \infty$ and $H \to \infty$ as $\nu \to \infty$. □

In the following lemma, we demonstrate some situations, when Assumption 4.2.2–*(i)* holds. Recall from the paragraph preceding Assumption 4.2.1 in Section 4.2 that $Q_y(p)$=$\inf\{t \in \mathbb{R} : F_y(t) \geq p\}$ and $Q_x(p)$=$\inf\{t \in \mathbb{R} : F_x(t) \geq p\}$ are superpopulation p^{th} quantiles of y and x, respectively, and $\mathbf{V}_i$=$\mathbf{R}_i - \sum_{i=1}^{N} \mathbf{R}_i/N$ for i=$1, \ldots, N$, where

$$\mathbf{R}_i = \big(\mathbb{1}_{[Y_i \leq Q_y(p_1)]}, \ldots, \mathbb{1}_{[Y_i \leq Q_y(p_k)]}, \mathbb{1}_{[X_i \leq Q_x(p_1)]} \ldots, \mathbb{1}_{[X_i \leq Q_x(p_k)]}\big)$$

for $p_1, \ldots, p_k \in (0, 1)$ and $k \geq 1$. Then, we state the following lemma.

Lemma 4.8.10. *Suppose that Assumptions 4.2.1, 4.2.4 and 4.2.5 hold. Then, Assumption 4.2.2–(i) holds under SRSWOR and LMS sampling design. Moreover, if* $X_i \leq b$ *a.s.* $[\boldsymbol{P}]$ *for some* $b > 0$, $E_{\boldsymbol{P}}(X_i)^{-1} < \infty$, *Assumption 4.2.1 holds with* $0 < \lambda < E_{\boldsymbol{P}}(X_i)/b$, *and Assumption 4.2.5 holds, then Assumption 4.2.2–(i) holds under any* πPS *sampling design.*

Proof. Given any $k \geq 1$ and $p_1, \ldots, p_k \in (0, 1)$ let us denote $(1/N^2) \sum_{i=1}^{N} (\mathbf{V}_i - \mathbf{T}_V \pi_i)^T (\mathbf{V}_i - \mathbf{T}_V \pi_i)(\pi_i^{-1} - 1)$ by Σ_N. Here, $\mathbf{T}_V$=$\sum_{i=1}^{N} \mathbf{V}_i(1-\pi_i)/\sum_{i=1}^{N} \pi_i(1-\pi_i)$, and the π_i's are inclusion probabilities. Note that

$$n\Sigma_N = (1 - n/N)\bigg(\sum_{i=1}^{N} \mathbf{V}_i^T \mathbf{V}_i/N - \overline{\mathbf{V}}^T \overline{\mathbf{V}}\bigg)$$

under SRSWOR. Then,

$$n\Sigma_N \to (1-\lambda)E_{\mathbf{P}}(\mathbf{R}_i - E_{\mathbf{P}}(\mathbf{R}_i))^T(\mathbf{R}_i - E_{\mathbf{P}}(\mathbf{R}_i)) \text{ as } \nu \to \infty \textit{ a.s. } [\mathbf{P}] \tag{4.8.26}$$

by Assumption 4.2.1 and SLLN. Note that $E_{\mathbf{P}}(\mathbf{R}_i - E_{\mathbf{P}}(\mathbf{R}_i))^T(\mathbf{R}_i - E_{\mathbf{P}}(\mathbf{R}_i))$ is p.d. by Assumption 4.2.5. Thus A4.2.2–*(i)* holds under SRSWOR.

Next, suppose that $\Sigma_N^{(1)}$ and $\Sigma_N^{(2)}$ denote $(1/N^2)\sum_{i=1}^N(\mathbf{V}_i - \mathbf{T}_V\pi_i)^T(\mathbf{V}_i - \mathbf{T}_V\pi_i)(\pi_i^{-1} - 1)$ under LMS sampling design and SRSWOR, respectively, and $\{\pi_i^{(1)}\}_{i=1}^N$ denote inclusion probabilities of LMS sampling design. Then, it follows from the proof of Lemma 2.7.1 in Section 2.7 of Chapter 2 that

$$\max_{1\leq i\leq N} |N\pi_i^{(1)}/n - 1| \to 0 \text{ as } \nu \to \infty \textit{ a.s. } [\mathbf{P}] \tag{4.8.27}$$

It can be shown using this latter result that $n(\Sigma_N^{(1)} - \Sigma_N^{(2)}) \to 0$ as $\nu \to \infty$ *a.s.* $[\mathbf{P}]$. Therefore, Assumption 4.2.2–*(i)* holds under LMS sampling design in view of (4.8.26).

Next, under any πPS sampling design (i.e., a sampling design with π_i=$nX_i/\sum_{i=1}^N X_i$), we have

$$\begin{aligned}\lim_{\nu\to\infty} n\Sigma_N = E_{\mathbf{P}}\big[&\big\{\mathbf{R}_i - E_{\mathbf{P}}(\mathbf{R}_i) + \lambda\chi^{-1}\mu_x^{-1}C_{xr}X_i\big\}^T \times \\ &\big\{\mathbf{R}_i - E_{\mathbf{P}}(\mathbf{R}_i) + \lambda\chi^{-1}\mu_x^{-1}C_{xr}X_i\big\}\big\{\mu_x/X_i - \lambda\big\}\big] \textit{ a.s. } [\mathbf{P}]\end{aligned} \tag{4.8.28}$$

by SLLN because $E_{\mathbf{P}}(X_i)^{-1} < \infty$ and Assumption 4.2.1 holds. Here, μ_x=$E_{\mathbf{P}}(X_i)$, χ=$\mu_x - \lambda(E_{\mathbf{P}}(X_i)^2/\mu_x)$ and C_{xr}=$E_{\mathbf{P}}[(\mathbf{R}_i - E_{\mathbf{P}}(\mathbf{R}_i))X_i]$. The matrix on the right hand side of (4.8.28) is p.d. because $X_i \leq b$ *a.s.* $[\mathbf{P}]$ for some $b > 0$, Assumption 4.2.5 holds and Assumption 4.2.1 holds with $0 < \lambda < E_{\mathbf{P}}(X_i)/b$. Thus Assumption 4.2.2–*(i)* holds under any πPS sampling design. This completes the proof of the lemma. □

Chapter 5

Regression analysis and related estimators in finite populations

In finite population problems, least square (LS) regression is used in the construction of several estimators (see [35], [19], [24], etc.). Some examples of these estimators are the GREG and the ratio estimators of the finite population mean (see Section 2.1 in Chapter 2). The GREG estimator is often considered for estimating the finite population mean because it turns out to be more efficient than several other estimators of the mean under various sampling designs (see Sections 2.1 and 2.2 in Chapter 2). Least square type regression analysis is also used for studying several estimators under sampling designs, which use the auxiliary information. Some examples of those sampling designs are πPS, LMS and RHC sampling designs (see the introduction).

[56], [37], [23], [81], [82], etc. considered quantile (QR) and robust regression in the context of sample survey. However, asymptotic behavior of the estimators obtained from these regression methods has not been studied in the above-mentioned articles, when the sample observations are drawn from a finite population using some sampling design. For i.i.d. sample observations, these estimators were studied in details in the earlier literature (see [46], [39], [50], [51], [59], [33], [21], [49], [42] etc.). It becomes challenging to show Bahadur type representations and asymptotic normality of these estimators, when the sample observations may neither be independent nor identical.

In this chapter, we construct estimators in regression analysis by optimizing convex loss functions. Examples of such estimators include estimators in regression methods like LS, asymmetric least square (ALS), truncated least square (TLS), least absolute deviation (LAD), QR or asymmetric least absolute deviation, etc. Bahadur type representations of these estimators are shown under a probability distribution generated by a sampling design and a superpopulation model. Asymptotic distributions of the above-mentioned estimators are then derived using these Bahadur type representations.

QR and TLS regression are used to construct estimators of the finite population mean. Asymptotic results related to regression analysis are applied to check whether a subset of the auxiliary variables has any influence on the study variable. Moreover, QR and ALS regression are used for detecting the heteroscedasticity present in the finite population observations.

Large sample comparisons of different estimators are carried out based on their asymptotic distributions. From these comparisons, we observe that HEπPS (see the introduction) and RHC sampling designs, which use the auxiliary information, sometimes may have an adverse effect on the performances of different estimators in regression analysis as well as different regression estimators of the finite population mean. We also observe that the estimators of the finite population mean constructed based on QR and TLS regression become more efficient than the GREG estimator under several sampling designs, whenever superpopulations satisfying linear models are considered, and errors in the linear models are generated from symmetric heavy-tailed superpopulation distributions (e.g., Laplace, Student's t, etc.).

In Section 5.1, estimators in regression analysis are constructed. Various asymptotic properties of these estimators are studied in Section 5.2. Covariance estimation for estimators in regression analysis is discussed in Section 5.3. Different applications of regression analysis in finite populations are discussed in Sections 5.4, 5.5 and 5.6. We make some remarks on our major findings in Section 5.7. The proofs of several results are given in Sections 5.8 and 5.9.

5.1. Regression analysis by minimizing loss functions in finite population

Suppose that y is a real-valued study variable and z is a $\mathbb{R}^d$-valued ($d \geq 1$) covariate. Recall from the introduction that (Y_i, Z_i, X_i) is the value of (y, z, x) for the i^{th} population unit, where i=$1, \ldots, N$, and x is a positive real-valued size variable. Also, recall from the introduction that the population total of z and the population values of x are assumed to be known. Moreover, z is used to construct estimators, and x is used to implement sampling designs as well as to construct estimators. As in the earlier chapters, here also we consider all vectors in Euclidean spaces as row vectors and use superscript T to denote their transpose.

Suppose that W_i=(Z_i, X_i) for i=$1, \ldots, N$ and $\rho : \mathbb{R} \to \mathbb{R}$ is a strictly convex function. Then, we define an estimator in regression analysis under a sampling design $P(s)$ as

$$\hat{\boldsymbol{\theta}}_n = (\hat{\alpha}_n, \hat{\beta}_n) = \arg \min_{(\alpha,\beta)\in\mathbb{R}^{d+2}} \sum_{i\in s} d(i,s)\rho(Y_i - \alpha - \beta W_i^T), \tag{5.1.1}$$

where $\{d(i,s) : i \in s\}$ are sampling design weights for the sampling design $P(s)$. Note that in the case of z=x, we take W_i=Z_i=X_i for i=$1, \ldots, N$. There is a unique solution to the minimization problem mentioned in (5.1.1) for any given $s \in \mathcal{S}$ *almost surely*, when ρ is strictly convex, and the population values $\{(Y_i, W_i) : 1 \leq i \leq N\}$ is a sample from some absolutely continuous distribution. Some examples of $\hat{\boldsymbol{\theta}}_n$ are given in Table 5.1 below. We consider $d(i,s)$=π_i^{-1} under

TABLE 5.1: Examples of $\hat{\boldsymbol{\theta}}_n$.

Regression procedure	$\rho(t)$
LS regression	t^2
ALS regression	$\lvert p - \mathbb{1}_{[t<0]}\rvert t^2$ for any fixed $p \in (0,1)$
TLS regression	$t^2\mathbb{1}_{[\lvert t\rvert\leq K]}/2 + K\big(\lvert t\rvert - K/2\big)\mathbb{1}_{[\lvert t\rvert>K]}$ for any fixed $K > 0$
LAD regression	$\lvert t\rvert$
QR	$\lvert t\rvert + (2p-1)t$ for any fixed $p \in (0,1)$

high entropy sampling designs and $d(i,s)$=$G_i X_i^{-1}$ under RHC sampling design. Here, $\{\pi_i\}_{i=1}^N$ are inclusion probabilities of high entropy sampling designs, and G_i is the x total of that group of population units formed in the first step of the RHC sampling design from which the i^{th} population unit is selected in the sample (see the beginning of Section 2.1 in Chapter 2). It is to be noted that $\hat{\boldsymbol{\theta}}_n$ can be viewed as an estimator of

$$\boldsymbol{\theta}_N = (\alpha_N, \beta_N) = \arg \min_{(\alpha,\beta)\in\mathbb{R}^{d+2}} \sum_{i=1}^{N} \rho(Y_i - \alpha - \beta W_i^T). \tag{5.1.2}$$

This is because $\sum_{i\in s} d(i,s)\rho(Y_i - \alpha - \beta W_i^T)$ is the HT estimator of $\sum_{i=1}^N \rho(Y_i - \alpha - \beta W_i^T)$ for $d(i,s)$=π_i^{-1}, and $\sum_{i\in s} d(i,s)\rho(Y_i - \alpha - \beta W_i^T)$ is the RHC estimator of $\sum_{i=1}^N \rho(Y_i - \alpha - \beta W_i^T)$ for $d(i,s)$=$G_i X_i^{-1}$.

5.2. Asymptotic behavior of estimators in regression analysis

In this section, we shall study the asymptotic behavior of $\hat{\boldsymbol{\theta}}_n$ for a general ρ under RHC and any high entropy sampling designs. In Chapter 2, we have derived the asymptotic distribution of $\hat{\beta}_n$ for $\rho(t)$=t^2 under RHC and several high entropy sampling designs in the case of z=x. We consider the asymptotic framework discussed in the earlier chapters. That is, we assume that $\{\mathcal{P}_\nu\}$ is a sequence of populations with $N_\nu, n_\nu \rightarrow \infty$ as $\nu \rightarrow \infty$, where N_ν and n_ν are, respectively, the population and the sample sizes corresponding to the ν^{th} population. As in the preceding chapters, here also we suppress the limiting index ν for the sake of notational simplicity. Moreover, we consider the following assumption mentioned in the earlier chapters (see Assumption 2.1.1 in Chapter 2, Assumption 3.2.1 in Chapter 3 and Assumption 4.2.1 in Chapter 4).

Assumption 5.2.1. $n/N \rightarrow \lambda$ *as* $\nu \rightarrow \infty$*, where* $0 \leq \lambda < 1$.

As in Chapters 2–4, we consider a superpopulation model, where $\{(Y_i, W_i) : 1 \leq i \leq N\}$ are i.i.d. random vectors on $(\Omega, \mathcal{F}, \mathbf{P})$ with some absolutely continuous distribution function. Also, as in Section 2.2 of Chapter 2, Section 3.1 of Chapter 3 and Section 4.2 of Chapter 4, we consider the function $P(s, \omega)$ that is defined on $\mathcal{S} \times \Omega$. Recall from these sections that for each $s \in \mathcal{S}$, $P(s, \omega)$ is a random variable on Ω, and for each $\omega \in \Omega$, $P(s, \omega)$ is a probability distribution on $\mathcal{S}$. It is to be noted that $P(s, \omega)$ is a sampling design for each $\omega \in \Omega$. Moreover, as in Section 4.2 of Chapter 4, we consider the probability measure $\mathbf{P}^*(B \times E)$=$\int_E \sum_{s \in B} P(s, \omega) d\mathbf{P}(\omega)$ defined on the product space ($\mathcal{S} \times \Omega$, $\mathcal{A} \times \mathcal{F}$), where $B \in \mathcal{A}$, $E \in \mathcal{F}$ and $B \times E$ is a cylinder subset of $\mathcal{S} \times \Omega$. Here, $\mathcal{A}$ is the power set of $\mathcal{S}$. As in Section 4.2 of Chapter 4, we denote expectations of random quantities with respect to $P(s, \omega)$, $\mathbf{P}$ and $\mathbf{P}^*$ by E, $E_{\mathbf{P}}$ and $E_{\mathbf{P}^*}$, respectively.

Note that ρ has left hand as well as right hand derivatives at all $t \in \mathbb{R}$ because ρ is convex on $\mathbb{R}$. Also, note that ρ is differentiable at all but at most countably many real numbers. Suppose that $\rho^+(t)$ denotes the right hand derivative of ρ at t. Let us also suppose that $\rho'(t)$ denotes the derivative of ρ at t, when ρ is differentiable at t. Then, we define a function $\psi : \mathbb{R} \rightarrow \mathbb{R}$ as follows.

$$\psi(t) = \begin{cases} \rho'(t), \text{ when } \rho \text{ is differentiable at } t, \\ \rho^+(t), \text{ otherwise.} \end{cases} \tag{5.2.1}$$

Note that $\psi(t)$=$\rho'(t)$ if ρ is differentiable at all $t \in \mathbb{R}$. One can also consider the left hand derivative of $\rho(t)$, say $\rho^-(t)$, in order to define ψ. Then, the results stated in the following

Theorems will remain the same. Let us also define

$$\boldsymbol{\theta} = \arg \min_{(\alpha,\beta)\in\mathbb{R}^{d+2}} E_{\mathbf{P}}\big(\rho(Y_i - \alpha - \beta W_i^T)\big), \text{ and} \tag{5.2.2}$$

$$\epsilon_i = Y_i - \boldsymbol{\theta}\mathbf{V}_i^T \text{ and } \phi(t, W_i) = E_{\mathbf{P}}\big(\psi(\epsilon_i - t)|W_i\big) \tag{5.2.3}$$

for i=$1, \ldots, N$ and $t \in \mathbb{R}$, where $\mathbf{V}_i$=$(1, W_i)$. Next, we consider the following assumptions on superpopulation distribution **P**.

Assumption 5.2.2. *ρ is such that $E_{\boldsymbol{P}}\big(\psi(\epsilon_i)\big)^4 < \infty$ and $\sup\big\{E_{\boldsymbol{P}}\big(\psi(\epsilon_i - \boldsymbol{u}\boldsymbol{V}_i^T/\sqrt{n} + h) - \psi(\epsilon_i - \boldsymbol{u}\boldsymbol{V}_i^T/\sqrt{n} - h)\big)/h : 0 < h \leq \delta\big\} < \infty$ for any given $\boldsymbol{u} \in \mathbb{R}^{d+2}$ and some $\delta > 0$. Further, $E_{\boldsymbol{P}}\big(\psi(\epsilon_i + h) - \psi(\epsilon_i)\big)^2$=$o(1)$, and $E_{\boldsymbol{P}}\big(\psi(\epsilon_i + h) - \psi(\epsilon_i)\big)^4$=$O(1)$, when $h \to 0$ as $\nu \to \infty$.*

Assumption 5.2.3. *ρ is such that $\phi(t, W_i)$ is differentiable with respect to t, $\phi'(t, W_i)$ is continuous with respect to t and $\sup_{t\in\mathbb{R}} |\phi'(t, W_i)|$ exists for any given $\omega \in \Omega$ and i=$1, \ldots, N$, where $\phi'(t, W_i)$ denotes the derivative of $\phi(t, W_i)$ with respect to t. Moreover, $E_{\boldsymbol{P}}\big(\sup_{t\in\mathbb{R}} |\phi'(t, W_i)|\big)^2 < \infty$.*

Assumption 5.2.4. *The distribution of W_i is supported on a compact set in $\mathbb{R}^{d+1}$ and $E_{\boldsymbol{P}}(Y_i)^4 < \infty$. Moreover, Σ=$E_{\boldsymbol{P}}\big(-\phi'(0, W_i)\boldsymbol{V}_i^T\boldsymbol{V}_i\big)$ is a positive definite (p.d.) matrix.*

Since (Y_i, W_i) has absolutely continuous distribution function, Assumptions 5.2.2, 5.2.3 and 5.2.4 hold for different choices of ρ in Table 5.1 in Section 5.1 under some weak regularity conditions as follows.

(i) For $\rho(t)$=t^2 (LS regression), we have $\psi(t)$=$2t$ and $\phi(t, W_i)$=$2(E_{\mathbf{P}}(\epsilon_i|W_i) - t)$ given any i=$1, \ldots, N$. Thus in this case, Assumptions 5.2.2 and 5.2.3 hold, whenever $E_{\mathbf{P}}(\epsilon_i)^4 < \infty$. Also, the condition that Σ=$E_{\mathbf{P}}\big(-\phi'(0, W_i)\mathbf{V}_i^T\mathbf{V}_i\big)$ is a p.d. matrix, which appears in Assumption 5.2.4, holds trivially in this case.

(ii) For $\rho(t)$=$|p - \mathbb{1}_{[t<0]}|t^2$ (ALS regression), we have $\psi(t)$=$2(1-2p)t\mathbb{1}_{[t<0]} + 2pt$ and $\phi(t, W_i)$= $2(1-2p)E_{\mathbf{P}}\big((\epsilon_i - t)\mathbb{1}_{[\epsilon_i<t]}\big|W_i\big) + 2p\big(E_{\mathbf{P}}(\epsilon_i|W_i) - t\big)$ given any i=$1, \ldots, N$. Then, the assumptions discussed in *(i)* above hold in this case if $E_{\mathbf{P}}(\epsilon_i)^4 < \infty$, $F(t, W_i)$ is differentiable with respect to t and $f(t, W_i)$ is continuous with respect to t for any given $\omega \in \Omega$, and $p + (1 - 2p)F(\boldsymbol{\theta}\mathbf{V}_i^T, W_i) > 0$ *a.s.* [**P**]. Here, $F(t, W_i)$ and $f(t, W_i)$, respectively, denote the conditional distribution and the conditional density functions of Y_i given W_i.

(iii) For $\rho(t)$=$t^2\mathbb{1}_{[|t|\leq K]}/2$+$K\big(|t|-K/2\big)\mathbb{1}_{[|t|>K]}$ (TLS regression), we have $\psi(t)$=$t\mathbb{1}_{[|t|\leq K]}$+ $K\mathbb{1}_{[t>K]} - K\mathbb{1}_{[t<-K]}$ and $\phi(t, W_i)$=$K\big(1 - F(t + \boldsymbol{\theta}\mathbf{V}_i^T + K, W_i)\big) - KF(t + \boldsymbol{\theta}\mathbf{V}_i^T - K, W_i) +$

$\int_{t+\boldsymbol{\theta}\mathbf{V}_i^T-K}^{t+\boldsymbol{\theta}\mathbf{V}_i^T+K}(y-t-\boldsymbol{\theta}\mathbf{V}_i^T)f(y,W_i)dy$ given any i=1, ..., N. Therefore, the assumptions discussed in *(i)* hold in this case, whenever $F(t,W_i)$ is differentiable with respect to t and $f(t,W_i)$ is continuous with respect to t for any given $\omega\in\Omega$, and $F(\boldsymbol{\theta}\mathbf{V}_i^T+K,W_i)-F(\boldsymbol{\theta}\mathbf{V}_i^T-K,W_i)>0$ *a.s.* [**P**].

(iv) For $\rho(t)$=$|t|+(2p-1)t$ (QR), we have $\psi(t)$=$2(p-\mathbb{1}_{[t<0]})$ and $\phi(t,W_i)$=$2\big(p-F(t+\boldsymbol{\theta}\mathbf{V}_i^T,W_i)\big)$ given any i=1, ..., N. Assumption 5.2.2 holds in this case, whenever $E_{\mathbf{P}}\big(\sup_{t\in\mathbb{R}} f(t,W_i)\big)<\infty$. Further, in this case, Assumption 5.2.3 is equivalent to Assumption 5.2.5 below. Moreover, the condition that Σ is p.d. holds if $f(\boldsymbol{\theta}\mathbf{V}_i^T,W_i)>0$ *a.s.* [**P**].

Assumption 5.2.5. *$F(t,W_i)$ is differentiable with respect to t, $f(t,W_i)$ is continuous with respect to t and $\sup_{t\in\mathbb{R}} f(t,W_i)$ exists for any given $\omega\in\Omega$ and i=1, ..., N. Moreover, $E_{\boldsymbol{P}}\big(\sup_{t\in\mathbb{R}} f(t,W_i)\big)^2<\infty$.*

Assumptions 5.2.1–5.2.4 are required to show that the results similar to (3.3) and (3.4) in [51] (see Lemmas 5.9.1 and 5.9.3 in Section 5.9) hold under rejective sampling designs (see [40]). Based on these results, we shall show the Bahadur type representation and the asymptotic normality of $\hat{\boldsymbol{\theta}}_n$ for $d(i,s)$=π_i^{-1} under high entropy sampling designs. Recall from Section 3.2 of Chapter 3 that a sampling design $P(s,\omega)$ is called high entropy sampling design, when

$$D(P||R)=\sum_{s\in\mathcal{S}}P(s,\omega)\log(P(s,\omega)/R(s,\omega))\to 0 \text{ as } \nu\to\infty \textit{ a.s. } [\mathbf{P}] \tag{5.2.4}$$

for some rejective sampling design $R(s,\omega)$ (for the description of the rejective sampling design, see the introduction). Some examples of high entropy sampling designs are SRSWOR, RS sampling design (see the introduction), LMS sampling design (see Lemma 3.6.1 in Section 3.6 of Chapter 3), etc.

Next, suppose that $\mathbf{H}_i$=$\psi(\epsilon_i)\mathbf{V}_i$ for i=1, ..., N. Further, suppose that $\mathbf{T}_H$=$\sum_{i=1}^N\mathbf{H}_i(1-\pi_i)/\sum_{i=1}^N\pi_i(1-\pi_i)$. Then, we consider the following assumption.

Assumption 5.2.6. *The inclusion probabilities $\{\pi_i\}_{i=1}^N$ are such that the following hold.*
*(i) There exist constants $K_1,K_2>0$ such that for any i=1, ..., N and all sufficiently large ν, $K_1\le N\pi_i/n\le K_2$ a.s. [**P**].*
*(ii) The matrices $(n/N^2)\sum_{i=1}^N(\boldsymbol{H}_i-\boldsymbol{T}_H\pi_i)^T(\boldsymbol{H}_i-\boldsymbol{T}_H\pi_i)(\pi_i^{-1}-1)\to\Gamma$ as $\nu\to\infty$ a.s. [**P**], where Γ is a p.d. matrix.*

A similar assumption like Assumption 5.2.6 is stated and discussed in Chapter 4 (see the discussion related to Assumption 4.2.2 in Section 4.2 of Chapter 4). It can be shown that Assumption 5.2.6-*(i)* holds under SRSWOR, LMS and any πPS sampling designs (see Lemma 3.6.1 in Chapter 3). It can also be shown using SLLN that Assumption 5.2.6-*(ii)* holds under the aforementioned sampling designs (see Lemma 5.9.5 in Section 5.9). Like Assumptions 5.2.1–5.2.4, Assumption 5.2.6 is also required to prove the results stated in Lemmas 5.9.1 and 5.9.3 in Section 5.9. Now, we state the following theorems.

Theorem 5.2.1. *Suppose that Assumptions 5.2.1–5.2.4 hold. Then, under the probability distribution $\boldsymbol{P}^*$, as $\nu \to \infty$,*

$$\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_N = \left[\sum_{i\in s} d(i,s)\psi(\epsilon_i)\boldsymbol{V}_i/N - \sum_{i=1}^{N}\psi(\epsilon_i)\boldsymbol{V}_i/N\right]\Sigma^{-1} + o_p(1/\sqrt{n}) \text{ and} \tag{5.2.5}$$

$$\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} = \left[\sum_{i\in s} d(i,s)\psi(\epsilon_i)\boldsymbol{V}_i/N\right]\Sigma^{-1} + o_p(1/\sqrt{n}) \tag{5.2.6}$$

for any high entropy sampling design satisfying Assumption 5.2.6, and $d(i,s)=\pi_i^{-1}$.

Theorem 5.2.2. *Suppose that Assumptions 5.2.1–5.2.4 hold. Then, under the probability distribution $\boldsymbol{P}^*$, as $\nu \to \infty$,*

$$\sqrt{n}\big(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_N\big) \xrightarrow{\mathcal{L}} N_{d+2}\big(0, \Sigma^{-1}\Gamma\Sigma^{-1}\big) \text{ and} \tag{5.2.7}$$

$$\sqrt{n}\big(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\big) \xrightarrow{\mathcal{L}} N_{d+2}\big(0, \Delta\big) \tag{5.2.8}$$

for any high entropy sampling design satisfying Assumption 5.2.6, and $d(i,s)=\pi_i^{-1}$, where $\Delta = \Sigma^{-1}\Gamma\Sigma^{-1} + \lambda\Sigma^{-1}E_{\boldsymbol{P}}\big(\psi^2(\epsilon_i)\boldsymbol{V}_i^T\boldsymbol{V}_i\big)\Sigma^{-1}$.

