INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Robust Statistical Inference in Survey Sampling

by

J. P. Gwet, B.Sc., M.Sc.

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Department of Mathematics and Statistics
Carleton University
Ottawa, Canada
April, 1997

© Copyright, 1997

J.P. Gwet

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-22168-7

The undersigned hereby recommend to
The Faculty of Graduate Studies and Research
acceptance of the thesis

Robust Statistical Inference in Survey Sampling

submitted by
J. P. Gwet
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

Chair, Department of Mathematics and Statistics

Thesis Supervisor

External Examiner

-

Carleton University
April, 1997
© Copyright, 1997
J.P. Gwet

I dedicate this thesis to my Father and my Mother

Abstract

This thesis deals with the classical problem of finite population mean estimation. Because standard procedures can only guarantee reliable estimates for large samples, we propose robust estimators based on M-statistics and L-statistics. Our main concern is to develop a general methodology for deriving robust and consistent estimators within the framework of Generalized REGression (GREG) superpopulation models.

The common mean superpopulation model is studied first. Under this simple model, the work done by pioneers of the theory of robust estimation such as P. Huber is extended to complex designs. The conditions under which M-estimators can work well for estimating finite population means are made more precise. For the M-estimator to work well, one of the key conditions is that a symmetrically distributed superpopulation provides a reasonable description of the finite population under study. Relaxing this constraint, we propose CM-estimators where CM stands for consistent M-estimators. It is shown that the CM-estimator is consistent for any design regardless of the superpopulation model. When the CM-estimator is not applicable, we propose the LM-estimator. This is obtained by adjusting the ordinary M-estimator for the bias with a robust L-statistic. It is shown that the LM-estimator is also consistent for any design and any superpopulation model.

The theory developed under the common mean superpopulation model is extended to multiple regression superpopulation models. The same asymptotic properties

are proved under more regularity conditions and estimators of the mean squared errors are developed using the linearization method. A simulation study conducted, indicates a nonnegligible underestimation of the CM-estimator mean squared error (MSE). In order to improve on the MSE estimator, a jackknife variance estimator as well as a jackknife MSE estimator which yield a substantial improvement are proposed.

M-statistics are finally applied to develop a robust alternative to the best linear unbiased prediction estimator of small area means. We confine ourselves to the case of known variance components.

Acknowledgements

I thank Professor J.N.K Rao, my thesis supervisor who spent so many hours guiding me through to the completion of this thesis. I wanted to apply the theory of robust statistics to survey sampling problems and he kindly accepted to supervise my research work. His commitment to insight and high standards has inspired me all along.

This work is a follow-up of a previous research work done in collaboration with Professor Louis-Paul Rivest. I am indebted to him for convincing me to study the application of robust statistics to survey sampling. I am grateful to Dr M. Alvo who provided helpful comments.

In Particular, I thank my wife Suzy for her understanding and support in all phases of my Ph.D. program. She supported tirelessly my absence during the long hours I used to spend digging in my books.

I express my gratitude to my father and my mother who wanted me to pay any price for the completion of my doctorate program.

Contents

Abstract	iv
Acknowledgements	vi
Table of Contents	vii
List of Tables	xiii
List of figures	xv
List of Appendices	xvi
Chapter 1. Introduction	1
1.1 Robustness in Finite Population Sampling	1
1.2 Review of the Literature	2
1.2.1 Methods where Robustness is Achieved at the Design Stage	4
1.2.2 Methods where Robustness is Achieved at the Weighting Stage	8
1.2.3 M-estimation Approach	15
1.3 The Generalized Regression Estimator	21
1.4 Robustness in Small Area Estimation	25
1.4.1 The Direct Methods	27
1.4.2 Estimators Making Implicit Use of Models	28
1.4.3 Model-Based Methods	31
1.4.4 Case-Deletion Diagnostics	33

1.4.5 Robust Methods	35
1.5 Thesis Outline	39
Chapter 2. Robust Estimation of the Population	
Mean under the Common Mean	
Superpopulation Model	44
2.1 Introduction	44
2.2 Review of Existing Procedures	48
2.2.1 The GREG Estimator	48
2.2.2 The Gwet-Rivest M-estimator	49
2.3 M-estimation for Complex Samples: the Symmetric Case	52
2.3.1 The M-estimator	52
2.3.2 Large-Sample Properties of the M-estimator	57
2.4 M-estimation for Complex Samples: the Asymmetric Case	65
2.4.1 The Consistent M-estimator	65
2.4.2 The LM-estimator	70
2.4.3 Asymptotic Properties of the CM-estimator	74
2.4.4 Asymptotic Properties of the LM-estimator	85
2.5 Simulation Study	102
2.5.1 Objectives	102
2.5.2 Description of the Experiment	104
2.5.3 Performance of the Robust Estimators on Symmetric Populations	106

2.5.4	Performance of the Robust Estimators on the MU284 Population	108
2.5.5	Accuracy of the Approximation to the MSE	110
2.5.6	Bias of the MSE's Estimators	111
 Chapter 3. Efficiency-Robust Alternatives to the		
GREG Estimator under the GREG		
Superpopulation Model 115		
3.1	Introduction	115
3.2	Review of Existing Procedures	117
3.2.1	The GREG Estimator	118
3.2.2	The Gwet-Rivest M-estimator	120
3.3	M-estimation under the Assumption of Symmetry: Extension to	
	Complex Designs	124
3.3.1	The M-estimator	124
3.3.2	Large-Sample Properties of the M-estimator	130
3.4	Robust Estimation for Skewed Populations	137
3.4.1	The CM-estimator	137
3.4.2	The LM-estimator	139
3.5	Asymptotic Properties of the Estimators	143
3.5.1	Asymptotic Properties of the CM-estimators	143
3.5.2	Asymptotic Properties of the LM-estimators	158
3.6	CM and LM Estimators under the Ratio Model	166

3.6.1 SRS Design	167
3.6.1.1 CM-estimator	167
3.6.1.2 LM-estimator	169
3.6.2 Stratified Random Sampling	171
3.6.2.1 CM-estimator	171
3.6.2.2 LM-estimator	173
3.7 CM and LM Estimators under the Simple Regression Model and the Simple Random Sampling Design	175
3.7.1 CM-estimator	175
3.7.2 LM-estimator	178

Chapter 4. Jackknife Variance Estimation on the

CM and LM Estimators	180
4.1 Introduction	180
4.2 An Overview of the Jackknife Methodology	183
4.3 Jackknife Variance Estimation Under the SRS-WOR Design	186
4.3.1 Jackknifing the CM-estimator	186
4.3.2 Jackknifing the LM-estimator	193
4.4 Large Sample Results	197
4.4.1 The CM-estimator	198
4.4.2 The LM-estimator	220
4.5 Monte-Carlo Study	229

4.5.1 Objectives	229
4.5.2 Description of the Experiment	231
4.5.3 Simulation Results	232
Chapter 5. Robust Estimation of Small Area Means	
Under the One-Fold Nested-Error	
Regression Model	237
5.1 Introduction	237
5.2 The Small Area Model	241
5.3 Robust Estimation of the Domain Mean Under	
The Assumption of Known Variance Components	242
5.3.1 The Infinite Population Case	243
5.3.2 The Finite Population Case	248
5.3.3 Computation of the Robust Estimators	249
5.3.4 Asymptotic Properties of the GM-estimator	251
5.4 Robust Estimation of the Domain Mean Under	
The Assumption of Unknown Variance Components	277
5.4.1 The Proposed Estimators	279
5.4.2 An Algorithm	281
5.5 Monte-Carlo Study	284

Chapter 6. Summary and Outstanding Issues for	
Future Research	291
6.1 Summary	291
6.2 Future Research	297
Appendix A. Technical Proofs of Chapter 3	300
Appendix B. Some Proofs and Results Related	
to Chapter 4	308
Appendix C. Some Proofs and Results Related	
to Chapter 5	318
References.	322

List of Tables

Table 2.1 Description of the Symmetric Finite Populations Generated From the Superpopulation Model $y_i = \mu + \epsilon_i$.