Bahadur type representations of $\hat{\boldsymbol{\theta}}_n$ (see Theorem 5.2.1 above) are first shown under rejective sampling designs using the idea of the proof of the result (3.11) in [51]. Then, these results are shown under high entropy sampling designs using the fact that any high entropy sampling design can be approximated by a rejective sampling design in Kullback-Liebler divergence. On the other hand, the asymptotic normality results of $\hat{\boldsymbol{\theta}}_n$ (see Theorem 5.2.2 above) are shown based on the results stated in Theorem 5.2.1 and the existing asymptotic normality results for the HT estimator.

Next, we shall show that asymptotic results similar to Theorems 5.2.1 and 5.2.2 hold under RHC sampling design. Recall from the introduction that in RHC sampling design, $\mathcal{P}$ is first divided randomly into n disjoint groups of sizes $\tilde{N}_1 \cdots, \tilde{N}_n$, respectively, by taking a sample of $\tilde{N}_1$ units from N units with SRSWOR, a sample of $\tilde{N}_2$ units from $N - \tilde{N}_1$ units with SRSWOR and so on. Then, one unit is selected in the sample from each of these groups independently with probability proportional to the size variable x. As in the earlier chapters, here also we consider the following assumption.

Assumption 5.2.7. *For the RHC sampling design,* $\{\tilde{N}_r\}_{r=1}^n$ *are such that*

$$\tilde{N}_r = \begin{cases} N/n, \textit{ for } r = 1, \cdots, n, \textit{ when } N/n \textit{ is an integer}, \\ \lfloor N/n \rfloor, \textit{ for } r = 1, \cdots, k, \textit{ and} \\ \lfloor N/n \rfloor + 1, \textit{ for } r = k+1, \cdots, n, \textit{ when } N/n \textit{ is not an integer}, \end{cases} \tag{5.2.9}$$

where k *is such that* $\sum_{r=1}^n \tilde{N}_r = N$. *Here,* $\lfloor N/n \rfloor$ *is the integer part of* N/n.

We also consider the following assumptions.

Assumption 5.2.8. $\max_{1\leq i\leq N} X_i / \min_{1\leq i\leq N} X_i = O(1)$ *as* $\nu \to \infty$ *a.s.* $[\mathbb{P}]$.

Assumption 5.2.9. *The matrix* $\Gamma^* = E_{\boldsymbol{P}}(X_i) E_{\boldsymbol{P}}\big\{\big(\boldsymbol{H}_i - X_i E_{\boldsymbol{P}}(\boldsymbol{H}_i)/E_{\boldsymbol{P}}(X_i)\big)^T \big(\boldsymbol{H}_i - X_i E_{\boldsymbol{P}}(\boldsymbol{H}_i)/ E_{\boldsymbol{P}}(X_i)\big) X_i^{-1}\big\}$ *is a p.d. matrix.*

Assumption 5.2.8 is stated and discussed in Chapters 2 and 3 (see Assumption 2.1.3 of Chapter 2 and Assumption 3.2.2 of Chapter 3). Similar kind of assumptions as Assumption 5.2.9 are often used in asymptotic analysis (see [50], [51], etc.). Assumptions 5.2.7–5.2.9 are required to show that the results similar to (3.3) and (3.4) in [51] hold under RHC sampling design (see the proof of Theorem 5.2.3 in Section 5.8). As in the case of high entropy sampling designs, here also we shall show the Bahadur type representation and the asymptotic normality of $\hat{\boldsymbol{\theta}}_n$ for $d(i,s) = G_i X_i^{-1}$ under RHC sampling design based on the aforementioned results.

Theorem 5.2.3. *Suppose that Assumptions 5.2.1–5.2.4 and 5.2.7–5.2.9 hold. Then, under the probability distribution* $\boldsymbol{P}^*$, *as* $\nu \to \infty$,

$$\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_N = \left[\sum_{i\in s} d(i,s)\psi(\epsilon_i)\boldsymbol{V}_i/N - \sum_{i=1}^N \psi(\epsilon_i)\boldsymbol{V}_i/N\right]\Sigma^{-1} + o_p(1/\sqrt{n}) \textit{ and} \tag{5.2.10}$$

$$\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} = \left[\sum_{i\in s} d(i,s)\psi(\epsilon_i)\boldsymbol{V}_i/N\right]\Sigma^{-1} + o_p(1/\sqrt{n}) \tag{5.2.11}$$

for RHC sampling design, and $d(i,s)$=$G_i X_i^{-1}$.

Theorem 5.2.4. *Suppose that Assumptions 5.2.1–5.2.4 and 5.2.7–5.2.9 hold. Then, under the probability distribution* $\boldsymbol{P}^*$*, as* $\nu \to \infty$,

$$\sqrt{n}\big(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_N\big) \xrightarrow{\mathcal{L}} N_{d+2}\big(0, c\Sigma^{-1}\Gamma^*\Sigma^{-1}\big) \text{ and} \tag{5.2.12}$$

$$\sqrt{n}\big(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\big) \xrightarrow{\mathcal{L}} N_{d+2}\big(0, \Delta^*\big) \tag{5.2.13}$$

for RHC sampling design, and $d(i,s)$=$G_i X_i^{-1}$*, where* c=$\lim_{\nu\to\infty} n\gamma$, γ=$\sum_{r=1}^{n} \tilde{N}_r(\tilde{N}_r-1)/N(N-1)$ *and* Δ^*=$c\Sigma^{-1}\Gamma^*\Sigma^{-1} + \lambda\Sigma^{-1}E_{\boldsymbol{P}}\big(\psi^2(\epsilon_i)\boldsymbol{V}_i^T\boldsymbol{V}_i\big)\Sigma^{-1}$.

The proof techniques of Theorems 5.2.3 and 5.2.4 are similar to the proof techniques of Theorems 5.2.3 and 5.2.4, respectively. It follows from Lemma 2.7.5 in Section 2.7 of Chapter 2 that c=1 for λ=0, c=$1-\lambda$ for λ^{-1} an integer, and c=$\lambda\lfloor\lambda^{-1}\rfloor(2-\lambda\lfloor\lambda^{-1}\rfloor-\lambda)$ when λ^{-1} is a non-integer.

5.2.1 Comparison of $\hat{\boldsymbol{\theta}}_n$ under different sampling designs

In this section, we shall first compare the performance of the estimator $\hat{\boldsymbol{\theta}}_n$ for a general ρ under SRSWOR, LMS, RHC and any HEπPS sampling designs in terms of asymptotic total variances (traces of asymptotic covariance matrices) of $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_N)$ under these sampling designs. Recall from the introduction that a sampling design is called HEπPS sampling design if it is a high entropy as well as a πPS sampling design (e.g., RS sampling). We shall carry out the above-mentioned comparison under superpopulations satisfying the linear model

$$Y_i = \boldsymbol{\theta}\mathbf{V}_i^T + \epsilon_i \text{ with } E_{\mathbf{P}}(\psi(\epsilon_i)) = 0 \text{ and } E_{\mathbf{P}}\big(\psi(\epsilon_i)\big)^2 > 0 \tag{5.2.14}$$

for i=$1,\cdots,N$, where $\mathbf{V}_i$=$(1,W_i)$, and $\{\epsilon_i\}_{i=1}^N$ are independent of $\{W_i\}_{i=1}^N$.

Theorem 5.2.5. *Suppose that* $X_i \le b$ *a.s.* $[\boldsymbol{P}]$ *for some* $b>0$, $E_{\boldsymbol{P}}(X_i)^{-2} < \infty$*, Assumption 5.2.1 holds with* $0 \le \lambda < E_{\boldsymbol{P}}(X_i)/b$*, and Assumptions 5.2.2–5.2.4 and 5.2.7–5.2.9 hold. Then, the asymptotic total variance of* $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_N)$ *under SRSWOR is the same as that of* $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_N)$ *under LMS sampling design. Further, the asymptotic total variance of* $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_N)$ *under SRSWOR is smaller than the asymptotic total variances of* $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_N)$ *under RHC and any*

HEπPS sampling designs (which use auxiliary information) if and only if

$$tr\left[\left(E_{\boldsymbol{P}}(\boldsymbol{V}_i^T\boldsymbol{V}_i)\right)^{-1}E_{\boldsymbol{P}}\left((\mu_x X_i^{-1}-1)\boldsymbol{V}_i^T\boldsymbol{V}_i\right)\left(E_{\boldsymbol{P}}(\boldsymbol{V}_i^T\boldsymbol{V}_i)\right)^{-1}\right]>0, \tag{5.2.15}$$

where tr denotes the trace, and μ_x=$E_{\boldsymbol{P}}(X_i)$.

The conditions that $X_i \leq b$ *a.s.* [$\mathbf{P}$] for some $b>0$, and $0<\lambda<E_{\mathbf{P}}(X_i)/b$ are discussed in Chapter 2 (see the discussion related to Assumption 2.2.1 in Chapter 2). The condition in (5.2.15) is an algebraic necessary and sufficient condition. This condition depends neither on the choice of ρ nor on the superpopulation distribution of $\epsilon_i's$. This condition involves superpopulation moments. In practice one can check the above-mentioned condition based on a pilot survey by estimating these superpopulation moments. However, in pilot surveys, the sample size sometimes may not be large enough to reliably estimate these superpopulation moments. Using (5.2.15), several statistical agencies and social-science pollsters can improve the sampling design of recurrently performed surveys. In Table 5.2 below, we consider some cases where this condition holds, and some cases where this condition does not hold. Theorem 5.2.5 implies that the use of the auxiliary information in the design stage may have an adverse effect on the performance of $\hat{\boldsymbol{\theta}}_n$.

TABLE 5.2: Discussion of the condition in (5.2.15).

w=(z,x)	Superpopulation distributions of W_i's	The condition in (5.2.15)
w=z=x	X_i's have log-normal distribution	holds for any parameter values
	X_i's have Pareto distribution with shape α and scale σ	fails to hold for $3\leq\alpha\leq 6$ & σ=1
$z\neq x$	X_i's have Pareto distribution with shape α and scale σ, and Z_i=log(X_i)	holds for $6\leq\alpha<10$ & σ=1
		fails to hold for $2<\alpha<5$ & σ=1

Now we try to demonstrate the result stated in Theorem 5.2.5 using synthetic data. For this, we choose N=5000 and consider the population values $\{(Y_i,X_i):1\leq i\leq N\}$ generated from the linear model Y_i=$1000+X_i+\epsilon_i$ for i=$1,\ldots,N$. Here, X_i's and ϵ_i's are independently generated from the standard log-normal and the standard normal distributions, respectively. Note that in this case, we have W_i=Z_i=X_i for any given i. We also consider the population values $\{(Y_i,W_i):1\leq i\leq N\}$ generated from the linear model Y_i=$1000+Z_i+X_i+\epsilon_i$ for i=$1,\ldots,N$. Here, we generate X_i's from the Pareto distribution with shape=3 and scale=1, and choose

Z_i=log(X_i) for i=$1,\ldots,N$. Then, we generate ϵ_i's independently of the X_i's from the standard normal distribution.

From each of the above data sets, we draw I=1000 samples each of size n=100 using SRSWOR, LMS, RS and RHC sampling designs. Based on these samples, we compare the performance of $\hat{\boldsymbol{\theta}}_n$ under the aforementioned sampling designs in terms of relative efficiencies. We carry out this comparison for each of LS, TLS and LAD regression techniques in the cases of both the data sets. We consider RS sampling design since it is a HEπPS sampling design, and it is easier to implement than other HEπPS sampling designs. Suppose that $P_1(s)$ and $P_2(s)$ denote any two sampling designs. Then, the relative efficiency of $\hat{\boldsymbol{\theta}}_n$ under $P_1(s)$ compared to $\hat{\boldsymbol{\theta}}_n$ under $P_2(s)$ is defined as

$$RE(\hat{\boldsymbol{\theta}}_n, P_1|\hat{\boldsymbol{\theta}}_n, P_2) = MSE(\hat{\boldsymbol{\theta}}_n, P_2)/MSE(\hat{\boldsymbol{\theta}}_n, P_1),$$

where $MSE(\hat{\boldsymbol{\theta}}_n, P)$=$I^{-1}\sum_{l=1}^{I}||\hat{\boldsymbol{\theta}}_{n,l} - \boldsymbol{\theta}_N||^2$ is the MSE of $\hat{\boldsymbol{\theta}}_n$ under any sampling design $P(s)$. Here, $\hat{\boldsymbol{\theta}}_{n,l}$ is an estimate of $\boldsymbol{\theta}_N$ based on the l^{th} sample, l=$1,\ldots,I$. We say that $\hat{\boldsymbol{\theta}}_n$ under $P_1(s)$ is more efficient than under $P_2(s)$ if $RE(\hat{\boldsymbol{\theta}}_n, P_1|\hat{\boldsymbol{\theta}}_n, P_1) > 1$. We use the R software for drawing samples as well as computing estimators. The conclusions drawn from the above data analysis are summarized as follows.

(i) For each of LS, TLS and LAD regression methods, $\hat{\boldsymbol{\theta}}_n$ has lower MSE under SRSWOR than under LMS, RS and RHC sampling designs (see Table 5.3 below) in the case of the first data set.

(ii) In the case of the second data set, $\hat{\boldsymbol{\theta}}_n$ has lower MSE under RS sampling design than under SRSWOR, LMS, and RHC sampling designs (see Table 5.4 below) for each of the above regression techniques.

(iii) The condition in (5.2.15) holds for the linear model Y_i=$1000 + X_i + \epsilon_i$, whereas it fails to hold for the linear model Y_i=$1000 + Z_i + X_i + \epsilon_i$ (see Table 5.2 above). Thus the above empirical results corroborate the theoretical result stated in Theorem 5.2.5.

Next, we try to demonstrate the result stated in Theorem 5.2.5 using real data. For this, as in Section 3.3.2 of Chapter 3, here also we consider Electricity Customer Behaviour Trial data available in Irish Social Science Data Archive (ISSDA, *https://www.ucd.ie/issda/*). Recall from Section 3.3.2 that in this data set, we have electricity consumption of Irish households from 14^{th} July in 2009 to 31^{st} December in 2010. Electricity consumption of these households were

TABLE 5.3: Relative efficiencies of $\hat{\boldsymbol{\theta}}_n$ for the synthetic data set generated from the linear model Y_i=1000 + $X_i + \epsilon_i$. Here, X_i's and ϵ_i's are independently generated from the standard log-normal and the standard normal distributions, respectively.

Regression technique	Relative efficiency	
LS	RE($\hat{\boldsymbol{\theta}}_n, SRSWOR \mid \hat{\boldsymbol{\theta}}_n, LMS$)	1.054569
	RE($\hat{\boldsymbol{\theta}}_n, SRSWOR \mid \hat{\boldsymbol{\theta}}_n, RS$)	2.844394
	RE($\hat{\boldsymbol{\theta}}_n, SRSWOR \mid \hat{\boldsymbol{\theta}}_n, RHC$)	2.897122
LAD	RE($\hat{\boldsymbol{\theta}}_n, SRSWOR \mid \hat{\boldsymbol{\theta}}_n, LMS$)	1.096166
	RE($\hat{\boldsymbol{\theta}}_n, SRSWOR \mid \hat{\boldsymbol{\theta}}_n, RS$)	2.844734
	RE($\hat{\boldsymbol{\theta}}_n, SRSWOR \mid \hat{\boldsymbol{\theta}}_n, RHC$)	3.028323
TLS	RE($\hat{\boldsymbol{\theta}}_n, SRSWOR \mid \hat{\boldsymbol{\theta}}_n, LMS$)	1.106733
	RE($\hat{\boldsymbol{\theta}}_n, SRSWOR \mid \hat{\boldsymbol{\theta}}_n, RS$)	1.356747
	RE($\hat{\boldsymbol{\theta}}_n, SRSWOR \mid \hat{\boldsymbol{\theta}}_n, RHC$)	1.65992

TABLE 5.4: Relative efficiencies of $\hat{\boldsymbol{\theta}}_n$ for the synthetic data set generated from the linear model Y_i=1000 + $Z_i + X_i + \epsilon_i$. , Here, X_i's are generated from the Pareto distribution with shape=3 and scale=1, and Z_i=log(X_i). ϵ_i's are generated from the standard normal distribution independent of the X_i's.

Regression technique	Relative efficiency	
LS	RE($\hat{\boldsymbol{\theta}}_n, RS \mid \hat{\boldsymbol{\theta}}_n, LMS$)	3.972501
	RE($\hat{\boldsymbol{\theta}}_n, RS \mid \hat{\boldsymbol{\theta}}_n, SRSWOR$)	3.697424
	RE($\hat{\boldsymbol{\theta}}_n, RS \mid \hat{\boldsymbol{\theta}}_n, RHC$)	1.015652
LAD	RE($\hat{\boldsymbol{\theta}}_n, RS \mid \hat{\boldsymbol{\theta}}_n, LMS$)	3.888212
	RE($\hat{\boldsymbol{\theta}}_n, RS \mid \hat{\boldsymbol{\theta}}_n, SRSWOR$)	4.094494
	RE($\hat{\boldsymbol{\theta}}_n, RS \mid \hat{\boldsymbol{\theta}}_n, RHC$)	1.148761
TLS	RE($\hat{\boldsymbol{\theta}}_n, RS \mid \hat{\boldsymbol{\theta}}_n, LMS$)	3.751654
	RE($\hat{\boldsymbol{\theta}}_n, RS \mid \hat{\boldsymbol{\theta}}_n, SRSWOR$)	4.789821
	RE($\hat{\boldsymbol{\theta}}_n, RS \mid \hat{\boldsymbol{\theta}}_n, RHC$)	1.125117

measured (in kWh) at the end of every half an hour during the entire time period mentioned above. We choose the mean electricity consumption in December of 2010 as the study variable y, and the mean electricity consumption in December of 2009 as both the covariate z and the size variable x. We have N=5092 households for which electricity consumption data are available during December of both 2009 and 2010. The scatter plot in Figure 5.1 below shows that y is approximately linearly related to x in this data set. Based on this data, we compare the performance of $\hat{\boldsymbol{\theta}}_n$ under SRSWOR, LMS, RS and RHC sampling designs in the same way as in the case

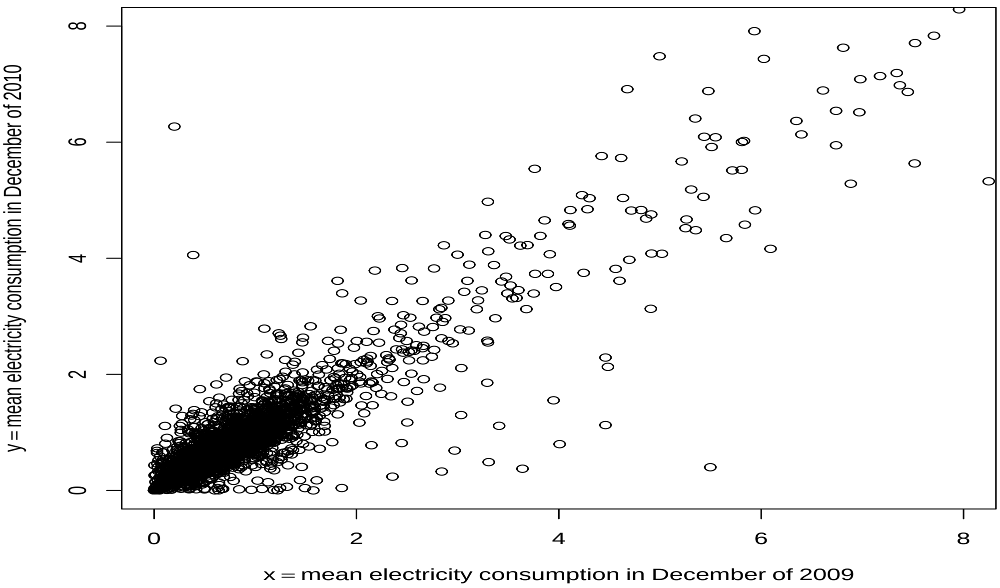

FIGURE 5.1: Scatter plot between y and x for the real data set consisting of mean electricity consumption in December of 2009 and 2010.

of synthetic data. We also approximate the superpopulation moments in (5.2.15) by their corresponding finite population moments based on all the population values in the above data set, and compute C_1=$tr\big[(\sum_{i=1}^N V_i^T V_i/N)^{-1}\big(\sum_{i=1}^N V_i^T V_i(\overline{X}X_i^{-1}-1)/N\big)(\sum_{i=1}^N V_i^T V_i/N)^{-1}\big]$. From this analysis, we observe that $C_1 > 0$. Further, for each of LS, TLS and LAD regression methods, $\hat{\boldsymbol{\theta}}_n$ has lower MSE under SRSWOR than under LMS, RS and RHC sampling designs (see Table 5.5 below). Thus the above empirical results are consistent with the asymptotic result stated in Theorem 5.2.5.

5.3. Covariance estimation for estimators in regression analysis

It follows from Theorem 5.2.2 that under $\mathbf{P}^*$, as $\nu \to \infty$,

$$\sqrt{n}\big(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_N\big) \xrightarrow{\mathcal{L}} N_{d+2}\big(0, \Sigma^{-1}\Gamma\Sigma^{-1}\big) \text{ and } \sqrt{n}\big(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\big) \xrightarrow{\mathcal{L}} N_{d+2}\big(0, \Delta\big) \tag{5.3.1}$$

for $d(i,s)$=π_i^{-1}, and any high entropy sampling design satisfying Assumption 5.2.6. Here, Γ_1=$\Sigma^{-1}\Gamma\Sigma^{-1}$, Σ is as in Assumption 5.2.4, and Γ is as in Assumption 5.2.6–(ii). Further, we

TABLE 5.5: Relative efficiencies of $\hat{\boldsymbol{\theta}}_n$ for the real data set consisting of mean electricity consumption in December of 2009 and 2010.

Regression technique	Relative efficiency	December in 2010
LS	RE($\hat{\boldsymbol{\theta}}_n, SRSWOR \mid \hat{\boldsymbol{\theta}}_n, LMS$)	1.025963
	RE($\hat{\boldsymbol{\theta}}_n, SRSWOR \mid \hat{\boldsymbol{\theta}}_n, RS$)	1.401591
	RE($\hat{\boldsymbol{\theta}}_n, SRSWOR \mid \hat{\boldsymbol{\theta}}_n, RHC$)	6.972742
LAD	RE($\hat{\boldsymbol{\theta}}_n, SRSWOR \mid \hat{\boldsymbol{\theta}}_n, LMS$)	1.507617
	RE($\hat{\boldsymbol{\theta}}_n, SRSWOR \mid \hat{\boldsymbol{\theta}}_n, RS$)	6.439307
	RE($\hat{\boldsymbol{\theta}}_n, SRSWOR \mid \hat{\boldsymbol{\theta}}_n, RHC$)	2.245872
TLS	RE($\hat{\boldsymbol{\theta}}_n, SRSWOR \mid \hat{\boldsymbol{\theta}}_n, LMS$)	1.024037
	RE($\hat{\boldsymbol{\theta}}_n, SRSWOR \mid \hat{\boldsymbol{\theta}}_n, RS$)	5.860129
	RE($\hat{\boldsymbol{\theta}}_n, SRSWOR \mid \hat{\boldsymbol{\theta}}_n, RHC$)	5.303686

have

$$\Delta = \Sigma^{-1}\Gamma\Sigma^{-1} + \lambda\Sigma^{-1}E_{\mathbf{P}}\big(\psi^2(\epsilon_i)\mathbf{V}_i^T\mathbf{V}_i\big)\Sigma^{-1} \text{ and}$$

Here, $\mathbf{V}_i$=$(1, W_i)$ and ϵ_i=Y_i−$\boldsymbol{\theta}\mathbf{V}_i^T$. Recall from Assumption 5.2.4 that Σ=$E_{\mathbf{P}}\big(-\phi'(0, W_i)\mathbf{V}_i^T\mathbf{V}_i\big)$, where $\phi'(0, W_i)$=$\partial\phi(t, W_i)/\partial t|_{t=0}$ for $\phi(t, W_i)$=$E_{\mathbf{P}}\big(\psi(\epsilon_i - t)|W_i\big)$. We estimate $\phi(t, W_i)$ under any high entropy sampling design by

$$\hat{\phi}_1(t, W_i) = \sum_{j\in s}\pi_j^{-1}\prod_{k=1}^{d+1}K_h\big(W_{ik} - W_{jk}\big)\int_{\mathbb{R}}\psi(y_1 - \hat{\boldsymbol{\theta}}_n\mathbf{V}_i^T - t)\times$$
$$K_h(y_1 - Y_j)dy_1\bigg/\sum_{j\in s}\pi_j^{-1}\prod_{k=1}^{d+1}K_h\big(W_{ik} - W_{jk}\big) \tag{5.3.2}$$

for any given i=$1, \ldots, N$, where W_{ik} and W_{jk} are k^{th} components of W_i and W_j, respectively, $K_h(t)$=$K(t/h)/h$, $K(t)$ is a bounded continuous density function, and $h > 0$ is the smoothing parameter. Here, $\hat{\boldsymbol{\theta}}_n$ is as defined in (5.1.1) in Section 5.1 for $d(i, s)$=π_i^{-1}. Note that $\hat{\phi}_1(t, W_i)$ is a Nadaraya-Watson type estimator of the conditional mean $E_{\mathbf{P}}\big(\psi(\epsilon_i - t)|W_i\big)$. Now, if we assume that $\int_{\mathbb{R}}\psi(hy_1 - t)K(y_1)dy_1$ is differentiable with respect to t, then an estimator of $\phi'(0, W_i)$ can be obtained as

$$\hat{\phi}_1'(0, W_i) = \partial\hat{\phi}_1(t, W_i)/\partial t|_{t=0} = \sum_{j\in s}\pi_j^{-1}\prod_{k=1}^{d+1}K_h\big(W_{ik} - W_{jk}\big)\times \tag{5.3.3}$$

$$\left\{\partial\left(\int_{\mathbb{R}}\psi(y_1-\hat{\boldsymbol{\theta}}_n\mathbf{V}_i^T-t)K_h(y_1-Y_j)dy_1\right)\Big/\partial t\Big|_{t=0}\right\}\Big/$$
$$\sum_{j\in s}\pi_j^{-1}\prod_{k=1}^{d+1}K_h\big(W_{ik}-W_{jk}\big).$$

Thus an estimator of Σ under any high entropy sampling design can be constructed by

$$\hat{\Sigma}_1=-\sum_{i\in s}\pi_i^{-1}\hat{\phi}_1'(0,W_i)\mathbf{V}_i^T\mathbf{V}_i/N. \tag{5.3.4}$$

Note that $\hat{\Sigma}_1$ is a HT type estimator of Σ. Also, note that for different choices of ρ in Table 5.1, $\hat{\phi}_1'(0,W_i)$ becomes as in Table 5.6 below. Thus $\hat{\Sigma}_1$ does not depend on the smoothing parameter h and the density function $K(t)$ for ρ as mentioned in 2^{nd} row of Table 5.6, whereas $\hat{\Sigma}_1$ depends on h and $K(t)$ for ρ as mentioned in 3^{rd}, 4^{th} and 5^{th} rows of Table 5.6. Now, following the approach of [16], Γ can be estimated under any high entropy sampling design by

$$\hat{\Gamma}=(n/N^2)\sum_{i\in s}(\hat{\mathbf{H}}_i-\hat{\mathbf{T}}_H\pi_i)^T(\hat{\mathbf{H}}_i-\hat{\mathbf{T}}_H\pi_i)(\pi_i^{-1}-1)\pi_i^{-1}, \tag{5.3.5}$$

where

$$\hat{\mathbf{T}}_H=\sum_{i\in s}\hat{\mathbf{H}}_i(\pi_i^{-1}-1)\Big/\sum_{i\in s}(1-\pi_i),\text{ and}$$
$$\hat{H}_i=\psi(\hat{\epsilon}_i)\mathbf{V}_i\text{ and }\hat{\epsilon}_i=Y_i-\hat{\boldsymbol{\theta}}_n\mathbf{V}_i^T\text{ for any }i\in s.$$

We also estimate $E_{\mathbf{P}}\big(\psi^2(\epsilon_i)\mathbf{V}_i^T\mathbf{V}_i\big)$ in the expression of Δ by $\sum_{i\in s}\pi_i^{-1}\psi^2(\hat{\epsilon}_i)\mathbf{V}_i^T\mathbf{V}_i/N$. There-

TABLE 5.6: Expressions of $\hat{\phi}_1'(0,W_i)$ for different ρ as mentioned in Table 5.1.