Table 2.2 Performance of the Estimators on the Symmetric Populations.

Table 2.3 Performance of the Estimators on the MU284 Population.

Table 2.4 Percent Relative Error of the MSE Approximations

Table 2.5 Percent Relative Bias of the MSE Estimators

Table 2.6 Choice of $K_n(\alpha)$ for the CM-estimator

Table 4.1 Monte Carlo Biases, Variances and Mean Squared Errors of the GREG, CM and LM estimators based on 10,000 replicates.

Table 4.2 Percentage Relative Biases of the Variance Estimators Based on the Jackknife and the Taylor Series Methods. Results Based on 10,000 Replicates.

Table 5.1 Efficiencies of Small Area Mean Estimators and Percent Relative Bias for the Corresponding MSE Estimators When $\mathcal{L}_v = \mathcal{N}(0, 64)$, $\mathcal{L}_e = \mathcal{N}(0, 5292)$ and $m = 15$

Table 5.2 Efficiencies of Small Area Mean Estimators and Percent Relative Bias for the Corresponding MSE Estimators when $\mathcal{L}_v = \mathcal{N}(0, 64)$, $\mathcal{L}_e =$

$\mathcal{D}(0, 5292/2)$ and $m = 15$

Table 5.3 Efficiencies of Small Area Mean Estimators and Percent Relative Bias
for the Corresponding MSE Estimators when $\mathcal{L}_v = \mathcal{N}(0, 64)$, $\mathcal{L}_e =$
 $(1 - \epsilon)\mathcal{N}(0, 292) + \epsilon\mathcal{N}(0, 100 * 292)$ and $m = 15$

Table 5.4 Percent Relative Bias for MSE Estimators when $\mathcal{L}_v = \mathcal{N}(0, 64)$ and
 $m = 60$

List of Figures

Figure 3.1: Contaminated Normal Distributions For Various Contamination
Rates ϵ

Figure 3.2: Influence Function of the M-estimator

Figure 3.3: Influence Function of the GM-estimator

List of Appendices

Appendix A: Technical Proofs of Chapter 3

Appendix B: Some Proofs and Results Related to Chapter 4

Appendix C: Some Proofs and Results Related to Chapter 5

CHAPTER 1

Introduction

1.1. Robustness in Finite Population Sampling

In survey sampling, the word “robustness” traditionally refers to the validity robustness. The main concern is the validity of the inference that guarantees confidence intervals with covering probabilities close to their nominal values no matter how the population units are distributed. Therefore it is possible to make a distribution-free statement about the accuracy of an estimator.

For the design-based confidence interval to be valid, it is usually necessary to have an unbiased (or approximately unbiased) estimator and a consistent variance estimator. For that reason, design-unbiasedness has become a compelling principle of design-based inference. As we will see later this limitation can affect the efficiency of the estimator dramatically.

The design-based confidence intervals, although valid are sometimes unduly large making the results useless. Therefore many *validity-robust* design-based procedures will not be *efficiency robust* very often. This will mainly occur when the sample size is not very large or the distribution of the target population has heavy tails. The lack of efficiency robustness of many standard estimation procedures is frequently handled in an informal way by survey practitioners as no other alternative is usually available. The reliability of the resulting estimates is questionable. The robust inference we plan to develop in the context of survey problems will face up to these limitations by providing robust estimation methods along with statistical measures for assessing the precision of the estimates.

In section 2, we review the existing robust methods proposed for survey problems. As will be seen, most of these methods have been proposed as robust versions of some specific standard estimators. The number of standard procedures is so high that it is more convenient to adopt a unifying approach that will use a common structure shared by most of them. Such a common structure is also discussed at the end of section 2. Section 3 deals with the robustness in small area estimation. A review of the literature will be provided as well.

1.2. Review of the Literature

The problem of robust estimation is usually encountered in survey sampling when some sample units are outlying. But estimation in the presence of outliers is a par-

ticular case of robust estimation. Within the framework of parametric statistical inference, some authors (see Barnett and Lewis, 1984) have mentioned the difference between robust methods designed to perform well under a variety of population distributions, and those that specifically refer to the outliers. The latter have the disadvantage of requiring a precise definition of what an outlier is. In general this is difficult and particularly in sample survey where weighting is involved. We intend to develop estimators with general robustness properties that will carry over to the outlier problem. Our procedures are closer in spirit to the type I procedures mentioned by Barnett and Lewis and are developed within the framework of superpopulation models.

Other definitions of robust estimation are available in the literature, Cassel, Särndal and Wretman (1977) say “the general aim in robust estimation is to find an estimator or a strategy that performs well in some broad sense allowing for our uncertainty about the real world”. Bickel (1976) says that robust procedures “... are procedures which may be mildly suboptimal at the original model, which pay an *insurance premium* in Anscombe’s (1960) terms but perform satisfactorily in the supermodel generally”.

Two main approaches to achieving robustness of efficiency in survey sampling have been proposed in the literature. The first approach is what can be called the *prevention approach* at the design stage, the second approach is the *weight-modification approach*. Note that many authors, Chambers & Kokic (1993) or Lee (1995) among

others often mention a third approach that is the value-modification approach such as trimming or winsorization. However that approach can be seen as a special case of the weight-modification approach as will be shown later in this section. The robust estimation techniques which apply the classical theory of robust estimation of Huber (1964) to survey problems is also seen as a particular case of the weight-modification approach. Recently, Fuller (1993) introduced a new technique that is based upon the idea of modelling only the tail of the distribution and incorporating the estimated parameters into the estimation procedure. All these methods are reviewed in the next paragraphs.

1.2.1. Methods where Robustness is Achieved at the Design Stage

As mentioned earlier, this approach aims at solving the outlier problem at the design stage. One such approach is to split the population into a take-all stratum and a number of take-some strata. Let $\mathcal{U}$ be a finite population of size N and y a variable of interest. The parameter of interest is the population mean $\bar{Y}$ given by

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N y_i.$$

Lavallée and Hidioglou (1988) suggested to split the population into H strata. One stratum (say H) is a take-all stratum made up of the largest units and the others ($h = 1, \dots, H - 1$) are take-some. All the units from the take-all stratum are selected with probability one and a sample s_h of size n_h is selected from each of the

take-some strata. Let us denote by N_h , the number of population units in stratum h and $s = s_1 \cup \dots \cup s_H$ the whole sample.

The population mean can be written as

$$\bar{Y} = \frac{1}{N} \sum_{h=1}^H \sum_{i=1}^{N_h} y_{hi}.$$

Lavallée and Hidirolou (1988) proposed the following estimator assuming that a simple random sample s is drawn without replacement within strata:

$$\bar{y}_{LH} = \frac{1}{N} \left[\sum_{h=1}^{H-1} \frac{N_h}{n_h} \sum_{i \in s_h} y_{hi} + \sum_{i=1}^{N_H} y_{Hi} \right].$$

The boundaries are determined so that the overall sample size n is minimized given a fixed precision and a specific allocation scheme. The main drawback of this approach is the difficulty to obtain the strata boundaries without having real data. In fact once the sample is selected, the y_{Hi} 's may not even be among the N_H largest units in the population. The method is complicated and many approximations are involved. The authors proposed to use an auxiliary variable that is correlated to the variable of interest. Such an auxiliary variable might be the same variable of interest observed during a previous census. However, as pointed out by Rivest (1993) some units might have increased dramatically by the time the survey is conducted. Thus, the use of an auxiliary variable might lead to a loss of efficiency that is difficult to assess. The whole procedure heavily relies upon the quality of the correlation between the variable of interest and the auxiliary variable. This is not a good feature for a robust procedure.

The prevention approach is also discussed by Glasser (1962) as well as by Kish (1965). Kish (1965) suggested the disproportionate allocation that is close to the take-all and take-some approach discussed earlier. He also suggested procedures for finding rare elements as well as truncation of large values when no reliable auxiliary information is available. The truncation idea will be discussed in the next section. But the other procedures still require auxiliary variables that are very correlated with the variable of interest.