$\rho(t)$	$\hat{\phi}_1'(0,W_i)$
t^2	-2
$\lvert p-\mathbb{1}_{[t<0]}\rvert t^2$ for any fixed $p\in(0,1)$	$-2\Big((1-2p)\big(\sum_{j\in s}\pi_j^{-1}\prod_{k=1}^{d+1}K_h\big(W_{ik}-W_{jk}\big)\times \int_{-\infty}^{\hat{\boldsymbol{\theta}}_n\mathbf{V}_i^T}K_h(y_1-Y_j)dy_1\big)/ \sum_{j\in s}\pi_j^{-1}\prod_{k=1}^{d+1}K_h\big(W_{ik}-W_{jk}\big)\Big)-2p$
$t^2\mathbb{1}_{[\lvert t\rvert\leq K]}/2+K\big(\lvert t\rvert-K/2\big)\mathbb{1}_{[\lvert t\rvert>K]}$ for any fixed $K>0$	$-\big(\sum_{j\in s}\pi_j^{-1}\prod_{k=1}^{d+1}K_h\big(W_{ik}-W_{jk}\big)\times \int_{\hat{\boldsymbol{\theta}}_n\mathbf{V}_i^T-K}^{\hat{\boldsymbol{\theta}}_n\mathbf{V}_i^T+K}K_h(y_1-Y_j)dy_1\big)/ \sum_{j\in s}\pi_j^{-1}\prod_{k=1}^{d+1}K_h\big(W_{ik}-W_{jk}\big)$
$\lvert t\rvert+(2p-1)t$ for any fixed $p\in(0,1)$	$-2\big(\sum_{j\in s}\pi_j^{-1}\prod_{k=1}^{d+1}K_h\big(W_{ik}-W_{jk}\big)\times K_h(\hat{\boldsymbol{\theta}}_n\mathbf{V}_i^T-Y_j)\big)/ \sum_{j\in s}\pi_j^{-1}\prod_{k=1}^{d+1}K_h\big(W_{ik}-W_{jk}\big)$

fore, estimators of Γ_1=$\Sigma^{-1}\Gamma\Sigma^{-1}$ and Δ are obtained as

$$\hat{\Gamma}_1 = \hat{\Sigma}_1^{-1}\hat{\Gamma}\hat{\Sigma}_1^{-1} \text{ and } \hat{\Delta} = \hat{\Gamma}_1 + (n/N)\hat{\Sigma}_1^{-1}\Big(\sum_{i\in s}\pi_i^{-1}\psi^2(\hat{\epsilon}_i)\mathbf{V}_i^T\mathbf{V}_i/N\Big)\hat{\Sigma}_1^{-1}. \tag{5.3.6}$$

Next, it follows from Theorem 5.2.4 that under $\mathbf{P}^*$, as $\nu \to \infty$,

$$\sqrt{n}\big(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_N\big) \xrightarrow{\mathcal{L}} N_{d+2}\big(0, c\Sigma^{-1}\Gamma^*\Sigma^{-1}\big) \text{ and } \sqrt{n}\big(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\big) \xrightarrow{\mathcal{L}} N_{d+2}\big(0, \Delta^*\big) \tag{5.3.7}$$

for $d(i,s)$=$G_iX_i^{-1}$, and RHC sampling design. Here, Γ_1^*=$c\Sigma^{-1}\Gamma^*\Sigma^{-1}$, Γ^* is as in Assumption 5.2.9, c=$\lim_{\nu\to\infty} n\gamma$, and γ=$\sum_{r=1}^{n}\tilde{N}_r(\tilde{N}_r-1)/N(N-1)$ with $\tilde{N}_r$ being the size of the r^{th} group formed randomly in RHC sampling design (see the paragraph following Theorem 5.2.2). Further, we have

$$\Delta^* = c\Sigma^{-1}\Gamma^*\Sigma^{-1} + \lambda\Sigma^{-1}E_{\mathbf{P}}\big(\psi^2(\epsilon_i)\mathbf{V}_i^T\mathbf{V}_i\big)\Sigma^{-1}.$$

Under RHC sampling design, we estimate Σ by

$$\hat{\Sigma}_2 = -\sum_{i\in s}(NX_i)^{-1}G_i\hat{\phi}_2'(0,W_i)\mathbf{V}_i^T\mathbf{V}_i, \tag{5.3.8}$$

where $\hat{\phi}_2'(0,W_i)$ is defined in the same way as $\hat{\phi}_1'(0,W_i)$ with π_i^{-1} replaced by $G_iX_i^{-1}$. Note that $\hat{\Sigma}_2$ is a RHC type estimator of Σ. As in the case of $\hat{\Sigma}_1$, $\hat{\Sigma}_2$ does not depend on h and $K(t)$ for ρ as mentioned in 2^{nd} row of Table 5.6, and $\hat{\Sigma}_2$ depends on h and $K(t)$ for ρ as mentioned in 3^{rd}, 4^{th} and 5^{th} rows of Table 5.6. Next, Γ^* can be estimated under RHC sampling design by

$$\hat{\Gamma}^* = (\overline{X}/N)\sum_{i\in s}G_iX_i^{-2}\hat{\mathbf{H}}_i^T\hat{\mathbf{H}}_i - (\hat{\overline{\mathbf{H}}})^T\hat{\overline{\mathbf{H}}}, \tag{5.3.9}$$

where

$$\hat{\overline{\mathbf{H}}} = \sum_{i\in s}(NX_i)^{-1}G_i\hat{\mathbf{H}}_i, \text{ and } \hat{H}_i = \psi(\hat{\epsilon}_i)\mathbf{V}_i \text{ and } \hat{\epsilon}_i = Y_i - \hat{\boldsymbol{\theta}}_n\mathbf{V}_i^T \text{ for any } i \in s.$$

Here, $\hat{\boldsymbol{\theta}}_n$ is as defined in (5.1.1) in Section 5.1 for $d(i,s)$=$G_iX_i^{-1}$. We also estimate $E_{\mathbf{P}}\big(\psi^2(\epsilon_i)\mathbf{V}_i^T\mathbf{V}_i\big)$ in the expression of Δ^* by $\sum_{i\in s}(NX_i)^{-1}G_i\psi^2(\hat{\epsilon}_i)\mathbf{V}_i^T\mathbf{V}_i$. Therefore, estimators of Γ_1^*=$c\Sigma^{-1}\Gamma^*\Sigma^{-1}$ and Δ^* are obtained as

$$\hat{\Gamma}_1^* = n\gamma\hat{\Sigma}_2^{-1}\hat{\Gamma}^*\hat{\Sigma}_2^{-1} \text{ and } \hat{\Delta}^* = \hat{\Gamma}_1^* + (n/N)\times \hat{\Sigma}_2^{-1}\left(\sum_{i\in s}(NX_i)^{-1}G_i\psi^2(\hat{\epsilon}_i)\mathbf{V}_i^T\mathbf{V}_i\right)\hat{\Sigma}_2^{-1}, \tag{5.3.10}$$

We shall now show that $\hat{\Gamma}_1$, $\hat{\Delta}$, $\hat{\Gamma}_1^*$ and $\hat{\Delta}^*$ are consistent estimators of Γ_1, Δ, Γ_1^* and Δ^*, respectively. Let us first consider the following assumptions.

Assumption 5.3.1. $h \to 0$ *and* $nh^{d+1} \to \infty$ *as* $\nu \to \infty$.

Assumption 5.3.2. *The density function* $K(t)$ *is such that* $\int_{\mathbb{R}} K^4(t)dt < \infty$ *and* $\int_{\mathbb{R}} tK(t)dt$=0. *Moreover,* $\partial\big(\int_{\mathbb{R}}\psi(hy_1 - y_2 - t)K(y_1)dy_1\big)/\partial t|_{t=0}$ *is continuous with respect to* y_2.

Assumption 5.3.1 is often considered in the literature for asymptotic analysis. The condition that $\partial\big(\int_{\mathbb{R}}\psi(hy_1 - y_2 - t)K(y_1)dy_1\big)/\partial t|_{t=0}$ is continuous with respect to y_2, which appears in Assumption 5.3.2, holds for different ρ in Table 5.6 because $K(t)$ is a continuous density function. Assumptions 5.3.1 and 5.3.2 are required to show the consistency of the asymptotic covariance matrices of $\hat{\boldsymbol{\theta}}_n$.

Theorem 5.3.1. *(i) Suppose that Assumptions 5.2.1–5.2.4 and 5.2.7–5.3.2 hold. Then, under the probability distribution* $\boldsymbol{P}^*$*, as* $\nu \to \infty$, $\hat{\Gamma}_1 \xrightarrow{p} \Gamma_1$ *for any high entropy sampling design satisfying Assumption 5.2.6, and* $\hat{\Gamma}_1^* \xrightarrow{p} \Gamma_1^*$ *for RHC sampling design.*
(ii) Further, suppose that Assumptions 5.2.1–5.2.4 and 5.2.7–5.3.2 hold. Then, under the probability distribution $\boldsymbol{P}^*$*, as* $\nu \to \infty$, $\hat{\Delta} \xrightarrow{p} \Delta$ *for any high entropy sampling design satisfying Assumption 5.2.6, and* $\hat{\Delta}^* \xrightarrow{p} \Delta^*$ *for RHC sampling design.*

5.4. Regression estimators of the population mean and their comparison

The GREG estimator of the finite population mean $\overline{Y}$=$\sum_{i=1}^N Y_i/N$ can be expressed as $\hat{\overline{Y}}_{GREG}$= $\hat{\boldsymbol{\theta}}_n\overline{\mathbf{V}}^T$, where $\hat{\boldsymbol{\theta}}_n$ is obtained from LS regression, and $\overline{\mathbf{V}}$=$\sum_{i=1}^N \mathbf{V}_i/N$ for $\mathbf{V}_i$=$(1, W_i)$. This motivates us to construct alternative estimators of $\overline{Y}$ based on QR and TLS regression. The estimators obtained from QR and TLS regression depends on p and K (see Table 5.1), respectively, where $p \in (0, 1)$ and $K > 0$. A special case of $\hat{\boldsymbol{\theta}}_n(p)$ is the estimator $\hat{\boldsymbol{\theta}}_n(0.5)$, which is obtained from LAD regression. For convenience, we shall denote these estimators by $\hat{\boldsymbol{\theta}}_n(p)$ and $\hat{\boldsymbol{\theta}}_n(K)$.

The finite population parameter $\boldsymbol{\theta}_N$ in (5.1.2) also depends on p for QR and therefore will be denoted by $\boldsymbol{\theta}_N(p)$. Now, we define

$$\begin{aligned}&\hat{\overline{Y}}_{QR} = \big(\hat{\boldsymbol{\theta}}_n(p_1),\ldots,\hat{\boldsymbol{\theta}}_n(p_l),\hat{\boldsymbol{\theta}}_n(0.5),\hat{\boldsymbol{\theta}}_n(1-p_1),\ldots,\hat{\boldsymbol{\theta}}_n(1-p_l)\big)H_1\overline{\mathbf{V}}^T \text{ and}\\&\hat{\overline{Y}}_{TLS} = \hat{\boldsymbol{\theta}}_n(K)\overline{\mathbf{V}}^T,\end{aligned} \tag{5.4.1}$$

where $l \geq 0, p_1,\ldots,p_l \in (0,0.5)$, H_1=$[m\mathbf{1}_l \boxtimes I_{d+2} \vdots (1-2lm)I_{d+2} \vdots m\mathbf{1}_l \boxtimes I_{d+2}]^T$, m=0 for l=0 and $0 < m < 1/2l$ for $l \geq 1$, $\mathbf{1}_l$ is a $1 \times l$ vector with all the elements equal to 1, and $\boxtimes$ denotes the Kronecker product. Since $\hat{\boldsymbol{\theta}}_n(p)$ is an estimator of $\boldsymbol{\theta}_N(p)$ (see Section 5.1), $\hat{\overline{Y}}_{QR}$ can be viewed as an estimator of $\big(\boldsymbol{\theta}_N(p_1),\ldots,\boldsymbol{\theta}_N(p_l),\boldsymbol{\theta}_N(0.5),\boldsymbol{\theta}_N(1-p_1),\ldots,\boldsymbol{\theta}_N(1-p_l)\big)H_1\overline{\mathbf{V}}^T$. Now, suppose that $\{(Y_i, W_i) : 1 \leq i \leq N\}$ are generated from the linear model

$$Y_i = \boldsymbol{\theta}\mathbf{V}_i^T + \epsilon_i, \tag{5.4.2}$$

where $\{\epsilon\}_{i=1}^N$ are independent of $\{W_i\}_{i=1}^N$ and are generated from some symmetric distribution with $E_{\mathbf{P}}(\epsilon_i)$=0. Then it can be shown that $\big(\boldsymbol{\theta}_N(p_1),\ldots,\boldsymbol{\theta}_N(p_l),\boldsymbol{\theta}_N(0.5),\boldsymbol{\theta}_N(1-p_1),\ldots,\boldsymbol{\theta}_N(1-p_l)\big)H_1\overline{\mathbf{V}}^T$ is close to $\overline{Y}$ for large N. Thus $\hat{\overline{Y}}_{QR}$ can be considered as an estimator of $\overline{Y}$. For a similar reason, $\hat{\overline{Y}}_{TLS}$ can also be considered as an estimator of $\overline{Y}$. Some special cases of $\hat{\overline{Y}}_{QR}$ are

$$\hat{\boldsymbol{\theta}}_n(0.5)\overline{V}^T \text{ and } \big(0.25\hat{\boldsymbol{\theta}}_n(0.25) + 0.5\hat{\boldsymbol{\theta}}_n(0.5) + 0.25\hat{\boldsymbol{\theta}}_n(0.75)\big)\overline{V}^T.$$

For superpopulations satisfying the linear model in (5.4.2), we have shown that the GREG estimator under SRSWOR has the lowest asymptotic variance among the HT, the Hájek, the ratio, the product and the GREG estimators under SRSWOR, LMS, RHC and any HEπPS sampling designs (see Sections 2.1 and 2.2 of Chapter 2). In this section, we shall compare $\hat{\overline{Y}}_{GREG}$, $\hat{\overline{Y}}_{QR}$ and $\hat{\overline{Y}}_{TLS}$ (see Section 5.4) under SRSWOR, LMS, RHC and any HEπPS sampling designs based on the asymptotic distributions of $\sqrt{n}(\hat{\overline{Y}}_{GREG} - E_{\mathbf{P}}(Y_i))$, $\sqrt{n}(\hat{\overline{Y}}_{QR} - E_{\mathbf{P}}(Y_i))$ and $\sqrt{n}(\hat{\overline{Y}}_{TLS} - E_{\mathbf{P}}(Y_i))$ under these sampling designs. We shall carry out the aforementioned comparison under the linear model in (5.4.2). Suppose that ϵ_i's in this linear model have a positive continuous density function f_ϵ. Further, suppose that $l \geq 0$, $p_1,\ldots,p_l \in (0,0.5)$, $(q_1,\ldots,q_{2l+1})$=$(p_1,\ldots,p_l,0.5,1-p_1,\ldots,1-p_l)$, D is a $(2l+1)\times(2l+1)$ matrix such that $((D))_{ij}$=$q_i \wedge q_j - q_iq_j$ for $1 \leq i,j \leq 2l+1$, and

$$\xi = \big(m/f_\epsilon(Q_\epsilon(p_1)), \ldots, m/f_\epsilon(Q_\epsilon(p_l)), (1-2lm)/f_\epsilon(Q_\epsilon(0.5)), \\ m/f_\epsilon(Q_\epsilon(1-p_1)), \ldots, m/f_\epsilon(Q_\epsilon(1-p_l))\big). \tag{5.4.3}$$

where $Q_\epsilon(p)$ is the p^{th} quantile of ϵ_i. Then, we state the following theorem.

Theorem 5.4.1. *Suppose that $X_i \le b$ a.s. $[\boldsymbol{P}]$ for some $b > 0$, $E_{\boldsymbol{P}}(X_i)^{-2} < \infty$, Assumption 5.2.1 holds with $0 \le \lambda < E_{\boldsymbol{P}}(X_i)/b$, and Assumptions 5.2.7 and 5.2.8 hold. Then, under any of SRSWOR, LMS, RHC and any HEπPS sampling designs, the asymptotic variance of $\sqrt{n}(\hat{\overline{Y}}_{QR} - E_{\boldsymbol{P}}(Y_i))$ becomes smaller than the asymptotic variance of $\sqrt{n}(\hat{\overline{Y}}_{GREG} - E_{\boldsymbol{P}}(Y_i))$ if and only if*

$$\sigma_\epsilon^2 > \xi D \xi^T, \tag{5.4.4}$$

the asymptotic variance of $\sqrt{n}(\hat{\overline{Y}}_{TLS} - E_{\boldsymbol{P}}(Y_i))$ becomes smaller than the asymptotic variance of $\sqrt{n}(\hat{\overline{Y}}_{GREG} - E_{\boldsymbol{P}}(Y_i))$ if and only if

$$\sigma_\epsilon^2 > \big(K^2\boldsymbol{P}(|\epsilon_i| > K) + E_{\boldsymbol{P}}(\epsilon_i)^2 \mathbb{1}_{[|\epsilon_i| \le K]}\big)\big/\big(\boldsymbol{P}(|\epsilon_i| \le K)\big)^2, \text{ and} \tag{5.4.5}$$

the asymptotic variance of $\sqrt{n}(\hat{\overline{Y}}_{QR} - E_{\boldsymbol{P}}(Y_i))$ becomes smaller than the asymptotic variance of $\sqrt{n}(\hat{\overline{Y}}_{TLS} - E_{\boldsymbol{P}}(Y_i))$ if and only if

$$\big(K^2\boldsymbol{P}(|\epsilon_i| > K) + E_{\boldsymbol{P}}(\epsilon_i)^2 \mathbb{1}_{[|\epsilon_i| \le K]}\big)\big/\big(\boldsymbol{P}(|\epsilon_i| \le K)\big)^2 > \xi D \xi^T, \tag{5.4.6}$$

where σ_ϵ^2 is the superpopulation variance of ϵ_i's.

The conditions in (5.4.4), (5.4.5) and (5.4.6) are algebraic necessary and sufficient conditions. These conditions involve superpopulation moments, quantiles and density function. In practice, one can check these conditions by estimating the above-mentioned parameters based on a pilot survey. For l=0 and K=1, we consider some cases where these conditions hold, and some cases where these conditions do not hold (see Tables 5.7, 5.8 and 5.9 below). Theorem 5.4.1 shows that $\hat{\overline{Y}}_{QR}$ as well as $\hat{\overline{Y}}_{TLS}$ is more efficient than $\hat{\overline{Y}}_{GREG}$, whenever ϵ_i's are generated from heavy-tailed distributions (e.g., Laplace, Student's t, etc.).

Under the linear model in (5.4.2), it is shown in Chapter 2 that $\hat{\overline{Y}}_{GREG}$ has the same asymptotic distribution around $\overline{Y}$ under SRSWOR and LMS sampling designs. It is also shown in Chapter 2 that RHC and HEπPS sampling designs, which use the auxiliary information, have an

TABLE 5.7: Discussion of the condition in (5.4.4).

Superpopulation distribution of ϵ_i's	The condition in (5.4.4)
Exponential power distribution with location μ=0, scale $\sigma > 0$ and shape $\alpha > 0$	*holds iff $\alpha^2\Gamma(3/\alpha) > \Gamma^3(1/\alpha)$
Student's t-distribution with degrees of freedom (df) $r > 2$	*holds iff $4\Gamma^2((r+1)/2) > (r-2)\pi\Gamma^2(r/2)$

* Here, $\Gamma(\cdot)$ denotes the gamma function.

TABLE 5.8: Discussion of the condition in (5.4.5).

Superpopulation distribution of ϵ_i's	The condition in (5.4.5)
Standard Laplace distribution	holds
Student's t-distribution with df r=3, 4 & 5	holds
Standard normal distribution	does not hold

TABLE 5.9: Discussion of the condition in (5.4.6).

Superpopulation distribution of ϵ_i's	The condition in (5.4.6)
Standard Laplace distribution	holds
Student's t-distribution with df r=3, 4 & 5	does not hold
Standard normal distribution	does not hold

adverse effect on the performance of $\hat{\overline{Y}}_{GREG}$. In the next theorem, we shall show that a similar result holds for $\hat{\overline{Y}}_{QR}$ and $\hat{\overline{Y}}_{TLS}$.

Theorem 5.4.2. *Suppose that the assumptions of Theorem 5.4.1 hold. Then, the asymptotic distribution of each of $\sqrt{n}(\hat{\overline{Y}}_{QR} - E_{\boldsymbol{P}}(Y_i))$ and $\sqrt{n}(\hat{\overline{Y}}_{TLS} - E_{\boldsymbol{P}}(Y_i))$ is the same under SRSWOR and LMS sampling designs. Further, the asymptotic variance of each of $\sqrt{n}(\hat{\overline{Y}}_{QR} - E_{\boldsymbol{P}}(Y_i))$ and $\sqrt{n}(\hat{\overline{Y}}_{TLS} - E_{\boldsymbol{P}}(Y_i))$ under SRSWOR is smaller than its asymptotic variance under RHC as well as any HEπPS sampling design, which uses the auxiliary information.*

Theorem 5.4.2 implies that the use of the auxiliary information in the design stage has an adverse effect on the performance of $\hat{\overline{Y}}_{QR}$ and $\hat{\overline{Y}}_{TLS}$.

As in Section 5.2.1, here also we try to demonstrate the results stated in Theorems 5.4.1 and 5.4.2 using synthetic and real data. For this, we consider z=x, and generate N=5000 population values on (y, x) from the linear model Y_i=1000 + X_i + ϵ_i for i=1, …, N. Here, X_i's are generated from the standard log-normal distribution, and ϵ_i's are generated independently of the

X_i's from the standard normal, the Student's t (with df 3) and the standard Laplace distributions. Based on these data sets, we compare $\hat{\overline{Y}}_{QR}$, $\hat{\overline{Y}}_{TLS}$ and $\hat{\overline{Y}}_{GREG}$ under SRSWOR, LMS, RS and RHC sampling designs in the same way as in Section 5.2.1. We consider $\hat{\overline{Y}}_{QR}$ for l=0, and $\hat{\overline{Y}}_{TLS}$ for K=1. The relative efficiency of an estimator $\hat{\overline{Y}}_1$ of $\overline{Y}$ under a sampling design $P_1(s)$ compared to another estimator $\hat{\overline{Y}}_2$ under another sampling design $P_2(s)$ is defined as

$$RE(\hat{\overline{Y}}_1, P_1|\hat{\overline{Y}}_2, P_2) = MSE(\hat{\overline{Y}}_2, P_2)/MSE(\hat{\overline{Y}}_1, P_1),$$

where $MSE(\hat{\overline{Y}}_k, P_k)$=$I^{-1}\sum_{l=1}^{I}(\hat{\overline{Y}}_{kl} - \overline{Y})^2$ is the MSE of $\hat{\overline{Y}}_k$ under $P_k(s)$ for k=$1, 2$. Here, $\hat{\overline{Y}}_{k1}$ is an estimate of $\overline{Y}$ based on the k^{th} estimator and the l^{th} sample, k=$1, 2$, l=$1, \ldots, I$=1000. The conclusions drawn from the above data analysis are summarized in Table 5.10 below (for further details, see Tables 5.11–5.13 below). We observe that the empirical results stated in Table 5.10 corroborate the theoretical results stated in Theorems 5.4.1 and 5.4.2.

TABLE 5.10: Most efficient regression estimators of $\overline{Y}$ in terms of relative efficiencies.

Superpopulation distribution of ϵ_i's	Most efficient estimators	Conditions in (5.4.4)–(5.4.6)
Standard normal distribution	$\hat{\overline{Y}}_{GREG}$ under SRSWOR	None of these holds
Student's t-distribution with df=3	$\hat{\overline{Y}}_{TLS}$ under SRSWOR	(5.4.4) & (5.4.5) hold but (5.4.6) does not hold
Standard Laplace distribution	$\hat{\overline{Y}}_{QR}$ under SRSWOR	All of these hold

Next, we carry out the above comparison based on the real data set considered in Section 5.2.1. We also approximate the superpopulation parameters in the conditions (5.4.4)–(5.4.6) for l=0 and K=1 based on all the population values in this real data set. Note that for l=0, we have $\xi D \xi^T$=$1/4f_\epsilon^2(0)$. Then, we approximate σ_ϵ^2, $1/4f_\epsilon^2(0)$ and $\big(\mathbf{P}(|\epsilon_i| > 1) + E_{\mathbf{P}}(\epsilon_i)^2 \mathbb{1}_{[|\epsilon_i|\leq 1]}\big)\big/\big(\mathbf{P}(|\epsilon_i| \leq 1)\big)^2$ by

$$C_2 = \sum_{i=1}^{N} e_{i,1}^2/N,\ C_3 = 1\Big/4\bigg(\sum_{i=1}^{N} K(e_{i,2}/h)/Nh\bigg)^2 \text{ and}$$
$$C_4 = \bigg(\sum_{i=1}^{N} \mathbb{1}_{[|e_{i,3}|>1]} + \sum_{i=1}^{N} e_{i,3}^2 \mathbb{1}_{[|e_{i,3}|\leq 1]}\bigg)\Big/N\bigg(\sum_{i=1}^{N} \mathbb{1}_{[|e_{i,3}|\leq 1]}/N\bigg)^2,$$

respectively, where $\{e_{i,1}\}_{i=1}^N$, $\{e_{i,2}\}_{i=1}^N$ and $\{e_{i,3}\}_{i=1}^N$ are the residuals obtained from LS, LAD and TLS regression, respectively, and $\sum_{i=1}^N K(e_{i,2}/h)/Nh$ is the kernel density estimator of $f_\epsilon(0)$. We choose $K(t)$ to be the uniform density function $\mathbb{1}_{[-1,1]}(t)$ and h by means of leave one out cross validation. We compute C_2, C_3 and C_4 based on LS, LAD and TLS regression, respectively, because σ_ϵ^2, $1/4f_\epsilon^2(0)$ and $\big(\mathbf{P}(|\epsilon_i| > 1) + E_{\mathbf{P}}(\epsilon_i)^2 \mathbb{1}_{[|\epsilon_i|\leq 1]}\big)/\big(\mathbf{P}(|\epsilon_i| \leq 1)\big)^2$ are involved in the asymptotic variances of $\hat{\overline{Y}}_{GREG}$, $\hat{\overline{Y}}_{QR}$ (for l=0) and $\hat{\overline{Y}}_{TLS}$ (for K=1), respectively. From the above analysis, we observe that $C_2 > C_4 > C_3$. Moreover, $\hat{\overline{Y}}_{QR}$ under SRSWOR has the lowest MSE among all the estimators and the sampling designs considered here (see Table 5.14 below). Thus the above empirical results are consistent with the asymptotic results stated in Theorems 5.4.1 and 5.4.2.

TABLE 5.11: Relative efficiencies of the regression estimators of $\overline{Y}$ for the synthetic data set generated from the linear model Y_i=1000 + $X_i + \epsilon_i$. Here, ϵ_i's have the standard normal distribution.

RE($\hat{\overline{Y}}_{GREG}, SRSWOR \mid \hat{\overline{Y}}_{TLS}, SRSWOR$)	1.079478
RE($\hat{\overline{Y}}_{GREG}, SRSWOR \mid \hat{\overline{Y}}_{QR}, SRSWOR$)	1.523295
RE($\hat{\overline{Y}}_{GREG}, SRSWOR \mid \hat{\overline{Y}}_{QR}, LMS$)	1.563709
RE($\hat{\overline{Y}}_{GREG}, SRSWOR \mid \hat{\overline{Y}}_{TLS}, LMS$)	1.118407
RE($\hat{\overline{Y}}_{GREG}, SRSWOR \mid \hat{\overline{Y}}_{GREG}, LMS$)	1.011407
RE($\hat{\overline{Y}}_{GREG}, SRSWOR \mid \hat{\overline{Y}}_{QR}, RS$)	4.233067
RE($\hat{\overline{Y}}_{GREG}, SRSWOR \mid \hat{\overline{Y}}_{TLS}, RS$)	2.774588
RE($\hat{\overline{Y}}_{GREG}, SRSWOR \mid \hat{\overline{Y}}_{GREG}, RS$)	2.173338
RE($\hat{\overline{Y}}_{GREG}, SRSWOR \mid \hat{\overline{Y}}_{QR}, RHC$)	4.04144
RE($\hat{\overline{Y}}_{GREG}, SRSWOR \mid \hat{\overline{Y}}_{TLS}, RHC$)	2.550825
RE($\hat{\overline{Y}}_{GREG}, SRSWOR \mid \hat{\overline{Y}}_{GREG}, RHC$)	2.166384

TABLE 5.12: Relative efficiencies of the regression estimators of $\overline{Y}$ for the synthetic data set generated from the linear model Y_i=1000 + $X_i + \epsilon_i$. Here, ϵ_i's have the t distribution with df 3.

RE($\hat{\overline{Y}}_{TLS}, SRSWOR \mid \hat{\overline{Y}}_{QR}, SRSWOR$)	1.14752
RE($\hat{\overline{Y}}_{TLS}, SRSWOR \mid \hat{\overline{Y}}_{GREG}, SRSWOR$)	1.88136
RE($\hat{\overline{Y}}_{TLS}, SRSWOR \mid \hat{\overline{Y}}_{QR}, LMS$)	1.28922
RE($\hat{\overline{Y}}_{TLS}, SRSWOR \mid \hat{\overline{Y}}_{TLS}, LMS$)	1.0916
RE($\hat{\overline{Y}}_{TLS}, SRSWOR \mid \hat{\overline{Y}}_{GREG}, LMS$)	1.924977
RE($\hat{\overline{Y}}_{TLS}, SRSWOR \mid \hat{\overline{Y}}_{QR}, RS$)	3.535446
RE($\hat{\overline{Y}}_{TLS}, SRSWOR \mid \hat{\overline{Y}}_{TLS}, RS$)	2.008073
RE($\hat{\overline{Y}}_{TLS}, SRSWOR \mid \hat{\overline{Y}}_{GREG}, RS$)	5.03639
RE($\hat{\overline{Y}}_{TLS}, SRSWOR \mid \hat{\overline{Y}}_{QR}, RHC$)	3.760415
RE($\hat{\overline{Y}}_{TLS}, SRSWOR \mid \hat{\overline{Y}}_{TLS}, RHC$)	1.973055
RE($\hat{\overline{Y}}_{TLS}, SRSWOR \mid \hat{\overline{Y}}_{GREG}, RHC$)	5.661026

TABLE 5.13: Relative efficiencies of the regression estimators of $\overline{Y}$ for the synthetic data set generated from the linear model Y_i=1000 + X_i + ϵ_i. Here, ϵ_i's have the standard Laplace distribution.

RE($\hat{\overline{Y}}_{QR}, SRSWOR \mid \hat{\overline{Y}}_{TLS}, SRSWOR$)	1.125307
RE($\hat{\overline{Y}}_{QR}, SRSWOR \mid \hat{\overline{Y}}_{GREG}, SRSWOR$)	1.690677
RE($\hat{\overline{Y}}_{QR}, SRSWOR \mid \hat{\overline{Y}}_{QR}, LMS$)	1.013656
RE($\hat{\overline{Y}}_{QR}, SRSWOR \mid \hat{\overline{Y}}_{TLS}, LMS$)	1.153869
RE($\hat{\overline{Y}}_{QR}, SRSWOR \mid \hat{\overline{Y}}_{GREG}, LMS$)	1.738247
RE($\hat{\overline{Y}}_{QR}, SRSWOR \mid \hat{\overline{Y}}_{QR}, RS$)	1.865937
RE($\hat{\overline{Y}}_{QR}, SRSWOR \mid \hat{\overline{Y}}_{TLS}, RS$)	2.9604
RE($\hat{\overline{Y}}_{QR}, SRSWOR \mid \hat{\overline{Y}}_{GREG}, RS$)	3.974535
RE($\hat{\overline{Y}}_{QR}, SRSWOR \mid \hat{\overline{Y}}_{QR}, RHC$)	1.837466
RE($\hat{\overline{Y}}_{QR}, SRSWOR \mid \hat{\overline{Y}}_{TLS}, RHC$)	3.073856
RE($\hat{\overline{Y}}_{QR}, SRSWOR \mid \hat{\overline{Y}}_{GREG}, RHC$)	4.074943

TABLE 5.14: Relative efficiencies of the regression estimators of $\overline{Y}$ for the real data set consisting of mean electricity consumption in December of 2009 and 2010.

Relative efficiency	December in 2010
RE($\hat{\overline{Y}}_{QR}, SRSWOR \mid \hat{\overline{Y}}_{TLS}, SRSWOR$)	1.170082
RE($\hat{\overline{Y}}_{QR}, SRSWOR \mid \hat{\overline{Y}}_{GREG}, SRSWOR$)	1.922412
RE($\hat{\overline{Y}}_{QR}, SRSWOR \mid \hat{\overline{Y}}_{QR}, LMS$)	1.070182
RE($\hat{\overline{Y}}_{QR}, SRSWOR \mid \hat{\overline{Y}}_{TLS}, LMS$)	1.298114
RE($\hat{\overline{Y}}_{QR}, SRSWOR \mid \hat{\overline{Y}}_{GREG}, LMS$)	2.100872
RE($\hat{\overline{Y}}_{QR}, SRSWOR \mid \hat{\overline{Y}}_{QR}, RS$)	2.793544
RE($\hat{\overline{Y}}_{QR}, SRSWOR \mid \hat{\overline{Y}}_{TLS}, RS$)	3.231571
RE($\hat{\overline{Y}}_{QR}, SRSWOR \mid \hat{\overline{Y}}_{GREG}, RS$)	4.081814
RE($\hat{\overline{Y}}_{QR}, SRSWOR \mid \hat{\overline{Y}}_{QR}, RHC$)	2.43444
RE($\hat{\overline{Y}}_{QR}, SRSWOR \mid \hat{\overline{Y}}_{TLS}, RHC$)	3.142127
RE($\hat{\overline{Y}}_{QR}, SRSWOR \mid \hat{\overline{Y}}_{GREG}, RHC$)	3.402416

5.5. Variable selection and related tests in sample survey

As discussed in Section 5.1, in sample survey, the auxiliary variables in w=(z, x) are used to construct estimators and to implement sampling designs. Therefore, it becomes significant to determine the variables in w, which have influence on the study variable y. In this section, we shall discuss a variable selection method based on LS regression under RHC and any high entropy sampling designs. The estimator in LS regression can be expressed as

$$\begin{aligned}
&\hat{\boldsymbol{\theta}}_n = (\hat{\alpha}_n, \hat{\beta}_n) \text{ with } \hat{\alpha}_n = \hat{\overline{Y}} - \hat{\beta}_n \hat{\overline{W}}^T, \hat{\beta}_n = \hat{S}_{wy} \hat{S}_{ww}^{-1},\\
&\hat{\overline{Y}} = \sum_{i \in s} d(i,s) Y_i \Big/ \sum_{i \in s} d(i,s), \hat{\overline{W}} = \sum_{i \in s} d(i,s) W_i \Big/ \sum_{i \in s} d(i,s),\\
&\hat{S}_{ww} = \left(\sum_{i \in s} d(i,s) W_i^T W_i \Big/ \sum_{i \in s} d(i,s) \right) - \hat{\overline{W}}^T \hat{\overline{W}} \text{ and}\\
&\hat{S}_{wy} = \left(\sum_{i \in s} d(i,s) Y_i W_i / \sum_{i \in s} d(i,s) \right) - \hat{\overline{Y}}\, \hat{\overline{W}}.
\end{aligned} \tag{5.5.1}$$

Now, suppose that the population values $\{(Y_i, W_i) : 1 \leq i \leq N\}$ are generated from a superpopulation satisfying the linear model

$$Y_i = \boldsymbol{\theta} \mathbf{V}_i^T + \epsilon_i \text{ with } E_{\mathbf{P}}(\epsilon_i | W_i) = 0, \tag{5.5.2}$$

where $\mathbf{V}_i$=$(1, W_i)$. One can carry out a step-wise selection of variables under high entropy and RHC sampling designs as follows. Suppose that $\boldsymbol{\theta}_{j+1}$ is the $(j+1)^{th}$ component of $\boldsymbol{\theta}$ and w_j is the j^{th} component of w for j=$1, \ldots, d+1$. Then, $H_{0,j} : \boldsymbol{\theta}_{j+1}$=0 is tested against $H_{A,j} : \boldsymbol{\theta}_{j+1} \neq 0$ based on the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}}_{n,j+1} - \boldsymbol{\theta}_{j+1})$ in the first step of the variable selection method for all j=$1, \ldots, d+1$. Here, $\hat{\boldsymbol{\theta}}_{n,j+1}$ is the $(j+1)^{th}$ component of $\hat{\boldsymbol{\theta}}_n$, and $\hat{\boldsymbol{\theta}}_n$ is the estimator obtained from LS regression. If the asymptotic p-value corresponding to the test $H_{0,k} : \boldsymbol{\theta}_{k+1}$=0 is the largest among the asymptotic p-values corresponding to the above-mentioned tests, and it exceeds a certain threshold C (e.g., 0.01 or 0.05), then w_k is dropped from the model. In the second step, the same procedure is followed with all the auxiliary variables except w_k. This step-wise selection of variables is continued until the maximum p-value at any step becomes less than the threshold C. In any given step, a large sample test for the hypothesis $H_{0,j} : \boldsymbol{\theta}_{j+1}$=0 is constructed based on the test statistic

$$\chi_{n,j} = (\sqrt{n} \hat{\boldsymbol{\theta}}_{n,j+1})^2 / \widehat{AV}\big(\sqrt{n}(\hat{\boldsymbol{\theta}}_{n,j+1} - \boldsymbol{\theta}_{j+1})\big), \tag{5.5.3}$$

where $\widehat{AV}\big(\sqrt{n}(\hat{\boldsymbol{\theta}}_{n,j+1} - \boldsymbol{\theta}_{j+1})\big)$ is a consistent estimator of the asymptotic variance of $\sqrt{n}(\hat{\boldsymbol{\theta}}_{n,j+1} - \boldsymbol{\theta}_{j+1})$ for j=$1, \ldots, d+1$. It follows from the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$ under high entropy and RHC sampling designs (see Theorems 5.2.2 and 5.2.4) that under $H_{0,j}$ and these sampling designs, the asymptotic distribution of $\chi_{n,j}$ is central chi-square with df 1 given any j. The variable selection method described above can also be carried out based on LAD regression under the assumption that the conditional distribution of ϵ_i given W_i is symmetric about 0.

We shall now demonstrate the variable selection method described above using synthetic data. For this, we choose N=5000 and consider the population values $\{(Y_i, W_i) : 1 \leq i \leq N\}$ generated from the linear models Y_i=$1000 + Z_i + X_i + \epsilon_i$ and Y_i=$1000 + X_i + \epsilon_i$ for i=$1, \ldots, N$. Here, we independently generate Z_i's and X_i's from the standard normal and the standard log-normal distributions, respectively. Then, we generate ϵ_i's independently of the (Z_i, X_i)'s from the standard normal distribution. From each of these data sets, we draw 1000 samples each of size n=100 using SRSWOR. Based on these samples, we carry out variable selection using LS regression for C=0.05 as discussed in the preceding paragraph. The conclusions drawn from the above data analysis are summarized as follows.

(i) For the data set generated from the first linear model, the variables z and x are always selected.

(ii) For the data set generated from the second linear model, although x is always selected, z is selected only 46 times out of 1000 repetitions.

Next, we consider the mean electricity consumption in the summer months (viz. June, July and August) of 2009 and 2010 from the Electricity Customer Behaviour Trial data (see Section 5.2.1), and demonstrate the variable selection method based on this data set. We choose the mean electricity consumption in the summer months of 2010 as the study variable y, the mean electricity consumption in July of 2009 as the first covariate z_1, and the mean electricity consumption in August of 2009 as the second covariate z_2. We have N=5372 households for which electricity consumption data are available during July and August of 2009 and all the summer months of 2010. Note that we have w=(z_1, z_2) in this case. Scatter plots in Figures 5.2 and 5.3 below show that y is approximately linearly related to each of z_1 and z_2 in this data set. Also, the finite population linear regression coefficient of y on z_1 and that of y on z_2 are 0.282 and 0.665, respectively. We observe that z_1 is selected 650 times and z_2 is selected 840 times out of 1000 times, when we perform the numerical experiment discussed in the preceding paragraph.

5.6. Detection of heteroscedasticity in finite populations

The presence of heteroscedasticity has an important influence on the performance of different estimators in sample survey. For instance, under superpopulations satisfying heteroscedastic linear

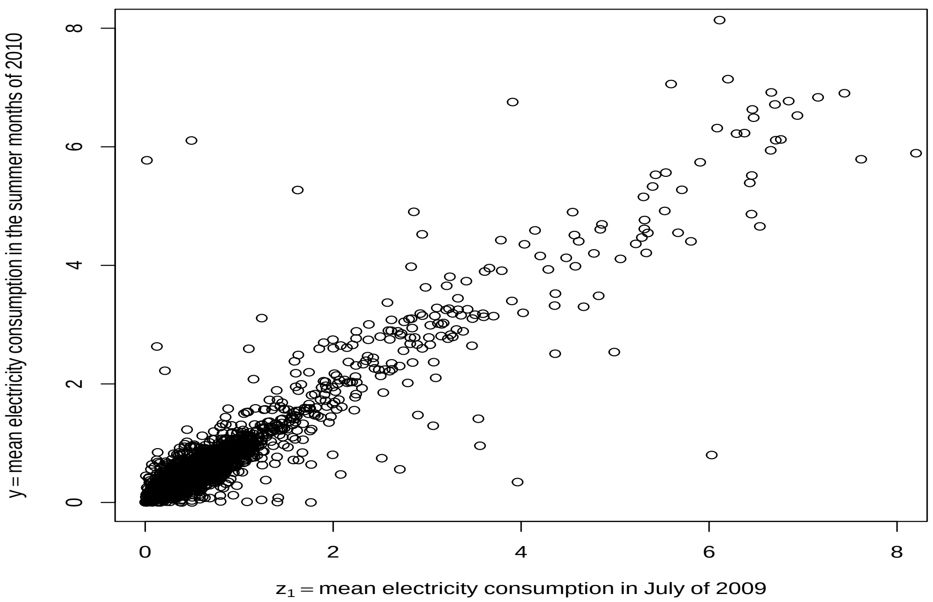

FIGURE 5.2: Scatter plot between y and z_1 for the real data set consisting of mean electricity consumption in the summer months of 2009 and 2010.

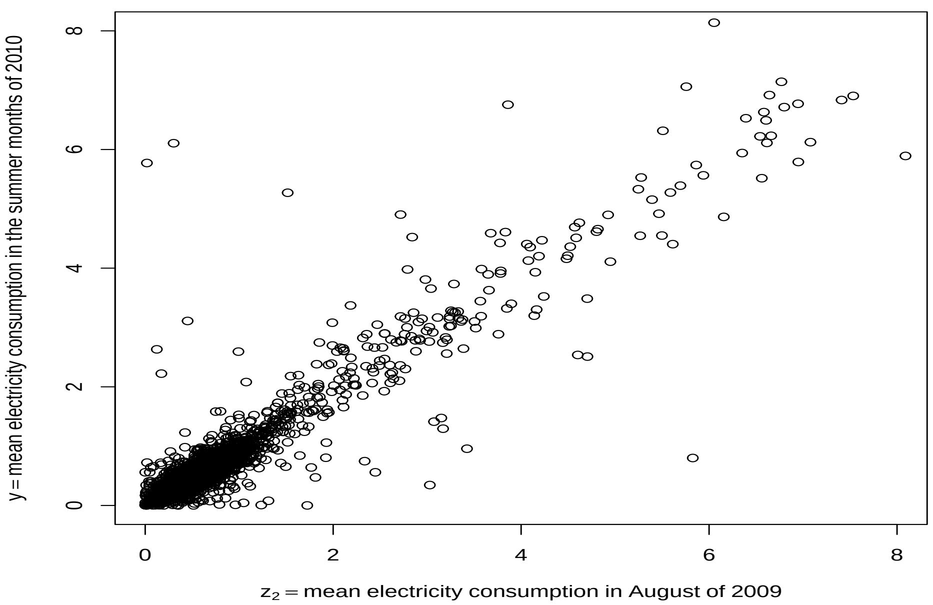

FIGURE 5.3: Scatter plot between y and z_2 for the real data set consisting of mean electricity consumption in the summer months of 2009 and 2010.

models, the performance of the GREG estimator of the finite population mean under different sampling designs depends on the degree of heteroscedasticity (see Chapter 3). Therefore, it is important to detect heteroscedasticity present in the data. [51] constructed statistical test for detecting heteroscedasticity based on QR in the classical set up involving i.i.d. sample observations. In this section, we shall construct similar tests under RHC and any high entropy sampling designs. Suppose that the population values $\{(Y_i, W_i) : 1 \leq i \leq N\}$ are generated from a superpopulation satisfying the linear model

$$Y_i = \boldsymbol{\theta}\mathbf{V}_i^T + (1 + \eta W_i^T)\epsilon_i, \tag{5.6.1}$$

where $\eta \in \mathbb{R}^{d+1}$, $\mathbf{V}_i$=$(1, W_i)$, and $\{\epsilon_i\}_{i=1}^N$ are i.i.d. random variables independent of $\{W_i\}_{i=1}^N$. This type of linear model was considered earlier in [51]. Under this linear model, one may be interested to check whether η=0. Note that the linear model in (5.6.1) can be expressed as

$$Y_i = \boldsymbol{\theta}(p)\mathbf{V}_i^T + (1 + \eta W_i^T)\epsilon_i(p), \tag{5.6.2}$$

where $\boldsymbol{\theta}(p)$=$\boldsymbol{\theta} + (Q_\epsilon(p), Q_\epsilon(p)\eta)$, $Q_\epsilon(p)$ is the p^{th} quantile of ϵ_i, and $\epsilon_i(p)$=$\epsilon_i - Q_\epsilon(p)$. Thus, if $l \geq 2$ and $p_1, \ldots, p_l \in (0,1)$, we have

$$\eta = 0 \Leftrightarrow \boldsymbol{\theta}(p_1)A^T = \cdots = \boldsymbol{\theta}(p_l)A^T \text{ for } A = [0^T \vdots I_{d+1}]. \tag{5.6.3}$$

Now, suppose that H_2=$B \boxtimes A^T$ with B being a $l \times (l-1)$ matrix such that

$$((B))_{ij} = \begin{cases} 1, \text{ if } j = i \text{ and } 1 \leq i \leq l-1, \\ -1, \text{ if } j = i-1 \text{ and } 2 \leq i \leq l, \\ 0, \text{ otherwise} . \end{cases} \tag{5.6.4}$$

Here, $\boxtimes$ denotes the Kronecker product. Then, for the diagnosis of heteroscedasticity present in the finite population observations, one can test the hypothesis (cf. [51])

$$\begin{aligned} H_0: \ & \big(\boldsymbol{\theta}(p_1), \ldots, \boldsymbol{\theta}(p_l)\big)H_2 = \big((\boldsymbol{\theta}(p_1) - \boldsymbol{\theta}(p_2))A^T, (\boldsymbol{\theta}(p_2) - \boldsymbol{\theta}(p_3))A^T, \\ & \ldots, (\boldsymbol{\theta}(p_{l-1}) - \boldsymbol{\theta}(p_l))A^T\big) = 0. \end{aligned} \tag{5.6.5}$$

A large sample test for the hypothesis mentioned above can be constructed based on the test statistic

$$\chi_n = n\big(\hat{\boldsymbol{\theta}}_n(p_1), \ldots, \hat{\boldsymbol{\theta}}_n(p_l)\big)H_2\big[H_2^T \hat{V} H_2\big]^{-1}H_2^T\big(\hat{\boldsymbol{\theta}}_n(p_1), \ldots, \hat{\boldsymbol{\theta}}_n(p_l)\big)^T, \tag{5.6.6}$$

where $\hat{\gamma}_n(p)$ is obtained from QR method, and $\hat{V}$ is a consistent estimator of the asymptotic covariance matrix of $\sqrt{n}\big(\hat{\boldsymbol{\theta}}_n(p_1) - \boldsymbol{\theta}(p_1), \ldots, \hat{\boldsymbol{\theta}}_n(p_l) - \boldsymbol{\theta}(p_l)\big)$. It follows from the proofs of Theorems 5.2.2 and 5.2.4 that under high entropy and RHC sampling designs, the asymptotic distribution of $\sqrt{n}\big(\hat{\boldsymbol{\theta}}_n(p_1) - \boldsymbol{\theta}(p_1), \ldots, \hat{\boldsymbol{\theta}}_n(p_l) - \boldsymbol{\theta}(p_l)\big)$ is normal with mean 0 and some p.d. covariance matrix. Hence, under H_0 and aforementioned sampling designs, the asymptotic distribution of χ_n is central chi-square with df $(l-1)(d+1)$.

The detection of heteroscedasticity can also be carried out based on the estimator obtained from ALS regression in the same way as above. For ALS regression, the p^{th} quantile of ϵ_i, $Q_\epsilon(p)$, in (5.6.2) is replaced by the p^{th} expectile of ϵ_i, $\mu_\epsilon(p)$, which is obtained by solving the equation (see [60])

$$\mu_\epsilon(p) - E_{\mathbf{P}}(\epsilon_i) = ((2p-1)/(1-p))\bigg(\int_{\mu_\epsilon(p)}^{\infty}(t - \mu_\epsilon(p))dF_\epsilon(t)\bigg), \tag{5.6.7}$$

where $F_\epsilon(t)$ is the distribution function of ϵ_i.

Now, we demonstrate the detection of heteroscedasticity discussed above based on synthetic data. For this, we choose N=5000 and generate the population values $\{(Y_i, X_i) : 1 \leq i \leq N\}$ from the heteroscedastic linear model Y_i=1000 $+ X_i + \epsilon_i(1 + X_i)$ and the homoscedastic linear model Y_i=1000 $+ X_i + \epsilon_i$ for i=$1, \ldots, N$. Here, X_i's and ϵ_i's are independently generated from the standard log-normal and the standard normal distributions, respectively. Note that we have W_i=Z_i=X_i for any given i. From these data sets, we draw I=1000 samples each of size n=100 using SRSWOR. Based on these samples, we perform the statistical tests discussed in the preceding paragraphs at 5% level. We choose l=3, and p_1=0.25, p_2=0.5 and p_3=0.75 in the cases of QR as well as ALS regression. For both the regression methods, we construct $\hat{V}$ in the same way as the consistent estimator of the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_n(p)$ (see Section 5.3). It follows from Section 5.3 that $\hat{V}$ depends on some smoothing parameter h and some density function $K(t)$. We choose $K(t)$ to be the uniform density function $\mathbb{1}_{[-1,1]}(t)$ and h by means of leave one out cross validation. Then, we compute proportions of times the tests reject the null hypothesis. The conclusions drawn from the above data analysis are summarized as follows.

(i) For the data set generated from the heteroscedastic model, the proportion of times the test based on QR reject the null hypothesis is 0.586, and the proportion of times the test based on ALS regression reject the null hypothesis is 0.59.

(ii) However, for the data set generated from the homoscedastic model, these proportions drop down to 0.048 and 0.042, respectively.

Next, based on the real data set considered in Section 5.2.1, we compute these proportions in the same way as in the case of synthetic data. The scatter plot in Figure 5.1 in Section 5.2.1 shows that there is heteroscedasticity present in this data set. We observe that for the above data set, the proportion of times the test based on QR reject the null hypothesis is 0.386, and the proportion of times the test based on ALS regression reject the null hypothesis is 0.414.

5.7. Concluding remarks

LS regression is extensively used to construct several estimators of finite population parameters. However, the use of regression methods like ALS, TLS, LAD, QR, etc. has been limited in the construction of different estimators in sample survey. Also, in the case of finite populations, large sample theory for the estimators obtained from different regression methods has not been adequately developed. In this chapter, asymptotic behavior of the estimators obtained from the above regression techniques is studied under high entropy and RHC sampling designs. Also, estimators of the finite population mean are constructed based on quantile and TLS regression. These estimators are then compared with the GREG estimator of the finite population mean, which is constructed using LS regression, based on their asymptotic distributions under several sampling designs.

As pointed out in the beginning of this chapter, it becomes challenging to derive different asymptotic results for the estimators obtained from various regression procedures, when the sample observations are neither independent nor identical. In this chapter, these results are first derived under rejective sampling designs using consistency and asymptotic normality of the HT estimator under these sampling designs following the ideas in [40] and [4]. Then, these results are derived under a high entropy sampling design using the fact that any high entropy sampling design can be approximated by a rejective sampling design in Kullback-Liebler divergence. Thus high entropy sampling designs play an important role in the study of the asymptotic behavior of the above-mentioned estimators, when the sample observations are neither independent nor identical.

It follows from the results discussed in Sections 5.2.1 and 5.4 that different estimators in regression analysis as well as different regression estimators of the finite population mean have the same performance under SRSWOR and LMS sampling designs. It also follows that these estimators sometimes may have worse performance under HEπPS and RHC sampling designs, which use the auxiliary information, than under SRSWOR. In practice, SRSWOR is easier to implement than the sampling designs that use the auxiliary information. Thus the above results are significant in view of selecting the appropriate sampling design.

As mentioned in the introduction and Section 5.4, the GREG estimator is more efficient than several other estimators (e.g., HT, RHC, ratio, product, etc.) of the finite population mean (see Chapter 2). However, it follows from an important result in Section 5.4 that the estimators of the finite population mean constructed based on quantile and TLS regression become more efficient than the GREG estimator under several sampling designs, whenever superpopulations satisfying linear models are considered, and errors in the linear models are generated from symmetric heavy-tailed superpopulation distributions like Laplace, Student's t, etc.

As discussed in Section 5.1, in sample survey, auxiliary variables are used to construct estimators and to implement sampling designs. Thus it becomes important to identify those auxiliary variables, which have more influence on the study variable than the others. On the other hand, heteroscedasticity influences the performance of the GREG estimator of the finite population mean under several sampling designs. In Chapter 3, it is shown that if the degree of heteroscedasticity present in linear regression models is not very large, then RHC and any HEπPS sampling designs, which use the auxiliary information, may have an adverse effect on the performance of the GREG estimator. It is also shown in Chapter 3 that if the degree of heteroscedasticity present in linear regression models is sufficiently large, then the aforementioned sampling designs improve the performance of the GREG estimator (see Theorem 3.2.3 in Chapter 3). Therefore, it also becomes important to detect heteroscedasticity present in the data. Variable selection and detection of heteroscadasticity were carried out in the earlier literature based on different regression techniques in the classical set up involving i.i.d. sample observations. In this chapter, we describe a variable selection method that uses the asymptotic results related to LS regression under high entropy and RHC sampling designs derived in this chapter. Under these sampling designs, we also construct a statistical test for detecting heteroscadasticity present in the data based on quantile regression.

5.8. Proofs of the main results

Suppose that

$$M_n(\mathbf{u}) = \sqrt{n}\sum_{i\in s}\pi_i^{-1}\mathbf{V}_i\psi_1\big(\epsilon_i - \mathbf{u}\mathbf{V}_i^T/\sqrt{n}\big)/N \text{ and}$$
$$L_n(\mathbf{u}) = M_n(\mathbf{u}) - M_n(0) - E_{\mathbf{P}^*}(M_n(\mathbf{u}) - M_n(0))$$

for any given $\mathbf{u} \in \mathbb{R}^{d+2}$, where ψ is as in (5.2.1) and ϵ_i is as in (5.2.3), and $\mathbf{V}_i$=$(1, W_i)$. Let us also suppose that $P(s,\omega)$ denotes a high entropy sampling design satisfying Assumption 5.2.6, and $Q(s,\omega)$ denotes a rejective sampling design having inclusion probabilities equal to those of $P(s,\omega)$. Such a rejective sampling design always exists (see [4]). Now, we give the proofs of the theorems.

Proof of Theorem 5.2.1. We shall first show that the result stated in (5.2.5) in Theorem 5.2.1 holds for the rejective sampling design $Q(s,\omega)$ and $d(i,s)$=π_i^{-1}. It is to be noted that $L(\boldsymbol{\theta})$= $\sum_{i\in s}\pi_i^{-1}\rho(Y_i - \boldsymbol{\theta}\mathbf{V}_i^T)$ is a convex function of $\boldsymbol{\theta}$ because ρ is a convex function. Therefore, $\nabla L(\hat{\boldsymbol{\theta}}_n)$=0 for $\hat{\boldsymbol{\theta}}_n = (\hat{\alpha}_n, \hat{\beta}_n) = \arg\min_{(\alpha,\beta)\in\mathbb{R}^{d+2}}\sum_{i\in s}\pi_i^{-1}\rho(Y_i - \alpha - \beta W_i^T)$ if $L(\boldsymbol{\theta})$ is differentiable at $\boldsymbol{\theta}$=$\hat{\boldsymbol{\theta}}_n$. Here, ∇L denotes the gradient of L. Recall from the paragraph containing (5.2.1) in Section 5.2 that ρ is differentiable at all but countably many $t \in \mathbb{R}$. Let $\{t_l\}$ be the real numbers, where ρ is not differentiable. Since (Y_i, W_i)'s have absolutely continuous distribution, we can say that *a.s.* [**P**],

$$\epsilon_i - \hat{\mathbf{u}}_n\mathbf{V}_i^T/\sqrt{n} - t_l = Y_i - \hat{\boldsymbol{\theta}}_n\mathbf{V}_i^T - t_l \neq 0 \text{ for any } i = 1,\ldots,N, s\in\mathcal{S}, \nu\geq 1 \text{ and } l = 1,2\ldots,$$

where $\hat{\mathbf{u}}_n$=$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$. Hence, *a.s.* [**P**], ρ is differentiable at $Y_i - \hat{\boldsymbol{\theta}}_n\mathbf{V}_i^T$ for any $1 \leq i \leq N$, $s \in \mathcal{S}$ and $\nu \geq 1$. Thus *a.s.* [**P**],

$$(\sqrt{n}/N)\nabla L(\hat{\boldsymbol{\theta}}_n) = -\sqrt{n}\sum_{i\in s}\pi_i^{-1}\mathbf{V}_i\psi(Y_i - \hat{\boldsymbol{\theta}}_n\mathbf{V}_i^T)/N$$
$$= -\sqrt{n}\sum_{i\in s}\pi_i^{-1}\mathbf{V}_i\psi(\epsilon_i - \hat{\mathbf{u}}_n\mathbf{V}_i^T/\sqrt{n})/N = 0$$

for any $s \in \mathcal{S}$ and $\nu \geq 1$. This is because $\psi(t)$=$\rho'(t)$, when ρ is differentiable at t (recall from the paragraph containing (5.2.1) in Section 5.2). Then, we have under $\mathbf{P}^*$, as $\nu\to\infty$,

$$M_n(\hat{\mathbf{u}}_n) = \sqrt{n}\sum_{i\in s}\pi_i^{-1}\mathbf{V}_i\psi(\epsilon_i - \hat{\mathbf{u}}_n\mathbf{V}_i^T/\sqrt{n})/N = o_p(1) \quad (5.8.1)$$

for $Q(s,\omega)$. Now, using (5.8.1), Lemma 5.9.3, and similar arguments as in the proof of Theorem 3.1 in [51], we can say that under $\mathbf{P}^*$, as $\nu \to \infty$, $\hat{\mathbf{u}}_n$=$O_p(1)$ for $Q(s,\omega)$. Then, using (5.9.14) in the proof of Lemma 5.9.3, one can show that under $\mathbf{P}^*$, as $\nu \to \infty$, $M_n(\hat{\mathbf{u}}_n) - M_n(0) + \hat{\mathbf{u}}_n\Sigma$=$o_p(1)$ for $Q(s,\omega)$. This result and (5.8.1) above further imply that under $\mathbf{P}^*$, as $\nu \to \infty$,

$$M_n(0) - \hat{\mathbf{u}}_n\Sigma = o_p(1) \tag{5.8.2}$$

for $Q(s,\omega)$. Therefore, under $\mathbf{P}^*$, as $\nu \to \infty$,

$$\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} = \left[\sum_{i\in s} \pi_i^{-1}\psi(\epsilon_i)\mathbf{V}_i/N\right]\Sigma^{-1} + o_p(1/\sqrt{n}) \tag{5.8.3}$$

for $Q(s,\omega)$. One can similarly show that under $\mathbf{P}$, as $\nu \to \infty$,

$$\boldsymbol{\theta}_N - \boldsymbol{\theta} = \left[\sum_{i=1}^{N} \psi(\epsilon_i)\mathbf{V}_i/N\right]\Sigma^{-1} + o_p(1/\sqrt{n}). \tag{5.8.4}$$

Hence, using (5.8.3) and (5.8.4), we can say that (5.2.5) in the statement of Theorem 5.2.1 holds for $Q(s,\omega)$ and $d(i,s)$=π_i^{-1}.

Now, we shall show that (5.2.5) in the statement of Theorem 5.2.1 holds for high entropy sampling design $P(s,\omega)$ (which satisfies Assumption 5.2.6) and $d(i,s)$=π_i^{-1}. Suppose that

$$\mathcal{S}_1 = \left\{s \in \mathcal{S} : \sqrt{n}\left|\left|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_N - \left(\sum_{i\in s} \pi_i^{-1}\psi(\epsilon_i)\mathbf{V}_i/N - \sum_{i=1}^{N}\psi(\epsilon_i)\mathbf{V}_i/N\right)\Sigma^{-1}\right|\right| > \delta\right\}$$

for any given $\delta > 0$. Then, for any $\omega \in \Omega$ and $\nu \geq 1$,

$$\left|\sum_{s\in\mathcal{S}_1}\big(P(s,\omega) - Q(s,\omega)\big)\right| \leq \sum_{s\in\mathcal{S}}\Big|P(s,\omega) - Q(s,\omega)\Big| \leq D(P||Q) \leq D(P||R)$$

by Lemmas 2 and 3 in [4], where $R(s,\omega)$ is such a rejective sampling design that $D(P||R) \to 0$ as $\nu \to \infty$ *a.s.* [$\mathbf{P}$]. Therefore,

$$\sum_{s\in\mathcal{S}_1}\big(P(s,\omega) - Q(s,\omega)\big) \to 0 \text{ as } \nu \to \infty \text{ } a.s. \text{ } [\mathbf{P}], \text{ and hence}$$
$$E_{\mathbf{P}}\left[\sum_{s\in\mathcal{S}_1}\big(P(s,\omega) - Q(s,\omega)\big)\right] \to 0 \text{ as } \nu \to \infty$$

by DCT. Now, since

$$E_{\mathbf{P}}\bigg[\sum_{s\in\mathcal{S}_1} Q(s,\omega)\bigg] = \mathbf{P}^*\bigg[\sqrt{n}\bigg|\bigg|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_N - \bigg(\sum_{i\in s}\pi_i^{-1}\psi(\epsilon_i)\mathbf{V}_i/N - \sum_{i=1}^{N}\psi(\epsilon_i)\mathbf{V}_i/N\bigg)\Sigma^{-1}\bigg|\bigg| > \delta\bigg] \to 0 \text{ as } \nu \to \infty \text{ for any given } \delta > 0,$$

$$E_{\mathbf{P}}\bigg[\sum_{s\in\mathcal{S}_1} P(s,\omega)\bigg] \to 0 \text{ as } \nu \to \infty \text{ for any given } \delta > 0.$$

Thus (5.2.5) in the statement of Theorem 5.2.1 holds for high entropy sampling design $P(s,\omega)$ and $d(i,s)$=π_i^{-1} because $P(s,\omega)$ and $Q(s,\omega)$ have same inclusion probabilities. Similarly, (5.2.6) in the statement of Theorem 5.2.1 holds for $P(s,\omega)$ and $d(i,s)$=π_i^{-1} based on the result stated in (5.8.3). □

Proof of Theorem 5.2.2. It is enough to show that the results stated in (5.2.7) and (5.2.8) in Theorem 5.2.1 hold for the rejective sampling design $Q(s,\omega)$ and $d(i,s)$=π_i^{-1}. Then, these results hold for high entropy sampling design $P(s,\omega)$ (which satisfies Assumption 5.2.6) and $d(i,s)$=π_i^{-1} in the same way as (5.2.5) and (5.2.6) in Theorem 5.2.1 hold for $P(s,\omega)$ and $d(i,s)$=π_i^{-1} in the 2^{nd} paragraph of the proof of Theorem 5.2.1. Let us fix $\mathbf{m} \in \mathbb{R}^{d+2}$ such that $\mathbf{m} \neq 0$. Then, it follows from Lemma 5.9.2 in Section 5.9 that under $Q(s,\omega)$, as $\nu \to \infty$,

$$\sqrt{n}\mathbf{m}\bigg[\sum_{i\in s}\pi_i^{-1}\psi(\epsilon_i)\mathbf{V}_i/N - \sum_{i=1}^{N}\psi(\epsilon_i)\mathbf{V}_i/N\bigg]^T \xrightarrow{\mathcal{L}} N(0,\mathbf{m}\Gamma\mathbf{m}^T) \tag{5.8.5}$$

a.s. [**P**]. Then, using DCT, one can show that under $\mathbf{P}^*$, as $\nu \to \infty$,

$$\sqrt{n}\mathbf{m}\bigg[\sum_{i\in s}\pi_i^{-1}\psi(\epsilon_i)\mathbf{V}_i/N - \sum_{i=1}^{N}\psi(\epsilon_i)\mathbf{V}_i/N\bigg]^T \xrightarrow{\mathcal{L}} N(0,\mathbf{m}\Gamma\mathbf{m}^T) \tag{5.8.6}$$

for $Q(s,\omega)$. It also follows from the 1^{st} paragraph in the proof of Theorem 5.2.1 that (5.2.5) in the statement of Theorem 5.2.1 holds for $Q(s,\omega)$ and $d(i,s)$=π_i^{-1}. Therefore, using (5.8.6), we can say that under $\mathbf{P}^*$, as $\nu \to \infty$,

$$\sqrt{n}\mathbf{m}\big(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_N\big)^T \xrightarrow{\mathcal{L}} N\Big(0, \mathbf{m}\big(\Sigma^{-1}\Gamma\Sigma^{-1}\big)\mathbf{m}^T\Big) \tag{5.8.7}$$

for $Q(s,\omega)$ and any given $\mathbf{m} \neq 0$. Thus (5.2.7) in the statement of Theorem 5.2.2 holds for $Q(s,\omega)$ and $d(i,s)$=π_i^{-1}.

Next, it follows from the paragraph containing (5.9.17) and (5.9.18) in the proof of Lemma 5.9.3 in Section 5.9 that $E_{\mathbf{P}}(\psi(\epsilon_i)\mathbf{V}_i)$=0. Then, under $\mathbf{P}$, as $\nu \to \infty$,

$$\sqrt{N}\mathbf{m}\Big[\sum_{i=1}^{N}\psi(\epsilon_i)\mathbf{V}_i/N\Big]^T \xrightarrow{\mathcal{L}} N\big(0, \mathbf{m}E_{\mathbf{P}}\big((\psi^2(\epsilon_i)\mathbf{V}_i^T\mathbf{V}_i\big)\mathbf{m}^T\big) \tag{5.8.8}$$

by CLT. Now, using (5.8.5), (5.8.8), Assumption 5.2.1, and (iii) of Theorem 5.1 in [69], one can show that under $\mathbf{P}^*$, as $\nu \to \infty$,

$$\sqrt{n}\mathbf{m}\Big[\sum_{i\in s}\pi_i^{-1}\psi(\epsilon_i)\mathbf{V}_i/N\Big]^T \xrightarrow{\mathcal{L}} N\Big(0, \mathbf{m}\big(\Gamma + \lambda E_{\mathbf{P}}\big((\psi^2(\epsilon_i)\mathbf{V}_i^T\mathbf{V}_i)\big)\mathbf{m}^T\Big) \tag{5.8.9}$$

for $Q(s,\omega)$. Therefore, it follows from (5.8.3) that under $\mathbf{P}^*$, as $\nu \to \infty$,

$$\sqrt{n}\mathbf{m}\big(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\big) \xrightarrow{\mathcal{L}} N\big(0, \mathbf{m}\Delta\mathbf{m}^T\big) \tag{5.8.10}$$

for $Q(s,\omega)$ and any given $\mathbf{m} \neq 0$. Thus (5.2.8) in the statement of Theorem 5.2.2 holds for $Q(s,\omega)$ and $d(i,s)$=π_i^{-1}. □

Proof of Theorem 5.2.3. Let us first define $\tilde{H}_{ij}$=$\big(\psi(\epsilon_i - \mathbf{u}\mathbf{V}_i^T/\sqrt{n}) - \psi(\epsilon_i)\big)V_{ij}$ for i=$1,\ldots,N$ and j=$1,\ldots,d+2$, where V_{ij} is the j^{th} component of $\mathbf{V}_i$. Then, note that (cf. [20], [66], cf. [61], etc.) given any $\omega \in \Omega$ and j=$1,\ldots,d+2$, under RHC sampling design,

$$var\Big(\sqrt{n}\sum_{i\in s}(NX_i)^{-1}G_i\tilde{H}_{ij}\Big) = (n\gamma)\Big[\overline{X}\sum_{i=1}^{N}(\tilde{H}_{ij})^2/NX_i - \Big(\sum_{i=1}^{N}\tilde{H}_{ij}/N\Big)^2\Big], \tag{5.8.11}$$

where $\overline{X}$=$\sum_{i=1}^{N}X_i/N$, γ=$\sum_{r=1}^{n}\tilde{N}_r(\tilde{N}_r-1)/N(N-1)$, and $\{\tilde{N}_r\}_{r=1}^{n}$ are as in the paragraph preceding Assumption 5.2.8. Since $n\gamma \to c$ as $\nu \to \infty$ for some $c \geq 1-\lambda > 0$ by Lemma 2.7.5 in Section 2.7 of Chapter 2, it can be shown using (5.8.11) and Assumption 5.2.8 that given any j=$1,\ldots,d+2$, under RHC sampling design,

$$var\Big(\sqrt{n}\sum_{i\in s}(NX_i)^{-1}G_i\tilde{H}_{ij}\Big) \leq K_1\sum_{i=1}^{N}(\tilde{H}_{ij})^2/N \tag{5.8.12}$$

for all sufficiently large ν and some constant $K_1 > 0$ (may depend on ω) *a.s.* [$\mathbf{P}$]. Now, if we consider $M_n(\mathbf{u})$ and $L_n(\mathbf{u})$ as mentioned in the paragraph preceding Lemma 5.9.1 with π_i^{-1} replaced by $G_iX_i^{-1}$, then using (5.8.12), it can be shown in the same way as the proof of Lemma 5.9.1 that the result stated in (5.9.1) in Lemma 5.9.1 holds for RHC sampling design. Next, it follows from Lemma 5.9.4 that for any given j=1, . . . , $d+2$, under RHC sampling design, as $\nu \to \infty$,

$$\sqrt{n}\mathbf{e}_j \Big[\sum_{i \in s} (NX_i)^{-1} G_i \psi(\epsilon_i) \mathbf{V}_i - \sum_{i=1}^{N} \psi(\epsilon_i) \mathbf{V}_i / N \Big]^T \xrightarrow{\mathcal{L}} N(0, \mathbf{e}_j \Gamma^* \mathbf{e}_j^T) \tag{5.8.13}$$

a.s. [**P**], where $\{\mathbf{e}_j : 1 \leq j \leq d+2\}$ are canonical basis vectors of $\mathbb{R}^{d+2}$. Then, using DCT, one can show that for any given j=1, . . . , $d+2$, under $\mathbf{P}^*$, as $\nu \to \infty$,

$$\begin{aligned} &\sqrt{n}\mathbf{e}_j \Big[\sum_{i \in s} (NX_i)^{-1} G_i \psi(\epsilon_i) \mathbf{V}_i - \sum_{i=1}^{N} \psi(\epsilon_i) \mathbf{V}_i / N \Big]^T = \\ &\sqrt{n} \Big[\sum_{i \in s} \pi_i^{-1} \psi(\epsilon_i) V_{ij} / N - \sum_{i=1}^{N} \psi(\epsilon_i) V_{ij} / N \Big] = O_p(1) \end{aligned} \tag{5.8.14}$$

for RHC sampling design. Now, if $M_n(\mathbf{u})$ is considered as in the paragraph preceding Lemma 5.9.1 with π_i^{-1} replaced by $G_i X_i^{-1}$, then using (5.8.14), one can show in the same way as the proof of Lemma 5.9.3 that (5.9.12) in Lemma 5.9.3 holds for RHC sampling design. Thus (5.2.10) and (5.2.11) in the statement of Theorem 3 hold for RHC sampling design and $d(i,s)$=$G_i X_i^{-1}$ in the same way as (5.2.5) and (5.2.6) in the statement of Theorem 5.2.1 hold for the rejective sampling design $Q(s,\omega)$ and $d(i,s)$=π_i^{-1} in the 1^{st} paragraph of the proof of Theorem 1 above. □

Proof of Theorem 5.2.4. Using Lemma 5.9.4, one can show that the conclusion of Theorem 5.2.4 holds for RHC sampling design and $d(i,s)$=$G_i X_i^{-1}$ in the same way as the conclusion of Theorem 5.2.2 holds for the rejective sampling design $Q(s,\omega)$ and $d(i,s)$=π_i^{-1} (see the proof of Theorem 5.2.2 above). □

Proof of Theorem 5.2.5. Let us denote the asymptotic covariance matrices of $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_N)$ under SRSWOR, LMS, RHC and any HEπPS sampling by Γ_{SRS}, Γ_{LMS}, Γ_{RHC} and $\Gamma_{HE\pi PS}$, respectively. It follows from (5.2.7) in Theorem 5.2.2, (5.2.12) in Theorem 5.2.4, and the proof of Lemma 5.9.5 in Section 5.9 that

$$\begin{aligned} &\Gamma_{SRS} = \Gamma_{LMS} = (1-\lambda) E_{\mathbf{P}}(\psi(\epsilon_i))^2 \Sigma^{-1} E_{\mathbf{P}}(\mathbf{V}_i^T \mathbf{V}_i) \Sigma^{-1}, \\ &\Gamma_{RHC} = c\mu_x E_{\mathbf{P}}(\psi(\epsilon_i))^2 \Sigma^{-1} E_{\mathbf{P}}(\mathbf{V}_i^T \mathbf{V}_i X_i^{-1}) \Sigma^{-1} \text{ and} \\ &\Gamma_{HE\pi PS} = E_{\mathbf{P}}(\psi(\epsilon_i))^2 \Sigma^{-1} E_{\mathbf{P}}(\mathbf{V}_i^T \mathbf{V}_i)(\mu_x X_i^{-1} - \lambda) \Sigma^{-1}, \end{aligned}$$

where

$$\mu_x = E_{\mathbf{P}}(X_i), \Sigma = -\partial\Big(E_{\mathbf{P}}\big(\psi(\epsilon_i - t)\big)\Big)\Big/\partial t\Big|_{t=0} \times E_{\mathbf{P}}(\mathbf{V}_i^T\mathbf{V}_i) \text{ and } c = \lim_{\nu\to\infty} n\gamma.$$

Thus the result that the asymptotic total variance of $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_N)$ under SRSWOR is the same as that of $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_N)$ under LMS sampling design follows. Next, we have

$$tr(\Gamma_{RHC} - \Gamma_{SRS}) = Ktr\Bigg[\Big(E_{\mathbf{P}}(\mathbf{V}_i^T\mathbf{V}_i)\Big)^{-1}E_{\mathbf{P}}\Big(\big(c\mu_x X_i^{-1} - (1-\lambda)\big)\mathbf{V}_i^T\mathbf{V}_i\Big)\times \Big(E_{\mathbf{P}}(\mathbf{V}_i^T\mathbf{V}_i)\Big)^{-1}\Bigg] \geq K(1-\lambda)tr\Bigg[\Big(E_{\mathbf{P}}(\mathbf{V}_i^T\mathbf{V}_i)\Big)^{-1}E_{\mathbf{P}}\Big(\big(\mu_x X_i^{-1} - 1\big)\mathbf{V}_i^T\mathbf{V}_i\Big)\times \Big(E_{\mathbf{P}}(\mathbf{V}_i^T\mathbf{V}_i)\Big)^{-1}\Bigg]$$

for some $K > 0$ because $c \geq 1 - \lambda$ by Lemma 2.7.5 in Section 2.7 of Chapter 2. Moreover, we have

$$tr(\Gamma_{HE\pi PS} - \Gamma_{SRS}) = Ktr\Bigg[\Big(E_{\mathbf{P}}(\mathbf{V}_i^T\mathbf{V}_i)\Big)^{-1}E_{\mathbf{P}}\Big(\big(\mu_x X_i^{-1} - 1\big)\mathbf{V}_i^T\mathbf{V}_i\Big)\Big(E_{\mathbf{P}}(\mathbf{V}_i^T\mathbf{V}_i)\Big)^{-1}\Bigg].$$

Therefore, $tr(\Gamma_{SRS}) < \min\{tr(\Gamma_{RHC}), tr(\Gamma_{HE\pi PS})\}$ if and only if the condition in (5.2.15) holds. This completes the proof of the theorem. □

Proof of Theorem 5.3.1. *(i)* We shall first show that under $\mathbf{P}^*$, as $\nu \to \infty$, $\hat{\Gamma}_1 \xrightarrow{p} \Gamma_1$ for the rejective sampling design $Q(s,\omega)$, where $Q(s,\omega)$ is as mentioned in the paragraph preceding the proof of Theorem 5.2.1. Then, this result will hold for high entropy sampling design $P(s,\omega)$ (which satisfies Assumption 5.2.6) in the same way as (5.2.5) in Theorem 5.2.1 holds for $P(s,\omega)$ and $d(i,s)$=π_i^{-1} in the 2^{nd} paragraph of the proof of Theorem 5.2.1 above. In order to show that under $\mathbf{P}^*$, $\hat{\Gamma}_1 \xrightarrow{p} \Gamma_1$ as $\nu \to \infty$ for $Q(s,\omega)$, we need to first show that under $\mathbf{P}^*$,

$$\hat{\Sigma}_1 \xrightarrow{p} \Sigma \text{ as } \nu \to \infty$$

for $Q(s,\omega)$. Let us define

$$\tilde{\Sigma}_1 = -\sum_{i\in s}\pi_i^{-1}\phi'(0,W_i)\mathbf{V}_i^T\mathbf{V}_i/N \text{ and } \Sigma_1^* = -\sum_{i=1}^{N}\phi'(0,W_i)\mathbf{V}_i^T\mathbf{V}_i/N$$

. We establish the consistency of $\hat{\Sigma}_1$ by showing that as $\nu \to \infty$, $\hat{\Sigma}_1 - \tilde{\Sigma}_1 \xrightarrow{p} 0$ and $\tilde{\Sigma}_1 - \Sigma_1^* \xrightarrow{p} 0$ under $\mathbf{P}^*$ for $Q(s,\omega)$, and $\Sigma_1^* \xrightarrow{p} \Sigma$ under $\mathbf{P}$.

The result

$$\Sigma_1^* \xrightarrow{p} \Sigma \text{ as } \nu \to \infty \text{ under } \mathbf{P}$$

holds by weak law of large numbers since $E_{\mathbf{P}}||\phi'(0,W_i)\mathbf{V}_i^T\mathbf{V}_i||| < \infty$ by Assumptions 3 and 4. Next, note that the $(j,l)^{th}$ element of $\tilde{\Sigma}_1$ is $((\tilde{\Sigma}_1))_{jl}$= $-\sum_{i\in s}\pi_i^{-1}\phi'(0,W_i)V_{ij}V_{il}/N$ for j,l=$1,\ldots,d+2$. Then, it follows from Theorem 6.1 in [40] that given any $\omega\in\Omega$ and j,l=$1,\ldots,d+2$, under $Q(s,\omega)$,

$$nvar\Big(((\tilde{\Sigma}_1))_{jl}\Big) = (n/N^2)\Big[\sum_{i=1}^{N}(\phi'(0,W_i)V_{ij}V_{il})^2(\pi_i^{-1}-1)- \Big(\sum_{i=1}^{N}\phi'(0,W_i)V_{ij}V_{il}(1-\pi_i)\Big)^2\Big/\sum_{i=1}^{N}\pi_i(1-\pi_i)\Big](1+e), \tag{5.8.15}$$

where $e\to 0$ if $\sum_{i=1}^{N}\pi_i(1-\pi_i)\to\infty$ as $\nu\to\infty$. Recall from the proof of Lemma 5.9.1 that under $Q(s,\omega)$, $\sum_{i=1}^{N}\pi_i(1-\pi_i)\to\infty$ as $\nu\to\infty$ *a.s.* [**P**]. Therefore, using (5.8.15) and Assumption 5.2.6–(i), we can show that given any j,l=$1,\ldots,d+2$, under $Q(s,\omega)$,

$$nvar\Big(((\tilde{\Sigma}_1))_{jl}\Big) \le (n/N^2)\sum_{i=1}^{N}(\phi'(0,W_i)V_{ij}V_{il})^2\pi_i^{-1} \le K_1\sum_{i=1}^{N}(\phi'(0,W_i)V_{ij}V_{il})^2/N \tag{5.8.16}$$

for all sufficiently large ν and some constant $K_1>0$ (may depend on ω) *a.s.* [**P**]. Now, $\sum_{i=1}^{N}(\phi'(0,W_i)V_{ij}V_{il})^2/N$=$O(1)$ as $\nu\to\infty$ *a.s.* [**P**] by SLLN since $E_{\mathbf{P}}(\phi'(0,W_i)V_{ij}V_{il})^2<\infty$ by Assumptions 3 and 4. Thus under $Q(s,\omega)$,

$$((\tilde{\Sigma}_1))_{jl}-((\Sigma_1^*))_{jl}\xrightarrow{p} 0 \text{ as } \nu\to\infty \textit{ a.s. } [\mathbf{P}]$$

for any given j,l=$1,\ldots,d+2$. Using DCT, one can then show that under $\mathbf{P}^*$,

$$\tilde{\Sigma}_1-\Sigma_1^*\xrightarrow{p} 0 \text{ as } \nu\to\infty \text{ for } Q(s,\omega).$$

Next, suppose that

$$\xi(y_2)=\partial\Big(\int_{\mathbb{R}}\psi(hy_1-y_2-t)K(y_1)dy_1\Big)/\partial t|_{t=0}$$

for $y_2\in\mathbb{R}$. Then, we have

$$\hat{\phi}'(0, W_i) = \sum_{j\in s} \pi_j^{-1} \prod_{k=1}^{d+1} K_h\big(W_{ik} - W_{jk}\big)\xi(\hat{\boldsymbol{\theta}}_n \mathbf{V}_i^T - Y_j)\Big/ \sum_{j\in s} \pi_j^{-1} \prod_{k=1}^{d+1} K_h\big(W_{ik} - W_{jk}\big)$$

for any given i=$1, \ldots, N$. Let us define

$$\tilde{\phi}'(0, W_i) = \sum_{j\in s} \pi_j^{-1} \prod_{k=1}^{d+1} K_h\big(W_{ik} - W_{jk}\big)\xi(\boldsymbol{\theta} \mathbf{V}_i^T - Y_j)\Big/ \sum_{j\in s} \pi_j^{-1} \prod_{k=1}^{d+1} K_h\big(W_{ik} - W_{jk}\big)$$

for i=$1, \ldots, N$. Now, suppose that **u**=$(1, \mathbf{u}_1)$, where $\mathbf{u}_1 \in \mathbb{R}^{d+1}$ and $\mathbf{u} \in \mathbb{R}^{d+2}$. Further, suppose that u_{1k} is the k^{th} component of $\mathbf{u}_1$ for k=$1, \ldots, d+1$. Then, it can be shown in the same way as in the preceding paragraph that under $\mathbf{P}^*$, as $\nu \to \infty$,

$$\sup_{||\mathbf{u}||\le K_1} \left| \tilde{\phi}'(0, \mathbf{u}_1) - \sum_{j=1}^{N} \prod_{k=1}^{d+1} K_h\big(u_{1k} - W_{jk}\big)\xi(\boldsymbol{\theta}\mathbf{u}^T - Y_j)\Big/ \sum_{j=1}^{N} \prod_{k=1}^{d+1} K_h\big(u_{1k} - W_{jk}\big) \right| \xrightarrow{p} 0$$

for $Q(s, \omega)$. It can also be shown that under **P**,

$$\sup_{||\mathbf{u}||\le K_1} \left| \sum_{j=1}^{N} \prod_{k=1}^{d+1} K_h\big(\mathbf{u}_1 - W_{jk}\big)\xi(\boldsymbol{\theta}\mathbf{u}^T - Y_j)\Big/ \sum_{j=1}^{N} \prod_{k=1}^{d+1} K_h\big(u_{1k} - W_{jk}\big) - \phi'(0, \mathbf{u}_1) \right| \xrightarrow{p} 0$$

as $\nu \to \infty$, and under $\mathbf{P}^*$,

$$\sup_{||\mathbf{u}||\le K_1} \left|\tilde{\phi}'(0, \mathbf{u}_1) - \hat{\phi}'(0, \mathbf{u}_1)\right| \xrightarrow{p} 0 \text{ as } \nu \to \infty \text{ for } Q(s, \omega).$$

Moreover, under $\mathbf{P}^*$, as $\nu \to \infty$, $\sum_{i\in s} \pi_i^{-1} ||\mathbf{V}_i||^2/N$=$O_p(1)$ for $Q(s, \omega)$. Thus under $\mathbf{P}^*$, as $\nu \to \infty$,

$$||\tilde{\Sigma}_1 - \hat{\Sigma}_1|| \le \sum_{i\in s} \pi_i^{-1} |\hat{\phi}'(0, W_i) - \phi'(0, W_i)|||\mathbf{V}_i||^2/N \le$$
$$\sup_{||\mathbf{u}||\le K_1} |\hat{\phi}'(0, \mathbf{u}_1) - \phi'(0, \mathbf{u}_1)| \sum_{i\in s} \pi_i^{-1} ||\mathbf{V}_i||^2/N \xrightarrow{p} 0$$

for $Q(s, \omega)$. Therefore, under $\mathbf{P}^*$, as $\nu \to \infty$, $\hat{\Sigma}_1 \xrightarrow{p} \Sigma$, and hence $\hat{\Sigma}_1^{-1} \xrightarrow{p} \Sigma^{-1}$ for $Q(s, \omega)$.

Next, note that we have

$$\hat{\Gamma} = (n/N^2)\Bigg[\sum_{i\in s} \hat{\mathbf{H}}_i^T \hat{\mathbf{H}}_i (\pi_i^{-1} - 1)\pi_i^{-1} - \qquad (5.8.17)$$

$$\Big(\sum_{i\in s}\hat{\mathbf{H}}_i^T(\pi_i^{-1}-1)\sum_{i\in s}\hat{\mathbf{H}}_i(\pi_i^{-1}-1)\Big)\Big/\sum_{i\in s}(1-\pi_i)\Big],$$

where $\hat{\mathbf{H}}_i$=$\psi(\hat{\epsilon}_i)\mathbf{V}_i$ for any given $i\in s$. The first term on the right hand side of (5.8.17) can be expressed as

$$\begin{aligned}&(n/N^2)\Big[\sum_{i\in s}\hat{\mathbf{H}}_i^T\hat{\mathbf{H}}_i(\pi_i^{-1}-1)\pi_i^{-1}\Big]=\\&(n/N^2)\Big[\sum_{i\in s}\big(\psi^2(\hat{\epsilon}_i)-\psi^2(\epsilon_i)\big)\mathbf{V}_i^T\mathbf{V}_i(\pi_i^{-1}-1)\pi_i^{-1}\Big]+\\&(n/N^2)\Big[\sum_{i\in s}\mathbf{H}_i^T\mathbf{H}_i(\pi_i^{-1}-1)\pi_i^{-1}\Big],\end{aligned}\tag{5.8.18}$$

where $\mathbf{H}_i$=$\psi(\epsilon_i)\mathbf{V}_i$ for i=$1,\ldots,N$. One can show that

$$(n/N^2)\Big[\sum_{i\in s}\mathbf{H}_i^T\mathbf{H}_i(\pi_i^{-1}-1)\pi_i^{-1}-\sum_{i=1}^{N}\mathbf{H}_i^T\mathbf{H}_i(\pi_i^{-1}-1)\Big]\xrightarrow{p}0$$

as $\nu\to\infty$ under $\mathbf{P}^*$ for $Q(s,\omega)$ in the same way as $\tilde{\Sigma}_1-\Sigma_1^*\xrightarrow{p}0$ as $\nu\to\infty$ under $\mathbf{P}^*$ for $Q(s,\omega)$ in the 2^{nd} paragraph of this proof. Moreover, we have

$$\begin{aligned}&\Big|\Big|(n/N^2)\Big[\sum_{i\in s}\big(\psi^2(\hat{\epsilon}_i)-\psi^2(\epsilon_i)\big)\mathbf{V}_i^T\mathbf{V}_i(\pi_i^{-1}-1)\pi_i^{-1}\Big]\Big|\Big|\leq\\&(n/N^2)\Big[\Big(\max_{1\leq i\leq N}|\psi^2(\hat{\epsilon}_i)-\psi^2(\epsilon_i)|\Big)\sum_{i\in s}||\mathbf{V}_i||^2(\pi_i^{-1}-1)\pi_i^{-1}\Big].\end{aligned}\tag{5.8.19}$$

Using (5.8.3) and (5.8.4) in the proof of Theorem 5.2.1 above, one can show that under $\mathbf{P}^*$, as $\nu\to\infty$,

$$\Big(\max_{1\leq i\leq N}|\psi^2(\hat{\epsilon}_i)-\psi^2(\epsilon_i)|\Big)\xrightarrow{p}0\text{ for }Q(s,\omega)$$

. Therefore, using Assumption 5.2.4, it can be shown in the same way as in the 2^{nd} paragraph of this proof that both the right hand side of (5.8.19) converges to 0 in probability under $\mathbf{P}^*$. Hence, it follows from (5.8.18) that under $\mathbf{P}^*$, as $\nu\to\infty$,

$$(n/N^2)\Big[\sum_{i\in s}\hat{\mathbf{H}}_i^T\hat{\mathbf{H}}_i(\pi_i^{-1}-1)\pi_i^{-1}-\sum_{i=1}^{N}\mathbf{H}_i^T\mathbf{H}_i(\pi_i^{-1}-1)\Big]\xrightarrow{p}0$$

for $Q(s,\omega)$. Similarly, one can show that under $\mathbf{P}^*$, as $\nu\to\infty$,

$$(n/N^2)\Big[\Big\{\sum_{i\in s}\hat{\mathbf{H}}_i^T(\pi_i^{-1}-1)\sum_{i\in s}\hat{\mathbf{H}}_i(\pi_i^{-1}-1)\Big\}\Big/\sum_{i\in s}(1-\pi_i)-$$

$$\left\{\sum_{i=1}^{N}\mathbf{H}_i(1-\pi_i)\sum_{i=1}^{N}\mathbf{H}_i(1-\pi_i)\right\}\Big/\sum_{i=1}^{N}\pi_i(1-\pi_i)\Big] \xrightarrow{p} 0$$

for $Q(s,\omega)$. Thus

$$\hat{\Gamma}-(n/N^2)\Big[\sum_{i=1}^{N}\mathbf{H}_i^T\mathbf{H}_i(\pi_i^{-1}-1)-$$
$$\left\{\sum_{i=1}^{N}\mathbf{H}_i^T(1-\pi_i)\sum_{i=1}^{N}\mathbf{H}_i(1-\pi_i)\right\}\Big/\sum_{i=1}^{N}\pi_i(1-\pi_i)\Big] \xrightarrow{p} 0,$$

and hence $\hat{\Gamma}\xrightarrow{p}\Gamma$ as $\nu\to\infty$ under $\mathbf{P}^*$ for $Q(s,\omega)$ by Assumption 5.2.6–(ii). Therefore, under $\mathbf{P}^*$, as $\nu\to\infty$,

$$\hat{\Gamma}_1\xrightarrow{p}\Gamma_1 \text{ for the rejective sampling design } Q(s,\omega).$$

Next, the result, $\hat{\Gamma}_1^*\xrightarrow{p}\Gamma_1^*$ as $\nu\to\infty$ for RHC sampling design under $\mathbf{P}^*$, will follow in the same way as the above result.

(ii) The proof follows exactly the same way as the proof of *(i)*. □

Proof of Theorem 5.4.1. Let us first assume that $\rho(t)$=t^2 or $t^2\mathbb{1}_{[|t|\leq K]}/2+K(|t|-K)\mathbb{1}_{[|t|>K]}$ for $t\in\mathbb{R}$ and $K>0$. Note that

$$\sqrt{n}(\hat{\overline{Y}}_{GREG}-E_{\mathbf{P}}(Y_i))=\sqrt{n}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta})\overline{\mathbf{V}}^T+\sqrt{n}\boldsymbol{\theta}(\overline{\mathbf{V}}-E_{\mathbf{P}}(\mathbf{V}_i))^T \text{ and}$$
$$\sqrt{n}(\hat{\overline{Y}}_{TLS}-E_{\mathbf{P}}(Y_i))=\sqrt{n}(\hat{\boldsymbol{\theta}}_n(K)-\boldsymbol{\theta})\overline{\mathbf{V}}^T+\sqrt{n}\boldsymbol{\theta}(\overline{\mathbf{V}}-E_{\mathbf{P}}(\mathbf{V}_i))^T$$

where $\mathbf{V}_i$=$(1,W_i)$, and $\hat{\boldsymbol{\theta}}_n$ and $\hat{\boldsymbol{\theta}}_n(K)$ are the estimators obtained from LS and TLS regression, respectively. Since $\{\epsilon_i\}_{i=1}^N$ in (5.4.2) are generated from some symmetric distribution with $E_{\mathbf{P}}(\epsilon_i)$=0, we have $E_{\mathbf{P}}\big(\psi(\epsilon_i)\big)$=0 for the above choices of ρ. Further, Assumptions 5.2.2–5.2.4 hold for these ρ's because ϵ_i's in (5.4.2) have a positive continuous density function. Then, it can be shown in the same way as the proof of the result in (5.8.9) that under $\mathbf{P}^*$, as $\nu\to\infty$,

$$\sqrt{n}(\hat{\overline{Y}}_{GREG}-E_{\mathbf{P}}(Y_i))\xrightarrow{\mathcal{L}}N(0,\Delta_1) \text{ and } \sqrt{n}(\hat{\overline{Y}}_{TLS}-E_{\mathbf{P}}(Y_i))\xrightarrow{\mathcal{L}}N(0,\Delta_2)$$

for $d(i,s)$=π_i^{-1} and SRSWOR, LMS and HEπPS sampling designs, and

$$\sqrt{n}(\hat{\overline{Y}}_{GREG}-E_{\mathbf{P}}(Y_i))\xrightarrow{\mathcal{L}}N(0,\Delta_1^*) \text{ and } \sqrt{n}(\hat{\overline{Y}}_{TLS}-E_{\mathbf{P}}(Y_i))\xrightarrow{\mathcal{L}}N(0,\Delta_2^*)$$

for $d(i,s)$=$G_iX_i^{-1}$ and RHC sampling design. Here, we have

$$\Delta_1 = \left(a\big(\Gamma_2/4 + \lambda\sigma_\epsilon^2 E_{\mathbf{P}}(\mathbf{V}_i^T\mathbf{V}_i)\big)a^T\right) + \lambda\boldsymbol{\theta} cov_{\mathbf{P}}(\mathbf{V}_i)\boldsymbol{\theta}^T$$
$$\Delta_2 = \left(a\big(\delta_\epsilon^2\Gamma_3 + \lambda\boldsymbol{\theta}_\epsilon^2\delta_\epsilon^2 E_{\mathbf{P}}(\mathbf{V}_i^T\mathbf{V}_i)\big)a^T\right) + \lambda\boldsymbol{\theta} cov_{\mathbf{P}}(\mathbf{V}_i)\boldsymbol{\theta}^T$$
$$\Delta_1^* = \left(a\big(\Gamma_2^*/4 + \lambda\sigma_\epsilon^2 E_{\mathbf{P}}(\mathbf{V}_i^T\mathbf{V}_i)\big)a^T\right) + \lambda\boldsymbol{\theta} cov_{\mathbf{P}}(\mathbf{V}_i)\boldsymbol{\theta}^T \text{ and}$$
$$\Delta_2^* = \left(a\big(\delta_\epsilon^2\Gamma_3^* + \lambda\boldsymbol{\theta}_\epsilon^2\delta_\epsilon^2 E_{\mathbf{P}}(\mathbf{V}_i^T\mathbf{V}_i)\big)a^T\right) + \lambda\boldsymbol{\theta} cov_{\mathbf{P}}(\mathbf{V}_i)\boldsymbol{\theta}^T,$$

where a is a $1\times(d+2)$ vector with first entry equals to 1 and other entries equal to 0, σ_ϵ^2=$E_{\mathbf{P}}(\epsilon_i)^2$, $\boldsymbol{\theta}_\epsilon^2$=$\big(K^2\mathbf{P}(|\epsilon_i|>K) + E_{\mathbf{P}}(\epsilon_i)^2\mathbb{1}_{[|\epsilon_i|\le K]}\big)$, δ_ϵ^2=$\big(\mathbf{P}(|\epsilon_i|\le K)\big)^{-2}$,

$$\Gamma_2 = \lim_{\nu\to\infty}(n/N^2)\sum_{i=1}^{N}(\mathbf{L}_{i,1}-\mathbf{T}_{L,1}\pi_i)^T(\mathbf{L}_{i,1}-\mathbf{T}_{L,1}\pi_i)(\pi_i^{-1}-1) \text{ and}$$
$$\Gamma_3 = \lim_{\nu\to\infty}(n/N^2)\sum_{i=1}^{N}(\mathbf{L}_{i,2}-\mathbf{T}_{L,2}\pi_i)^T(\mathbf{L}_{i,2}-\mathbf{T}_{L,2}\pi_i)(\pi_i^{-1}-1) \text{ a.s. } [\mathbf{P}], \text{ and}$$
$$\Gamma_2^* = cE_{\mathbf{P}}(X_i)E_{\mathbf{P}}\big(\mathbf{L}_{i,1}^T\mathbf{L}_{i,1}X_i^{-1}\big) = 4c\sigma_\epsilon^2 E_{\mathbf{P}}(X_i)E_{\mathbf{P}}\big(\mathbf{V}_i\mathbf{V}_iX_i^{-1}\big) \text{ and}$$
$$\Gamma_3^* = cE_{\mathbf{P}}(X_i)E_{\mathbf{P}}\big(\mathbf{L}_{i,2}^T\mathbf{L}_{i,2}X_i^{-1}\big) = c\boldsymbol{\theta}_\epsilon^2 E_{\mathbf{P}}(X_i)E_{\mathbf{P}}\big(\mathbf{V}_i^T\mathbf{V}_iX_i^{-1}\big).$$

Here, $\mathbf{L}_{i,1}$=$2\epsilon_i\mathbf{V}_i$ and $\mathbf{L}_{i,2}$=$\big(K\mathbb{1}_{[\epsilon_i>K]} - K\mathbb{1}_{[\epsilon_i<-K]} + \epsilon_i\mathbb{1}_{[|\epsilon_i|\le K]}\big)\mathbf{V}_i$ for i=$1,\ldots,N$, $\mathbf{T}_{L,k}$ =$\sum_{i=1}^N\mathbf{L}_{i,k}(1-\pi_i)/\sum_{i=1}^N\pi_i(1-\pi_i)$ for k=$1,2$, and c=$\lim_{\nu\to\infty}n\gamma$ (see Theorem 5.2.4). Moreover, it can be shown in the same way as the proof of Lemma 5.9.5 in Section 5.9 that

$$\Gamma_2 = \begin{cases} 4(1-\lambda)\sigma_\epsilon^2 E_{\mathbf{P}}(\mathbf{V}_i^T\mathbf{V}_i) \text{ under SRSWOR and LMS sampling designs, and} \\ 4\sigma_\epsilon^2 E_{\mathbf{P}}(\mathbf{V}_i^T\mathbf{V}_i)(E_{\mathbf{P}}(X_i)X_i^{-1}-\lambda) \text{ under any HE}\pi\text{PS sampling design,} \end{cases} \tag{5.8.20}$$

and

$$\Gamma_3 = \begin{cases} (1-\lambda)\boldsymbol{\theta}_\epsilon^2 E_{\mathbf{P}}(\mathbf{V}_i^T\mathbf{V}_i) \text{ under SRSWOR and LMS sampling designs, and} \\ \boldsymbol{\theta}_\epsilon^2 E_{\mathbf{P}}(\mathbf{V}_i^T\mathbf{V}_i)(E_{\mathbf{P}}(X_i)X_i^{-1}-\lambda) \text{ under any HE}\pi\text{PS sampling design.} \end{cases} \tag{5.8.21}$$

Let us next assume that $\rho(t)$=$|t| + (2p-1)t$ for $t\in\mathbb{R}$ and $p\in(0,1)$. Note that the linear model in (5.4.2) can be expressed as

$$Y_i = \boldsymbol{\theta}(p)\mathbf{V}_i^T + \epsilon_i(p) \text{ for } i = 1,\ldots,N,$$

where $\boldsymbol{\theta}(p)$=$\boldsymbol{\theta}$+$(Q_\epsilon(p), 0, \ldots, 0)$, $Q_\epsilon(p)$ is the superpopulation p^{th} quantile of ϵ_i's, and $\epsilon_i(p)$=$\epsilon_i-Q_\epsilon(p)$. Also, note that $E_{\mathbf{P}}\big(\psi(\epsilon_i(p))\big)$=$2E_{\mathbf{P}}\big(p - \mathbb{1}_{[\epsilon_i(p)\leq 0]}\big)$=0. Let us recall $(q_1, \ldots, q_{2l+1})$ from Section 5.4. Now, since ϵ_i's have a symmetric distribution about 0, we have

$$\sqrt{n}(\hat{\overline{Y}}_{QR} - E_{\mathbf{P}}(Y_i)) = \sqrt{n}\big(\hat{\boldsymbol{\theta}}_n(q_1) - \boldsymbol{\theta}(q_1), \ldots, \hat{\boldsymbol{\theta}}_n(q_{2l+1}) - \boldsymbol{\theta}(q_{2l+1})\big)H_1\overline{\mathbf{V}}^T +$$
$$\big(\boldsymbol{\theta}(q_1), \ldots, \boldsymbol{\theta}(q_{2l+1})\big)H_1\big(\overline{\mathbf{V}} - E_{\mathbf{P}}(\mathbf{V}_i)\big)^T,$$

where H_1 is as defined in the paragraph containing (5.4.1) in Section 5.4. Further, Assumptions 5.2.2–5.2.4 hold for the above-mentioned ρ because ϵ_i's have a positive continuous density function. Then, it can be shown in the same way as the proof of the result in (5.8.9) that under $\mathbf{P}^*$, as $\nu \to \infty$,

$$\sqrt{n}(\hat{\overline{Y}}_{QR} - E_{\mathbf{P}}(Y_i)) \xrightarrow{\mathcal{L}} N(0, \Delta_3)$$

for $d(i,s)$=π_i^{-1} and SRSWOR, LMS and HEπPS sampling designs, and

$$\sqrt{n}(\hat{\overline{Y}}_{QR} - E_{\mathbf{P}}(Y_i)) \xrightarrow{\mathcal{L}} N(0, \Delta_3^*)$$

for $d(i,s)$=$G_iX_i^{-1}$ and RHC sampling design. Here, we have

$$\Delta_3 = \bigg((\xi \otimes a)\big(\Gamma_4/4 + \lambda D \otimes E_{\mathbf{P}}(\mathbf{V}_i^T\mathbf{V}_i)\big)(\xi \otimes a)^T\bigg) + \lambda\boldsymbol{\theta} cov_{\mathbf{P}}(\mathbf{V}_i)\boldsymbol{\theta}^T \text{ and}$$
$$\Delta_3^* = \bigg((\xi \otimes a)\big(\Gamma_4^*/4 + \lambda D \otimes E_{\mathbf{P}}(\mathbf{V}_i^T\mathbf{V}_i)\big)(\xi \otimes a)^T\bigg) + \lambda\boldsymbol{\theta} cov_{\mathbf{P}}(\mathbf{V}_i)\boldsymbol{\theta}^T,$$

where D and ξ are as defined in Section 5.4. $\otimes$ denotes the Kronecker product, Γ_4=$\lim_{\nu\to\infty}(n/N^2)$ $\times\sum_{i=1}^N(\mathbf{L}_{i,3} - \mathbf{T}_{L,3}\pi_i)^T(\mathbf{L}_{i,3} - \mathbf{T}_{L,3}\pi_i)(\pi_i^{-1} - 1)$ *a.s.* [$\mathbf{P}$], $\mathbf{T}_{L,3}$=$\sum_{i=1}^N \mathbf{L}_{i,3}(1-\pi_i)/\sum_{i=1}^N \pi_i(1-\pi_i)$, and Γ_4^*=$cE_{\mathbf{P}}(X_i)E_{\mathbf{P}}\big(\mathbf{L}_{i,3}^T\mathbf{L}_{i,3}X_i^{-1}\big)$. Here,

$$\mathbf{L}_{i,3} = 2\bigg(\mathbf{V}_i(p_1 - \mathbb{1}_{[\epsilon_i(p_1)<0]}), \ldots, \mathbf{V}_i(p_l - \mathbb{1}_{[\epsilon_i(p_l)<0]}), \mathbf{V}_i(0.5 - \mathbb{1}_{[\epsilon_i(0.5)<0]}),$$
$$\mathbf{V}_i(1-p_1 - \mathbb{1}_{[\epsilon_i(1-p_1)<0]}), \ldots, \mathbf{V}_i(1-p_l - \mathbb{1}_{[\epsilon_i(1-p_l)<0]})\bigg)$$

for i=$1, \ldots, N$. Moreover, it can be shown in the same way as the proof of Lemma 5.9.5 in Section 5.9 that

$$\Gamma_4 = \begin{cases} (1-\lambda)E_{\mathbf{P}}(\mathbf{L}_{i,3}^T\mathbf{L}_{i,3}) \text{ under SRSWOR and LMS sampling designs, and} \\ E_{\mathbf{P}}(\mathbf{L}_{i,3}^T\mathbf{L}_{i,3})(E_{\mathbf{P}}(X_i)X_i^{-1} - \lambda) \text{ under any HE}\pi\text{PS sampling design.} \end{cases} \quad (5.8.22)$$

In view of (5.8.20), (5.8.21) and (5.8.22), it follows that

$$\Delta_1 - \Delta_3 = \begin{cases} \sigma_\epsilon^2 - \xi D \xi^T \text{ under SRSWOR and LMS sampling designs, and} \\ (\sigma_\epsilon^2 - \xi D \xi^T) E_{\mathbf{P}}(X_i) E_{\mathbf{P}}(X_i^{-1}) \text{ under any HE}\pi\text{PS sampling design,} \end{cases} \tag{5.8.23}$$

$$\Delta_1 - \Delta_2 = \begin{cases} \sigma_\epsilon^2 - \boldsymbol{\theta}_\epsilon^2 \delta_\epsilon^2 \text{ under SRSWOR and LMS sampling designs, and} \\ (\sigma_\epsilon^2 - \boldsymbol{\theta}_\epsilon^2 \delta_\epsilon^2) E_{\mathbf{P}}(X_i) E_{\mathbf{P}}(X_i^{-1}) \text{ under any HE}\pi\text{PS sampling design,} \end{cases} \tag{5.8.24}$$

and

$$\Delta_2 - \Delta_3 = \begin{cases} \boldsymbol{\theta}_\epsilon^2 \delta_\epsilon^2 - \xi D \xi^T \text{ under SRSWOR and LMS sampling designs, and} \\ (\boldsymbol{\theta}_\epsilon^2 \delta_\epsilon^2 - \xi D \xi^T) E_{\mathbf{P}}(X_i) E_{\mathbf{P}}(X_i^{-1}) \text{ under any HE}\pi\text{PS} \\ \text{sampling design.} \end{cases} \tag{5.8.25}$$

It also follows that

$$\Delta_1^* - \Delta_3^* = (\sigma_\epsilon^2 - \xi D \xi^T)(c E_{\mathbf{P}}(X_i) E_{\mathbf{P}}(X_i^{-1}) + \lambda) \text{ under RHC sampling design,}$$
$$\Delta_1^* - \Delta_2^* = (\sigma_\epsilon^2 - \boldsymbol{\theta}_\epsilon^2 \delta_\epsilon^2)(c E_{\mathbf{P}}(X_i) E_{\mathbf{P}}(X_i^{-1}) + \lambda) \text{ under RHC sampling design, and}$$
$$\Delta_2^* - \Delta_3^* = (\boldsymbol{\theta}_\epsilon^2 \delta_\epsilon^2 - \xi D \xi^T)(c E_{\mathbf{P}}(X_i) E_{\mathbf{P}}(X_i^{-1}) + \lambda) \text{ under RHC sampling design.}$$

Therefore, the conclusion of Theorem 5.4.1 holds. □

Proof of Theorem 5.4.2. It follows from the 1^{st} paragraph in the proof of Theorem 5.4.1 that the asymptotic distribution of $\sqrt{n}(\hat{\bar{Y}}_{TLS} - E_{\mathbf{P}}(Y_i))$ is the same under SRSWOR and LMS sampling designs. Further, it follows from the 1^{st} paragraph in the proof of Theorem 5.4.1 that the asymptotic variance of $\sqrt{n}(\hat{\bar{Y}}_{TLS} - E_{\mathbf{P}}(Y_i))$ under SRSWOR is smaller than its asymptotic variance under RHC as well as any HEπPS sampling design because $E_{\mathbf{P}}(X_i) E_{\mathbf{P}}(X_i)^{-1} > 1$ and $c \geq 1 - \lambda$ (see 2.7.5 in Section 2.7 of Chapter 2).

It follows from the 2^{nd} paragraph in the proof of Theorem 5.4.1 that the asymptotic distribution of $\sqrt{n}(\hat{\bar{Y}}_{QR} - E_{\mathbf{P}}(Y_i))$ is the same under SRSWOR and LMS sampling designs. Further, it follows from the 2^{nd} paragraph in the proof of Theorem 5.4.1 that the asymptotic variance of $\sqrt{n}(\hat{\bar{Y}}_{QR} - E_{\mathbf{P}}(Y_i))$ under SRSWOR is smaller than its asymptotic variance under RHC as well as any HEπPS sampling design. □

5.9. Proofs of additional results required to prove the main results

In this section, we state and prove some lemmas, which will be required to prove the theorems in this chapter. Let us first recall expressions for $M_n(\mathbf{u})$ and $L_n(\mathbf{u})$ from the paragraph preceding the proof of Theorem 5.2.1 in Section 5.8. Next, suppose that $P(s,\omega)$ denotes a high entropy sampling design satisfying Assumption 5.2.6, and $Q(s,\omega)$ denotes a rejective sampling design having inclusion probabilities equal to those of $P(s,\omega)$. Recall from the paragraph preceding the proof of Theorem 5.2.1 in Section 5.8 that such a rejective sampling design always exists. Now, we state the following lemma.

Lemma 5.9.1. *Suppose that Assumptions 5.2.1, 5.2.2 and 5.2.4 hold. Then, for any $K > 0$, under the probability distribution $\boldsymbol{P}^*$,*

$$\sup_{||\boldsymbol{u}||\leq K} ||L_n(\boldsymbol{u})|| = o_p(1) \text{ as } \nu \to \infty \tag{5.9.1}$$

for the rejective sampling design $Q(s,\omega)$.

Proof. We write the proof using similar arguments in the proof of Lemma 4.1 in [5]. Note that $L_n(\mathbf{u})$=$L_n^*(\mathbf{u}) + \tilde{L}_n(\mathbf{u})$, where

$$L_n^*(\mathbf{u}) = M_n(\mathbf{u}) - M_n(0) - (\tilde{M}_n(\mathbf{u}) - \tilde{M}_n(0)) \text{ with}$$
$$\tilde{M}_n(\mathbf{u}) = \sqrt{n}\sum_{i=1}^{N} \mathbf{V}_i\psi_1\big(\epsilon_i - \mathbf{u}\mathbf{V}_i^T/\sqrt{n}\big)/N, \text{ and}$$
$$\tilde{L}_n(\mathbf{u}) = \tilde{M}_n(\mathbf{u}) - \tilde{M}_n(0) - E_{\mathbf{P}}(\tilde{M}_n(\mathbf{u}) - \tilde{M}_n(0)).$$

Suppose that V_{ij} and u_j are the j^{th} components of $\mathbf{V}_i$ and $\mathbf{u}$, respectively, for j=$1,\ldots,d+2$. Further, suppose that

$$\mathbf{C} = \left\{\mathbf{u} \in \mathbb{R}^{d+2} : \max_{1\leq j\leq d+2} |u_j| \leq K\right\} \text{ for some } K > 0,$$

and $\{\mathbf{e}_j : 1 \leq j \leq d+2\}$ is the canonical basis of $\mathbb{R}^{d+2}$. Also, recall from the proof of Theorem 5.2.3 that

$$\tilde{H}_{ij} = \big(\psi(\epsilon_i - \mathbf{u}\mathbf{V}_i^T/\sqrt{n}) - \psi(\epsilon_i)\big)V_{ij}$$

for i=$1,\ldots,N$ and j=$1,\ldots,d+2$. Now, fix $\mathbf{u}\in\mathbf{C}$. Then, it follows from Theorem 6.1 in [40] that given any $\omega\in\Omega$ and j=$1,\ldots,d+2$, under $Q(s,\omega)$,

$$var\bigg(L_n^*(\mathbf{u})\mathbf{e}_j^T\bigg)=var\bigg(\sqrt{n}\sum_{i\in s}\pi_i^{-1}\tilde{H}_{ij}/N\bigg)= \\ (n/N^2)\bigg[\sum_{i=1}^N(\tilde{H}_{ij})^2(\pi_i^{-1}-1)-\bigg(\sum_{i=1}^N\tilde{H}_{ij}(1-\pi_i)\bigg)^2\bigg/\sum_{i=1}^N\pi_i(1-\pi_i)\bigg](1+e), \tag{5.9.2}$$

where $e\to0$ if $\sum_{i=1}^N\pi_i(1-\pi_i)\to\infty$ as $\nu\to\infty$. Note that $Q(s,\omega)$ satisfies Assumption 5.2.6–(i) since $P(s,\omega)$ and $Q(s,\omega)$ have same inclusion probabilities, and $P(s,\omega)$ satisfies Assumption 5.2.6–(i). Then, under $Q(s,\omega)$, $\sum_{i=1}^N\pi_i(1-\pi_i)\to\infty$ as $\nu\to\infty$ *a.s.* $[\mathbf{P}]$ by Assumption 5.2.1. Therefore, using (5.9.2), one can show that given any j=$1,\ldots,d+2$, under $Q(s,\omega)$,

$$var\bigg(\sqrt{n}\sum_{i\in s}\pi_i^{-1}\tilde{H}_{ij}/N\bigg)\leq(n/N^2)\sum_{i=1}^N(\tilde{H}_{ij})^2\pi_i^{-1}\leq K_1\sum_{i=1}^N(\tilde{H}_{ij})^2/N \tag{5.9.3}$$

for all sufficiently large ν and some constant $K_1>0$ (may depend on ω) *a.s.* $[\mathbf{P}]$. Next, there exists a constant K_2 such that $\max_{1\leq i\leq N}||\mathbf{V}_i||\leq K_2$ *a.s.* $[\mathbf{P}]$ by Assumption 5.2.4. Since, ψ is a non-decreasing function, we have

$$\sum_{i=1}^N E_{\mathbf{P}}(\tilde{H}_{ij})^2/N=E_{\mathbf{P}}\big\{\big(\psi(\epsilon_i-\mathbf{u}\mathbf{V}_i^T/\sqrt{n})-\psi(\epsilon_i)\big)^2V_{ij}^2\big\}\leq K_2^2\times \\ E_{\mathbf{P}}\big(\psi(\epsilon_i+KK_2\sqrt{d+2}/\sqrt{n})-\psi(\epsilon_i-KK_2\sqrt{d+2}/\sqrt{n})\big)^2\to0. \tag{5.9.4}$$

as $\nu\to\infty$ by Assumption 5.2.2. Hence, by Markov inequality, we have $\sum_{i=1}^N(\tilde{H}_{ij})^2/N\xrightarrow{p}0$ as $\nu\to\infty$ under $\mathbf{P}$ for any j=$1,\ldots,d+2$. This result and (5.9.3) imply that under $\mathbf{P}^*$,

$$var\bigg(\sqrt{n}\sum_{i\in s}\pi_i^{-1}\tilde{H}_{ij}/N\bigg)\xrightarrow{p}0\text{ as }\nu\to\infty \tag{5.9.5}$$

for the rejective sapling design $Q(s,\omega)$ and any j=$1,\ldots,d+2$. Suppose that

$$\mathcal{S}_j=\{s\in\mathcal{S}:\sqrt{n}|\sum_{i\in s}\pi_i^{-1}\tilde{H}_{ij}-\sum_{i=1}^N\tilde{H}_{ij}|/N>\delta\}$$

for any given $\delta>0$ and j=$1,\ldots,d+2$. Then, (5.9.5) implies that under $\mathbf{P}$,

$$\sum_{s\in\mathcal{S}_j} Q(s,\omega) \leq var\big(\sqrt{n}\sum_{i\in s}\pi_i^{-1}\tilde{H}_{ij}/N\big)/\delta^2 \xrightarrow{p} 0 \text{ as } \nu\to\infty$$

for $\delta > 0$ and j=$1,\ldots,d+2$. Since, $\sum_{s\in\mathcal{S}_j} Q(s,\omega)$ is bounded, we have $E_{\mathbf{P}}\big(\sum_{s\in\mathcal{S}_j} Q(s,\omega)\big)$= $\mathbf{P}^*\big\{\sqrt{n}|\sum_{i\in s}\pi_i^{-1}\tilde{H}_{ij} - \sum_{i=1}^N \tilde{H}_{ij}|/N > \delta\big\} \to 0$ as $\nu\to\infty$. In other words, under $\mathbf{P}^*$, as $\nu\to\infty$,

$$L_n^*(\mathbf{u})\mathbf{e}_j^T = \sqrt{n}\bigg(\sum_{i\in s}\pi_i^{-1}\tilde{H}_{ij} - \sum_{i=1}^N \tilde{H}_{ij}\bigg)/N \xrightarrow{p} 0 \tag{5.9.6}$$

for $Q(s,\omega)$ and any given j=$1,\ldots,d+2$. Next, recall $\tilde{L}_n(\mathbf{u})$ from the 1^{st} paragraph of this proof and note that

$$var_{\mathbf{P}}\bigg(\tilde{L}_n(\mathbf{u})\mathbf{e}_j^T\bigg) = var_{\mathbf{P}}\bigg(\sqrt{n}\sum_{i=1}^N \tilde{H}_{ij}/N\bigg) \leq (n/N)\sum_{i=1}^N E_{\mathbf{P}}(\tilde{H}_{ij})^2/N \to 0$$

as $\nu\to\infty$ under $\mathbf{P}$ by (5.9.4) and Assumption 5.2.1. Therefore, under $\mathbf{P}$, as $\nu\to\infty$, $\tilde{L}_n(\mathbf{u})\mathbf{e}_j^T \xrightarrow{p} 0$. Hence, under $\mathbf{P}^*$, as $\nu\to\infty$,

$$L_n(\mathbf{u})\mathbf{e}_j^T = L_n^*(\mathbf{u})\mathbf{e}_j^T + \tilde{L}_n(\mathbf{u})\mathbf{e}_j^T \xrightarrow{p} 0 \tag{5.9.7}$$

for $Q(s,\omega)$ and any given j=$1,\ldots,d+2$.

Now, we consider the cube

$$\mathbf{C}_a = \{\mathbf{u}\in\mathbb{R}^{d+2} : \max_{1\leq j\leq d+2}|u_j| \leq ([1/a]+1)aK\} \text{ for any given } a>0,$$

and decompose it into the cubes with vertices $(r_1aK,\ldots,r_{d+2}aK)$, where r_j=$0,\pm1,\ldots,$ $\pm([1/a]+1)$ for j=$1,\ldots,d+2$. Let $\mathcal{C}_a$ be the collection of all such cubes. Suppose that for any $\mathbf{C}_a^*\in\mathcal{C}_a$, $\mathbf{u}^*$ denotes the lowest vertex of $\mathbf{C}_a^*$. We say that a vertex v of any cube in $\mathbb{R}^{d+2}$ is its lowest vertex if $v_j\leq w_j$ for all j=$1\ldots,d+2$ and any other vertex w of that cube. Note that $\mathbf{u}^*\in\mathbf{C}_a$ for any given $\mathbf{C}_a^*\in\mathcal{C}_a$. Then, it follows in the same way as the derivation of the result in (5.9.7) that under $\mathbf{P}^*$, as $\nu\to\infty$,

$$\max_{\mathbf{C}_a^*\in\mathcal{C}_a}|L_n(\mathbf{u}^*)\mathbf{e}_j^T| \xrightarrow{p} 0 \tag{5.9.8}$$

for $Q(s,\omega)$ and any given j=$1,\ldots,d+2$. Next, note that

$$\sup_{\mathbf{u}\in\mathbf{C}}\left|L_n(\mathbf{u})\mathbf{e}_j^T\right| \leq \sup_{\mathbf{u}\in\mathbf{C}_a}\left|L_n(\mathbf{u})\mathbf{e}_j^T\right| \leq \max_{\mathbf{C}_a^*\in\mathcal{C}_a}\sup_{\mathbf{u}\in\mathbf{C}_a^*}\left|\big(L_n(\mathbf{u})-L_n(\mathbf{u}^*)\big)\mathbf{e}_j^T\right|+ \max_{\mathbf{C}_a^*\in\mathcal{C}_a}\left|L_n(\mathbf{u}^*)\mathbf{e}_j^T\right| \tag{5.9.9}$$

for any given j=1, . . . , $d+2$. Also, note that

$$\begin{aligned}\sup_{\mathbf{u}\in\mathbf{C}_a^*}\left|\big(L_n(\mathbf{u})-L_n(\mathbf{u}^*)\big)\mathbf{e}_j^T\right| \leq \sqrt{n}\Big\{\sum_{i\in s}\pi_i^{-1}\Big(\psi(\epsilon_i-\mathbf{u}^*\mathbf{V}_i^T/\sqrt{n}+\\ KaS_i/\sqrt{n})-\psi(\epsilon_i-\mathbf{u}^*\mathbf{V}_i^T/\sqrt{n}-KaS_i/\sqrt{n})\Big)|V_{ij}|\Big\}\Big/N+\\ \sqrt{n}E_{\mathbf{P}^*}\Big\{\sum_{i\in s}\pi_i^{-1}\Big(\psi(\epsilon_i-\mathbf{u}^*\mathbf{V}_i^T/\sqrt{n}+KaS_i/\sqrt{n})-\\ \psi(\epsilon_i-\mathbf{u}^*\mathbf{V}_i^T/\sqrt{n}-KaS_i/\sqrt{n})\Big)|V_{ij}|\Big\}\Big/N\end{aligned} \tag{5.9.10}$$

for any given $\mathbf{C}_a^*\in\mathcal{C}_a$ because ψ is a non-decreasing function. Here, S_i=$\sum_{j=1}^{d+2}|V_{ij}|$ for i=1, . . . , N. It can be shown in the same way as the derivation of the result in (5.9.7) that under $\mathbf{P}^*$, as $\nu\to\infty$,

$$\begin{aligned}\sqrt{n}\big\{\sum_{i\in s}\pi_i^{-1}\big(\psi(\epsilon_i-\mathbf{u}^*\mathbf{V}_i^T/\sqrt{n}+KaS_i/\sqrt{n})-\psi(\epsilon_i-\mathbf{u}^*\mathbf{V}_i^T/\sqrt{n}-\\ KaS_i/\sqrt{n})\big)|V_{ij}|\big\}/N-\sqrt{n}E_{\mathbf{P}^*}\big\{\sum_{i\in s}\pi_i^{-1}\big(\psi(\epsilon_i-\mathbf{u}^*\mathbf{V}_i^T/\sqrt{n}+KaS_i/\sqrt{n})-\\ \psi(\epsilon_i-\mathbf{u}^*\mathbf{V}_i^T/\sqrt{n}-KaS_i/\sqrt{n})\big)|V_{ij}|\big\}/N\xrightarrow{p}0\end{aligned}$$

for $Q(s,\omega)$ and any given $\mathbf{C}_a^*\in\mathcal{C}_a$. Now given any $\delta>0$, we have $KK_2a\sqrt{d+2}/\sqrt{n}\leq\delta$ for all sufficiently large ν. Then, it follows from Assumptions 2 and 4 that as $\nu\to\infty$

$$\begin{aligned}&\sqrt{n}E_{\mathbf{P}^*}\Big\{\sum_{i\in s}\pi_i^{-1}\Big(\psi(\epsilon_i-\mathbf{u}^*\mathbf{V}_i^T/\sqrt{n}+KaS_i/\sqrt{n})-\\ &\psi(\epsilon_i-\mathbf{u}^*\mathbf{V}_i^T/\sqrt{n}-KaS_i/\sqrt{n})\Big)|V_{ij}|\Big\}\Big/N\\ &=\sqrt{n}\sum_{i=1}^{N}E_{\mathbf{P}}\Big\{\Big(\psi(\epsilon_i-\mathbf{u}^*\mathbf{V}_i^T/\sqrt{n}+KaS_i/\sqrt{n})-\\ &\psi(\epsilon_i-\mathbf{u}^*\mathbf{V}_i^T/\sqrt{n}-KaS_i/\sqrt{n})\Big)|V_{ij}|\Big\}\Big/N\\ &\leq K_2\sqrt{n}E_{\mathbf{P}}\Big(\psi(\epsilon_i-\mathbf{u}^*\mathbf{V}_i^T/\sqrt{n}+KK_2a\sqrt{d+2}/\sqrt{n})-\\ &\psi(\epsilon_i-\mathbf{u}^*\mathbf{V}_i^T/\sqrt{n}-KK_2a\sqrt{d+2}/\sqrt{n})\Big)\end{aligned} \tag{5.9.11}$$

$$\leq KK_2^2 a\sqrt{d+2}\sup\Big\{E_{\mathbf{P}}\Big(\psi(\epsilon_i-\mathbf{u}^*\mathbf{V}_i^T/\sqrt{n}+h)-$$
$$\psi(\epsilon_i-\mathbf{u}^*\mathbf{V}_i^T/\sqrt{n}-h)\Big)\Big/h: 0<h\leq\delta\Big\}=aO(1)$$

under $Q(s,\omega)$ for any given $\mathbf{C}_a^*\in\mathcal{C}_a$. Therefore, it follows from (5.9.10) that under $\mathbf{P}^*$, as $\nu\to\infty$,

$$\sup_{\mathbf{u}\in\mathbf{C}_a^*}\big|\big(L_n(\mathbf{u})-L_n(\mathbf{u}^*)\big)\mathbf{e}_j^T\big|=aO_p(1)$$

for $Q(s,\omega)$ and any given $\mathbf{C}_a^*\in\mathcal{C}_a$. Hence, using (5.9.8) and (5.9.9), one can show that under $\mathbf{P}^*$, as $\nu\to\infty$, $\sup_{\mathbf{u}\in\mathbf{C}}\big|L_n(\mathbf{u})\mathbf{e}_j^T\big|=aO_p(1)$ for $Q(s,\omega)$, and any given $a>0$ and j=$1,\ldots,d+2$. On taking $a\to 0$, we obtain that under $\mathbf{P}^*$, as $\nu\to\infty$,

$$\sup_{||\mathbf{u}||\leq K}\big|L_n(\mathbf{u})\mathbf{e}_j^T\big|\leq\sup_{\mathbf{u}\in\mathbf{C}}\big|L_n(\mathbf{u})\mathbf{e}_j^T\big|=o_p(1)$$

for $Q(s,\omega)$ and any given j=$1,\ldots,d+2$. Then the proof of the result in (5.9.1) follows in a straight-forward way. □

Next, suppose that $\hat{\overline{\mathbf{H}}}_1$=$\sum_{i\in s}(N\pi_i)^{-1}\mathbf{H}_i$ and $\overline{\mathbf{H}}$=$\sum_{i=1}^N\mathbf{H}_i/N$, where $\mathbf{H}_i$=$\psi(\epsilon_i)\mathbf{V}_i$ for i= $1,\ldots,N$. Then, we state the following lemma.

Lemma 5.9.2. *Fix $\boldsymbol{m}\in\mathbb{R}^{d+2}$ such that $\boldsymbol{m}\neq 0$. Suppose that Assumption 5.2.1 holds. Then, under $Q(s,\omega)$, we have $\sqrt{n}\boldsymbol{m}\big(\hat{\overline{\boldsymbol{H}}}_1-\overline{\boldsymbol{H}}\big)^T\xrightarrow{\mathcal{L}}N(0,\boldsymbol{m}\Gamma\boldsymbol{m}^T)$ as $\nu\to\infty$ a.s. $[\boldsymbol{P}]$, where Γ is as mentioned in Assumption 5.2.6-(ii).*

Proof. The proof follows exactly the same way as the derivation of the result, which appears in the proof of Lemma 2.7.2 in Section 2.7 of Chapter 2, that $\sqrt{n}\mathbf{m}_1(\hat{\overline{\mathbf{V}}}_1-\overline{\mathbf{V}})^T\xrightarrow{\mathcal{L}}N(0,\mathbf{m}_1\Gamma\mathbf{m}_1^T)$ as $\nu\to\infty$ under each of SRSWOR, LMS and any HEπPS sampling designs for any $\mathbf{m}_1\in\mathbb{R}^p$, $\mathbf{m}_1\neq 0$ and Γ=$\lim_{\nu\to\infty}\Sigma$. □

Lemma 5.9.3. *Suppose that Assumptions 5.2.1–5.2.4 hold. Then, given any $\delta>0$, there exist ζ_1, ζ_2 and ν_0 such that*

$$\boldsymbol{P}^*\Big\{\inf_{||\boldsymbol{u}||\geq\zeta_2}||M_n(\boldsymbol{u})||<\zeta_1\Big\}<\delta \text{ for all } \nu\geq\nu_0 \tag{5.9.12}$$

and the rejective sampling design $Q(s,\omega)$.

Proof. Recall ϕ from (5.2.3) in Section 5.2. Then, we note that under $Q(s,\omega)$,

$$E_{\mathbf{P}^*}\big(M_n(\mathbf{u}) - M_n(0)\big) = \sqrt{n}\sum_{i=1}^{N} E_{\mathbf{P}}\Big\{\big(\psi(\epsilon_i - \mathbf{u}\mathbf{V}_i^T/\sqrt{n}) - \psi(\epsilon_i)\big)\mathbf{V}_i\Big\}\Big/ N$$
$$= \sqrt{n}E_{\mathbf{P}}\Big\{\big(\phi(\mathbf{u}\mathbf{V}_1^T/\sqrt{n}, W_1) - \phi(0, W_1)\big)\mathbf{V}_1\Big\} = E_{\mathbf{P}}\big\{\phi'(\xi_1, W_1)\mathbf{u}\mathbf{V}_1^T\mathbf{V}_1\big\} \tag{5.9.13}$$

by Taylor expansion and Assumption 5.2.3. Here, ξ_1 lies between 0 and $\mathbf{u}\mathbf{V}_1^T/\sqrt{n}$. This implies that

$$|\xi_1| \leq |\mathbf{u}\mathbf{V}_1^T|/\sqrt{n} \leq ||\mathbf{u}||||\mathbf{V}_1||/\sqrt{n}.$$

Now, if we fix any $K > 0$, then $|\xi_1| \to 0$ uniformly over $\{\mathbf{u} \in \mathbb{R}^{d+2} : ||\mathbf{u}|| \leq K\}$ as $\nu \to \infty$ *a.s.* [**P**] by Assumption 5.2.4. By Assumption 5.2.3, $\phi'(t, W_1)$ is continuous, and hence uniformly continuous on $[-\delta_1, \delta_1]$ for any given $\omega \in \Omega$ and any $\delta_1 > 0$. Therefore,

$$\sup_{||\mathbf{u}|| \leq K} |\phi'(\xi_1, W_1) - \phi'(0, W_1)| \to 0 \text{ as } \nu \to \infty \textit{ a.s. } [\mathbf{P}].$$

Moreover, for any $\nu \geq 1$,

$$\sup_{||\mathbf{u}|| \leq K} |\phi'(\xi_1, W_1) - \phi'(0, W_1)| \leq 2\sup_{t\in\mathbb{R}} |\phi'(t, W_1)| \text{ and } E_{\mathbf{P}}\big(\sup_{t\in\mathbb{R}} |\phi'(t, W_1)|\big)^2 < \infty$$

by Assumption 5.2.3. Hence,

$$\sup_{||\mathbf{u}|| \leq K} ||E_{\mathbf{P}}\big\{\phi'(\xi_1, W_1)\mathbf{u}\mathbf{V}_1^T\mathbf{V}_1\big\} - E_{\mathbf{P}}\big\{\phi'(0, W_1)\mathbf{u}\mathbf{V}_1^T\mathbf{V}_1\big\}|| \to 0 \text{ as } \nu \to \infty$$

by Assumption 5.2.4 and DCT. Thus $\sup_{||\mathbf{u}|| \leq K} ||E_{\mathbf{P}^*}\big(M_n(\mathbf{u}) - M_n(0)\big) + \mathbf{u}\Sigma|| \to 0$ as $\nu \to \infty$ by Assumption 5.2.4. This result and Lemma 5.9.1 imply that under $\mathbf{P}^*$, as $\nu \to \infty$,

$$\sup_{||\mathbf{u}|| \leq K} ||M_n(\mathbf{u}) - M_n(0) + \mathbf{u}\Sigma|| = o_p(1) \tag{5.9.14}$$

for $Q(s, \omega)$ and any $K > 0$.

Next, it follows from Lemma 5.9.2 that for any given j=$1, \ldots, d+2$, under $Q(s, \omega)$, as $\nu \to \infty$,

$$\sqrt{n}\mathbf{e}_j\Bigg[\sum_{i\in s}\pi_i^{-1}\psi(\epsilon_i)\mathbf{V}_i/N - \sum_{i=1}^{N}\psi(\epsilon_i)\mathbf{V}_i/N\Bigg]^T \xrightarrow{\mathcal{L}} N(0, \mathbf{e}_j\Gamma\mathbf{e}_j^T) \tag{5.9.15}$$

a.s. [**P**], where $\{\mathbf{e}_j : 1 \leq j \leq d+2\}$ are canonical basis vectors of $\mathbb{R}^{d+2}$. Then, using DCT, one can show that for any given j=$1, \ldots, d+2$, under $\mathbf{P}^*$, as $\nu \to \infty$,

$$\begin{aligned}&\sqrt{n}\mathbf{e}_j\Big[\sum_{i\in s}\pi_i^{-1}\psi(\epsilon_i)\mathbf{V}_i/N - \sum_{i=1}^{N}\psi(\epsilon_i)\mathbf{V}_i/N\Big]^T = \\ &\sqrt{n}\Big[\sum_{i\in s}\pi_i^{-1}\psi(\epsilon_i)V_{ij}/N - \sum_{i=1}^{N}\psi(\epsilon_i)V_{ij}/N\Big] = O_p(1)\end{aligned} \tag{5.9.16}$$

for $Q(s, \omega)$, where V_{ij} is the j^{th} component of $\mathbf{V}_i$. Moreover, we have

$$var_{\mathbf{P}}\Big(\sqrt{n}\sum_{i=1}^{N}\psi(\epsilon_i)V_{ij}/N\Big) \leq (n/N)E_{\mathbf{P}}\big(\psi(\epsilon_i)V_{ij}\big)^2 = O(1) \tag{5.9.17}$$

as $\nu \to \infty$ for any j=$1, \ldots, d+2$ by Assumptions 5.2.2 and 5.2.4. One can also show that $E_{\mathbf{P}}\big(\psi(\epsilon_i)\mathbf{V}_i\big)$=$E_{\mathbf{P}}\big(\psi(Y_i - \boldsymbol{\theta}\mathbf{V}_i^T)\mathbf{V}_i\big)$=0 because (Y_i, Z_i, X_i) have absolutely continuous distribution and $\rho(t)$ is differentiable at all but at most countably many t. Therefore, under $\mathbf{P}^*$, as $\nu \to \infty$,

$$\sqrt{n}\sum_{i\in s}\pi_i^{-1}\psi(\epsilon_i)V_{ij}/N = O_p(1) \text{ for any } j = 1, \ldots, d+2, \text{ and hence } ||M_n(0)|| = O_p(1)$$

for $Q(s, \omega)$. This implies that given any δ, there exist $\nu_0 \in \mathbb{N}$ and $K_1 > 0$ such that

$$\mathbf{P}^*\big\{||M_n(0)|| > K_1\big\} < \delta/2 \text{ for all } \nu \geq \nu_0. \tag{5.9.18}$$

Now, suppose that λ_1 is the minimum eigenvalue of Σ. Let us choose $\zeta_1 > 0$ and $\zeta_2 > 0$ such that $\zeta_2 > 2K_1/\lambda_1$ and $\zeta_1 < K_1/2$. Further, suppose that ζ_3=$\zeta_1\zeta_2$. Then, we have

$$\begin{aligned}&\mathbf{P}^*\Big\{\inf_{||\mathbf{u}||=\zeta_2}\big(-M_n(\mathbf{u})\mathbf{u}^T\big) < \zeta_3\Big\} \leq \mathbf{P}^*\Big\{\inf_{||\mathbf{u}||=\zeta_2}\big(-M_n(\mathbf{u})\mathbf{u}^T\big) < \zeta_3, \\ &\inf_{||\mathbf{u}||=\zeta_2}\big(-M_n(0)\mathbf{u}^T + \mathbf{u}\Sigma\mathbf{u}^T\big) \geq 2\zeta_3\Big\} + \mathbf{P}^*\Big\{\inf_{||\mathbf{u}||=\zeta_2}\big(-M_n(0)\mathbf{u}^T + \\ &\mathbf{u}\Sigma\mathbf{u}^T\big) < 2\zeta_3\Big\}.\end{aligned} \tag{5.9.19}$$

Further, we have

$$\begin{aligned}&\mathbf{P}^*\Big\{\inf_{||\mathbf{u}||=\zeta_2}\big(-M_n(\mathbf{u})\mathbf{u}^T\big) < \zeta_3, \inf_{||\mathbf{u}||=\zeta_2}\big(-M_n(0)\mathbf{u}^T + \mathbf{u}\Sigma\mathbf{u}^T\big) \geq 2\zeta_3\Big\} \\ &\leq \mathbf{P}^*\Big\{\sup_{||\mathbf{u}||=\zeta_2}\big((M_n(\mathbf{u}) - M_n(0))\mathbf{u}^T + \mathbf{u}\Sigma\mathbf{u}^T\big) \geq \zeta_3\Big\} \leq\end{aligned} \tag{5.9.20}$$

$$\mathbf{P}^*\bigg\{\sup_{||\mathbf{u}||=\zeta_2} ||(M_n(\mathbf{u}) - M_n(0) + \mathbf{u}\Sigma|| \geq \zeta_1\bigg\} \to 0$$

as $\nu \to \infty$ by (5.9.14). Next, it follows that

$$\begin{aligned} &\mathbf{P}^*\bigg\{\inf_{||\mathbf{u}||=\zeta_2} \big(-M_n(0)\mathbf{u}^T + \mathbf{u}\Sigma\mathbf{u}^T\big) < 2\zeta_3\bigg\} \leq \mathbf{P}^*\bigg\{\inf_{||\mathbf{u}||=\zeta_2} \big(-M_n(0)\mathbf{u}^T\big) \\ &+ \zeta_2^2\lambda_1 < 2\zeta_3\bigg\} \leq \mathbf{P}^*\bigg\{-\zeta_2||M_n(0)|| < 2\zeta_3 - \zeta_2^2\lambda_1\bigg\} \leq \\ &\mathbf{P}^*\bigg\{||M_n(0)|| > K_1\bigg\} < \delta/2, \end{aligned} \tag{5.9.21}$$

for all $\nu \geq \nu_0$ by (5.9.18). Thus, one can choose ν_0 large enough such that

$$\mathbf{P}^*\bigg\{\inf_{||\mathbf{u}||=\zeta_2} \big(-M_n(\mathbf{u})\mathbf{u}^T\big) < \zeta_3\bigg\} < \delta \tag{5.9.22}$$

for all $\nu \geq \nu_0$ by (5.9.19), (5.9.20) and (5.9.21). Next, note that

$$-M_n(\tau\mathbf{u}_1)\mathbf{u}_1^T \geq -M_n(\mathbf{u}_1)\mathbf{u}_1^T \tag{5.9.23}$$

for any given $\tau \geq 1$ and $\mathbf{u}_1 \in \mathbb{R}^{d+2}$. Now, if $||\mathbf{u}|| \geq \zeta_2$ and $\mathbf{u}_1$=$\zeta_2\mathbf{u}/||\mathbf{u}||$, then $||\mathbf{u}_1||$=ζ_2 and $\mathbf{u}$=$\tau\mathbf{u}_1$ with τ=$||\mathbf{u}||/\zeta_2 \geq 1$. Then, using (5.9.22) and (5.9.23), one can show that

$$\begin{aligned} &\mathbf{P}^*\bigg\{\inf_{||\mathbf{u}||\geq\zeta_2} ||M_n(\mathbf{u})|| < \zeta_1\bigg\} \leq \mathbf{P}^*\bigg\{\inf_{||\mathbf{u}||\geq\zeta_2} \big(-M_n(\mathbf{u})\mathbf{u}^T\big)\zeta_2/||\mathbf{u}|| < \\ &\zeta_1\zeta_2\bigg\} \leq \mathbf{P}^*\bigg\{\inf_{||\mathbf{u}_1||=\zeta_2} \big(-M_n(\mathbf{u}_1)\mathbf{u}_1^T\big) < \zeta_3\bigg\} < \delta \end{aligned} \tag{5.9.24}$$

for all $\nu \geq \nu_0$. Hence, the result in (5.9.12) holds. □

Next, suppose that $\hat{\overline{\mathbf{H}}}_2$=$\sum_{i\in s}(NX_i)^{-1}G_i\mathbf{H}_i$, where $\mathbf{H}_i$=$\psi(\epsilon_i)\mathbf{V}_i$ for i=$1,\ldots,N$, and G_i's are as in the paragraph containing (5.1.1) and (5.1.2) in Section 5.1. Recall from the paragraph preceding Lemma 5.9.2 that $\overline{\mathbf{H}}$=$\sum_{i=1}^N \mathbf{H}_i/N$. Also, recall from the statement of Theorem 5.2.4 that γ=$\sum_{r=1}^n \tilde{N}_r(\tilde{N}_r-1)/N(N-1)$ with $\tilde{N}_r$ being the size of the i^{th} group formed randomly in RHC sampling design. Now, we state the following lemma.

Lemma 5.9.4. *Fix* $\boldsymbol{m} \in \mathbb{R}^{d+2}$ *such that* $\boldsymbol{m} \neq 0$. *Suppose that Assumptions 5.2.1 and 5.2.7–5.2.9 hold. Then, under RHC sampling design, we have* $\sqrt{n}\boldsymbol{m}\big(\hat{\overline{\boldsymbol{H}}}_2 - \overline{\boldsymbol{H}}\big)^T \xrightarrow{\mathcal{L}} N(0, c\boldsymbol{m}\Gamma^*\boldsymbol{m}^T)$ *as* $\nu \to \infty$ *a.s.* $[\boldsymbol{P}]$, *where* Γ^* *is as mentioned in Assumption 5.2.9 and* c=$\lim_{\nu\to\infty} n\gamma$.

Note that the limit $\lim_{\nu\to\infty} n\gamma$ exists by Lemma 2.7.5 in Section 2.7 of Chapter 2.

Proof. The proof follows exactly the same way as the derivation of the result, which appears in the proof of Lemma 2.7.2 in Section 2.7 of Chapter 2, that $\sqrt{n}\mathbf{m}_1(\hat{\overline{\mathbf{V}}}_2 - \overline{\mathbf{V}})^T \xrightarrow{\mathcal{L}} N(0, \mathbf{m}_1\Gamma_2\mathbf{m}_1^T)$ as $\nu \to \infty$ under RHC sampling design for any $\mathbf{m}_1 \in \mathbb{R}^p$, $\mathbf{m}_1 \neq 0$ and Γ_2=$\lim_{\nu\to\infty} \Sigma_2$. □

Next, we show that Assumption 5.2.6–*(ii)* holds under SRSWOR, LMS and any πPS sampling designs. Recall ψ from (5.2.1), and ϵ_i from (5.2.3). Also, recall from the paragraph preceding Assumption 5.2.6 that $\mathbf{H}_i$=$\psi(\epsilon_i)\mathbf{V}_i$ for i=$1, \ldots, N$, and $\mathbf{T}_H$=$\sum_{i=1}^N \mathbf{H}_i(1-\pi_i)/\sum_{i=1}^N \pi_i(1-\pi_i)$. Here, $\mathbf{V}_i$=$(1, W_i)$. Now, we state the following lemma.

Lemma 5.9.5. *Suppose that Assumptions 5.2.1 and 5.2.8 hold, and $E_{\boldsymbol{P}}||\boldsymbol{H}_i||^2 < \infty$. Then, Assumption 5.2.6–(ii) holds under SRSWOR and LMS sampling designs. Moreover, if $X_i \leq K$ a.s. $[\boldsymbol{P}]$ for some $0 < K < \infty$, $E_{\boldsymbol{P}}(X_i)^{-2} < \infty$, and Assumption 5.2.1 holds with $0 \leq \lambda < E_{\boldsymbol{P}}(X_i)/K$, then Assumption 5.2.6–(ii) holds under any πPS sampling design.*

Proof. Let us denote $(1/N^2)\sum_{i=1}^N(\mathbf{H}_i - \mathbf{T}_H\pi_i)^T(\mathbf{H}_i - \mathbf{T}_H\pi_i)(\pi_i^{-1} - 1)$ by Σ_N. Here, π_i's are inclusion probabilities. Note that

$$n\Sigma_N = (1 - n/N)\bigg(\sum_{i=1}^N \mathbf{H}_i^T\mathbf{H}_i/N - \overline{\mathbf{H}}^T\overline{\mathbf{H}}\bigg) \text{ with } \overline{\mathbf{H}} = \sum_{i=1}^N \mathbf{H}_i/N$$

under SRSWOR. Then,

$$n\Sigma_N \to (1-\lambda)E_{\mathbf{P}}(\mathbf{H}_i - E_{\mathbf{P}}(\mathbf{H}_i))(\mathbf{H}_i - E_{\mathbf{P}}(\mathbf{H}_i)) \text{ as } \nu \to \infty \text{ a.s. } [\mathbf{P}] \tag{5.9.25}$$

by Assumption 5.2.1 and SLLN. Note that $E_{\mathbf{P}}(\mathbf{H}_i - E_{\mathbf{P}}(\mathbf{H}_i))(\mathbf{H}_i - E_{\mathbf{P}}(\mathbf{H}_i))$ is p.d. because $\{(Y_i, W_i) : 1 \leq i \leq N\}$ have absolutely continuous distribution. Thus Assumption 5.2.6–*(ii)* holds under SRSWOR.

Next, suppose that $\Sigma_N^{(1)}$ and $\Sigma_N^{(2)}$ denote $(1/N^2)\sum_{i=1}^N(\mathbf{H}_i - \mathbf{T}_H\pi_i)^T(\mathbf{H}_i - \mathbf{T}_H\pi_i)(\pi_i^{-1} - 1)$ under LMS sampling design and SRSWOR, respectively, and $\{\pi_i^{(1)}\}_{i=1}^N$ denote inclusion probabilities of LMS sampling design. Then, we have

$$\begin{aligned} &\pi_i^{(1)} = (n-1)/(N-1) + \bigg(X_i \Big/ \sum_{i=1}^N X_i\bigg)((N-n)/(N-1)) \text{ and} \\ &\pi_i^{(1)} - n/N = -(N-n)(N(N-1))^{-1}(X_i/\overline{X} - 1). \end{aligned} \tag{5.9.26}$$

Further,

$$\frac{|\pi_i^{(1)} - n/N|}{n/N} = \frac{N-n}{n(N-1)} \left| \frac{X_i}{\overline{X}} - 1 \right| \leq \frac{N-n}{n(N-1)} \left(\frac{\max_{1 \leq i \leq N} X_i}{\min_{1 \leq i \leq N} X_i} + 1 \right).$$

Therefore,

$$\max_{1 \leq i \leq N} |N\pi_i^{(1)}/n - 1| \to 0 \text{ as } \nu \to \infty \textit{ a.s. } [\mathbf{P}] \tag{5.9.27}$$

by Assumption 5.2.8. Now, it can be shown using the result in (5.9.27) that $n(\Sigma_N^{(1)} - \Sigma_N^{(2)}) \to 0$ as $\nu \to \infty$ *a.s.* [$\mathbf{P}$]. Therefore, Assumption 5.2.6–*(ii)* holds under LMS sampling design in view of (5.9.25).

Next, under any πPS sampling design, we have

$$\begin{aligned} \lim_{\nu \to \infty} n\Sigma_N = E_{\mathbf{P}}\big[&\big\{ \mathbf{H}_i + \chi^{-1}\mu_x^{-1}(\lambda E_{\mathbf{P}}(\mathbf{H}_i X_i) - E_{\mathbf{P}}(\mathbf{H}_i)\mu_x)X_i \big\}^T \times \\ &\big\{ \mathbf{H}_i + \chi^{-1}\mu_x^{-1}(\lambda E_{\mathbf{P}}(\mathbf{H}_i X_i) - E_{\mathbf{P}}(\mathbf{H}_i)\mu_x)X_i \big\} \big\{ \mu_x/X_i - \lambda \big\} \big] \textit{ a.s. } [\mathbf{P}] \end{aligned} \tag{5.9.28}$$

by Assumption 5.2.1 and SLLN. Here, μ_x=$E_{\mathbf{P}}(X_i)$ and χ=$\mu_x - \lambda(E_{\mathbf{P}}(X_i)^2/\mu_x)$. The matrix on the right hand side of (5.9.28) is p.d. because $X_i \leq K$ *a.s.* [$\mathbf{P}$] for some $0 < K < \infty$, Assumption 5.2.1 holds with $0 < \lambda < E_{\mathbf{P}}(X_i)/K$, and $\{(Y_i, W_i) : 1 \leq i \leq N\}$ have absolutely continuous distribution. Thus Assumption 5.2.6–*(ii)* holds under any πPS sampling design. This completes the proof of the lemma. □

Bibliography

[1] M. Avadhani and A. Srivastava. A comparison of midzuno-sen scheme with pps sampling without replacement and its applicaton to successive sampling. *Annals of the Institute of Statistical Mathematics*, 24(1):153–164, 1972.

[2] M. Avadhani and B. Sukhatme. A comparison of two sampling procedures with an application to successive sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 19(3):251–259, 1970.

[3] V. Barnett. Outliers in sample surveys. *Journal of applied statistics*, 21(5):373–381, 1994.

[4] Y. G. Berger. Rate of convergence to normal distribution for the Horvitz-Thompson estimator. *J. Statist. Plann. Inference*, 67(2):209–226, 1998.

[5] P. J. Bickel. One-step huber estimates in the linear model. *Journal of the American Statistical Association*, 70(350):428–434, 1975.

[6] P. Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.

[7] H. Boistard, H. P. Lopuhaä, and A. Ruiz-Gazen. Functional central limit theorems for single-stage sampling designs. *Ann. Stat.*, 45(4):1728–1758, 2017. ISSN 0090-5364. doi: 10.1214/16-AOS1507. URL https://doi.org/10.1214/16-AOS1507.

[8] H. Boistard, H. P. Lopuhaä, and A. Ruiz-Gazen. Supplement to "functional central limit theorems for single-stage samplings designs". *Ann. Stat.*, 2017.

[9] L. Bondesson, I. Traat, and A. Lundqvist. Pareto sampling versus sampford and conditional poisson sampling. *Scand. J. Stat.*, 33(4):699–720, 2006.

[10] N. E. Breslow and J. A. Wellner. Weighted likelihood for semiparametric models and two-phase stratified samples, with application to cox regression. *Scandinavian Journal of Statistics*, 34(1):86–102, 2007.

[11] T. S. Breusch and A. R. Pagan. A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the econometric society*, pages 1287–1294, 1979.

[12] H. Cardot and E. Josserand. Horvitz–thompson estimators for functional data: Asymptotic confidence bands and optimal allocation for stratified sampling. *Biometrika*, 98(1):107–118, 2011.

[13] H. Cardot, D. Degras, and E. Josserand. Confidence bands for horvitz–thompson estimators using sampled noisy functional data. *Bernoulli*, 19(5A):2067–2097, 2013.

[14] H. Cardot, A. Dessertaine, C. Goga, É. Josserand, and P. Lardin. Comparison of different sample designs and construction of confidence bands to estimate the mean of functional data: An illustration on electricity consumption. *Surv. Methodol.*, 39(2):283–301, 2013.

[15] H. Cardot, C. Goga, and P. Lardin. Uniform convergence and asymptotic confidence bands for model-assisted estimators of the mean of sampled functional data. *Electron. J. Stat.*, 7: 562–596, 2013.

[16] H. Cardot, C. Goga, and P. Lardin. Variance estimation and asymptotic confidence bands for the mean estimator of sampled functional data with high entropy unequal probability sampling designs. *Scand. J. Stat.*, 41(2):516–534, 2014.

[17] C. M. Cassel, C. E. Särndal, and J. H. Wretman. Some results on generalized difference estimation and generalized regression estimation for finite populations. *Biometrika*, 63(3): 615–620, 1976.

[18] A. Chatterjee. Asymptotic properties of sample quantiles from a finite population. *Ann. Inst. Statist. Math.*, 63(1):157–179, 2011. ISSN 0020-3157. doi: 10.1007/s10463-008-0210-4. URL `https://doi.org/10.1007/s10463-008-0210-4`.

[19] A. Chaudhuri. *Modern survey sampling*. Chapman and Hall/CRC, 2014.

[20] A. Chaudhuri, K. Dihidar, and M. Bose. On the feasibility of basing horvitz and thompson's estimator on a sample by rao, hartley, and cochran's scheme. *Commun. Stat. - Theory Methods*, 35(12):2239–2244, 2006.

[21] P. Chaudhuri, K. Doksum, and A. Samarov. On average derivative quantile regression. *Ann. Statist.*, 25(2):715–744, 1997. ISSN 0090-5364. doi: 10.1214/aos/1031833670. URL `https://doi.org/10.1214/aos/1031833670`.

[22] J. Chen and R. R. Sitter. A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys. *Statist. Sinica*, 9(2):385–406, 1999. ISSN 1017-0405.

[23] S. Chen and Y. D. Zhao. Quantile regression analysis of survey data under informative sampling. *Journal of Survey Statistics and Methodology*, 7:157–174, 2019.

[24] W. G. Cochran. *Sampling techniques*. John Wiley & Sons, New York-London-Sydney, third edition, 1977. Wiley Series in Probability and Mathematical Statistics.

[25] P. L. Conti. On the estimation of the distribution function of a finite population under high entropy sampling designs, with applications. *Sankhya B*, 76(2):234–259, 2014.

[26] P. L. Conti and D. Marella. Inference for quantiles of a finite population: asymptotic versus resampling results. *Scandinavian Journal of Statistics*, 42(2):545–561, 2015.

[27] M. del Mar Rueda, A. Arcos, and M. D. Martínez. Difference estimators of quantiles in finite populations. *Test*, 12(2):481–496, 2003.

[28] J.-C. Deville and C.-E. Särndal. Calibration estimators in survey sampling. *Journal of the American statistical Association*, 87(418):376–382, 1992.

[29] A. Dey and P. Chaudhuri. A comparison of estimators of mean and its functions in finite populations. To appear in *Statistica Sinica*, 2024. doi: 10.5705/ss.202022.0181.

[30] A. Dey and P. Chaudhuri. On estimators of the mean of infinite dimensional data in finite populations. *Bernoulli*, 30(1):797–824, 2024.

[31] A. Dey and P. Chaudhuri. Quantile processes and their applications in finite populations. Revision submitted to the *Annals of Statistics*, 2024.

[32] A. Dey and P. Chaudhuri. Regression analysis and related estimators in finite populations. Submitted to a journal for publication, 2024.

[33] B. Efron. Regression percentiles using asymmetric squared error loss. *Statistica Sinica*, 1: 93–125, 1991.

[34] P. J. Farrella and M. Salibian-Barrerab. A comparison of several robust estimators for a finite population mean. *Journal of Statistical Studies ISSN*, 1022:4734, 2006.

[35] C. A. Francisco and W. A. Fuller. Quantile estimation with a complex survey design. *Ann. Statist.*, 19(1):454–469, 1991. ISSN 0090-5364. doi: 10.1214/aos/1176347993. URL `https://doi.org/10.1214/aos/1176347993`.

[36] W. A. Fuller. *Sampling statistics*. John Wiley & Sons, 2011.

[37] M. Geraci. Estimation of regression quantiles in complex surveys with data missing at random: An application to birthweight determinants. *Statistical methods in medical research*, 25:1393–1421, 2016.

[38] H. Glejser. A new test for heteroskedasticity. *Journal of the American Statistical Association*, 64(325):316–323, 1969.

[39] A. S. Goldberger et al. Econometric theory. *Econometric theory.*, 1964.

[40] J. Hájek. Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Ann. Math. Stat.*, 35:1491–1523, 1964.

[41] J. Hájek. Comment on "an essay on the logical foundations of survey sampling, part one". *The foundations of survey sampling*, 236, 1971.

[42] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust statistics: the approach based on influence functions*. John Wiley & Sons, 2011.

[43] Q. Han and J. A. Wellner. Complex sampling designs: Uniform limit theorems and applications. *The Annals of Statistics*, 49(1):459–485, 2021.

[44] D. G. Horvitz and D. J. Thompson. A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.*, 47(260):663–685, 1952.

[45] T. Hsing and R. Eubank. *Theoretical foundations of functional data analysis, with an introduction to linear operators*. John Wiley & Sons, 2015.

[46] P. J. Huber. Robust estimation of a location parameter. *Ann. Math. Statist.*, 35:73–101, 1964.

[47] B. Hulliger. Outlier robust horvitz-thompson estimators. *Survey Methodology*, 21(1):79–87, 1995.

[48] C. T. Isaki and W. A. Fuller. Survey design under the regression superpopulation model. *J. Amer. Statist. Assoc.*, 77(377):89–96, 1982. ISSN 0162-1459. URL `http://links.jstor.org/sici?sici=0162-1459(198203)77:377<89:SDUTRS>2.0.CO;2-8&origin=MSN`.

[49] R. Koenker. *Quantile regression*, volume 38. Cambridge university press, 2005.

[50] R. Koenker and G. Bassett, Jr. Regression quantiles. *Econometrica*, 46(1):33–50, 1978. ISSN 0012-9682. doi: 10.2307/1913643. URL `https://doi.org/10.2307/1913643`.

[51] R. Koenker and G. Bassett, Jr. Robust tests for heteroscedasticity based on regression quantiles. *Econometrica*, 50(1):43–61, 1982. ISSN 0012-9682. doi: 10.2307/1912528. URL `https://doi.org/10.2307/1912528`.

[52] A. Y. Kuk. Estimation of distribution functions and medians under sampling with unequal probabilities. *Biometrika*, 75(1):97–103, 1988.

[53] A. Y. C. Kuk and T. K. Mak. Median estimation in the presence of auxiliary information. *J. Roy. Statist. Soc. Ser. B*, 51(2):261–269, 1989. ISSN 0035-9246.

[54] S. Kundu, S. Majumdar, and K. Mukherjee. Central limit theorems revisited. *Stat. Probab. Lett.*, 47(3):265–275, 2000.

[55] D. B. Lahiri. A method of sample selection providing unbiased ratio estimates. *Bull. Int. Stat. Inst.*, 33(2):133–140, 1951.

[56] Y. Li, B. I. Graubard, and E. L. Korn. Application of nonparametric quantile regression to body mass index percentile curves from survey data. *Statistics in medicine*, 29:558–572, 2010.

[57] H. Midzuno. On the sampling system with probabilities proportionate to the sum of sizes. *Ann. Inst. Stat. Math.*, 3:99–107, 1952.

[58] M. N. Murthy. *Sampling theory and methods*. Statistical Publishing Society, Calcutta, 1967.

[59] W. K. Newey and J. L. Powell. Asymmetric least squares estimation and testing. *Econometrica*, 55(4):819–847, 1987.

[60] W. K. Newey and J. L. Powell. Asymmetric least squares estimation and testing. *Econometrica: Journal of the Econometric Society*, pages 819–847, 1987.

[61] E. Ohlsson. Asymptotic normality of the Rao-Hartley-Cochran estimator: an application of the martingale CLT. *Scand. J. Stat.*, 13(1):17–28, 1986. ISSN 0303-6898.

[62] E. Ohlsson. Asymptotic normality for two-stage sampling from a finite population. *Probab. Theory Related Fields*, 81(3):341–352, 1989. ISSN 0178-8051. doi: 10.1007/BF00340058. URL https://doi.org/10.1007/BF00340058.

[63] D. Pfeffermann. New important developments in small area estimation. *Statist. Sci.*, 28 (1):40–68, 2013. ISSN 0883-4237,2168-8745. doi: 10.1214/12-STS395. URL https://doi.org/10.1214/12-STS395.

[64] D. Raj. On sampling with probabilities proportional to size. *Ganita*, 5:175–182, 1954.

[65] J. Rao. Small area estimation, a john wiley & sons. *Inc, New Jersey*, 2003.

[66] J. Rao, H. Hartley, and W. Cochran. On a simple procedure of unequal probability sampling without replacement. *J. R. Stat. Soc. Ser. B Methodol.*, 24(2):482–491, 1962.

[67] J. Rao, J. Kovar, and H. Mantel. On estimating distribution functions and quantiles from survey data using auxiliary information. *Biometrika*, pages 365–375, 1990.

[68] P. Robinson and C. E. Särndal. Asymptotic properties of the generalized regression estimator in probability sampling. *Sankhyā: The Indian Journal of Statistics, Series B*, pages 240–248, 1983.

[69] S. Rubin-Bleuer and I. S. Kratina. On the two-phase framework for joint model and design-based inference. *The Annals of Statistics*, 33(6):2789–2810, 2005.

[70] M. Rueda, A. Arcos, M. Martınez-Miranda, and Y. Román. Some improved estimators of finite population quantile using auxiliary information in sample surveys. *Computational statistics & data analysis*, 45(4):825–848, 2004.

[71] C. E. Särndal. On π-inverse weighting versus best linear unbiased weighting in probability sampling. *Biometrika*, 67(3):639–650, 1980.

[72] C. E. Särndal, B. Swensson, and J. Wretman. *Model assisted survey sampling*. Springer Science & Business Media, 1992.

[73] C. E. Särndal, B. Swensson, and J. Wretman. *Model assisted survey sampling*. Springer Science & Business Media, 2003.

[74] A. Scott and C.-F. Wu. On the asymptotic distribution of ratio and regression estimators. *Journal of the American Statistical Association*, 76(373):98–102, 1981.

[75] A. Sen. On the estimator of the variance in sampling with varying probabilities. *Jour. Ind. Soc. Ag. Statistics*, 5(2):119–127, 1953.

[76] R. J. Serfling. *Approximation theorems of mathematical statistics*, volume 162. John Wiley & Sons, 2009.

[77] J. Shao. L-statistics in complex survey problems. *Ann. Statist.*, 22(2):946–967, 1994. ISSN 0090-5364. doi: 10.1214/aos/1176325505. URL `https://doi.org/10.1214/aos/1176325505`.

[78] X. Shi, C. Wu, and J. Chen. Weak and strong representations for quantile processes from finite populations with application to simulation size in resampling inference. *Canadian Journal of Statistics*, 18(2):141–148, 1990.

[79] G. R. Shorack and J. A. Wellner. *Empirical processes with applications to statistics*. SIAM, 2009.

[80] M. Stefan and M. A. Hidiroglou. Jackknife bias-corrected generalized regression estimator in survey sampling. *Journal of Survey Statistics and Methodology*, 2022.

[81] M. Subzar, C. N. Bouz, and A. I. Al-Omari. Utilization of different robust regression techniques for estimation of finite population mean in srswor in case of presence of outliers through ratio method of estimation. *Revista Investigación Operacional*, 40:600–609, 2019.

[82] M. Subzar, A. I. Al-Omari, and A. Alanzi. The robust regression methods for estimating of finite population mean based on srswor in case of outliers. *Computers, Materials & Continua*, 65:125–138, 2020.

[83] R. D. Tuddenham and M. M. Snyder. Physical growth of california boys and girls from birth to eighteen years. *University of California publications in child development*, 1:183—-364, 1954.

[84] A. W. Van Der Vaart and J. A. Wellner. *Weak convergence and empirical processes*. Springer, 1996.

[85] J. C. Wang and J. D. Opsomer. On asymptotic normality and variance estimation for nondifferentiable survey estimators. *Biometrika*, 98(1):91–106, 2011.

[86] H. White. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: journal of the Econometric Society*, pages 817–838, 1980.

[87] C. Wu. Algorithms and r codes for the pseudo empirical likelihood method in survey sampling. *Survey Methodology*, 31(2):239, 2005.