Another well-known approach to achieving robustness of efficiency is the probability proportional-to-size (PPS) sampling with or without replacement. Many estimation strategies are available in this context. We will only mention the PPS sampling without replacement and the Horvitz-Thompson estimator. Let x be an auxiliary variable that is proportional to y and known for all population units. Assume that the selection is without replacement and that unit i is included in the sample with probability π_i . Two units i and j are included in the sample with probability π_{ij} .

We have that

$$\pi_i = \frac{nx_i}{\sum_{k=1}^N x_k},$$

The Horvitz-Thompson estimator $\hat{Y}_{HT}$ is given by

$$\hat{Y}_{HT} = \frac{1}{N} \sum_{i=1}^n \frac{y_i}{\pi_i}.$$

Its variance is given by,

$$V(\hat{Y}_{HT}) = \frac{1}{N^2} \sum_{i < j}^N (\pi_i \pi_j - \pi_{ij}) \left(y_i / \pi_i - y_j / \pi_j \right)^2.$$

If x and y are highly correlated, then robustness of efficiency will be achieved as y_i/π_i and y_j/π_j will have approximately the same value. Again the robustness of this estimator heavily relies upon the underlying model of proportionality between y and x . As pointed out by Rao (1966) this estimator will be less efficient than the ordinary sample mean if the correlation is poor. This was shown by comparing the expected variances of both estimators under the following superpopulation model.

$$(\xi) \quad \begin{cases} y_i = \beta + \epsilon_i, \\ E_{\xi}(\epsilon_i|x_i) = 0; \quad E_{\xi}(\epsilon_i^2|x_i) = \lambda, \quad \lambda > 0, \\ E_{\xi}(\epsilon_i\epsilon_j|x_ix_j) = 0 \end{cases}$$

and by proving that the bias of the sample mean is fairly small under this model. Therefore when the ratio model breaks down, the Horvitz-Thompson estimator might be very inefficient. If $\bar{y}$ is the sample mean, then

$$V(\bar{y}) = (1/n^2) \sum_{i < j}^N \sum_{j < i}^N (\pi_i\pi_j - \pi_{ij})(y_i - y_j)^2.$$

It follows from this variance formula that the sample mean is clearly much less sensitive to the correlation between the characteristics y and x . It is however a biased estimator with a bias given by,

$$\begin{aligned} B(\bar{y}) &= (1/n) \sum_{i=1}^N y_i\pi_i - \bar{Y}, \\ &= \frac{1}{\bar{X}} \times \frac{1}{N} \sum_{i=1}^N x_i y_i - \bar{Y}. \end{aligned}$$

This bias will be small if the correlation between x and y is weak.

The main drawback of the prevention approach is observed in large-scale surveys when the estimates must be produced for several variables of interest. As pointed out by Lee (1995), only a few key variables are used at the design stage for stratification or as size measures for the PPS sampling. Therefore, robust estimation cannot be performed on the other variables of interest.

1.2.2. Methods where Robustness is Achieved at the Weighting Stage

The outlier problem is usually solved in many Statistical agencies such as Statistics Canada or Statistics Sweden by reducing the weight associated with the large units. Hidioglou and Srinath (1981) as well as Dalèn (1987) proposed many estimators that appeal to this technique. This approach is likely to increase the estimator bias. But the resulting mean squared error (MSE) will usually be smaller than that of the standard procedures. The *Poststratified (PS) estimator* under the simple random sampling (SRS) design is given by

$$\bar{y}_{PS} = \frac{1}{N} \left(\frac{T}{t} \sum_{i \in s_0} + \frac{N-T}{n-t} \sum_{i \in \bar{s}_0} y_i \right),$$

where s_0 is the subsample containing the outliers, $\bar{s}_0 = s/s_0$, T is the number of population outliers and t the number of sample outliers. Suppose that a population unit i is outlying with respect to the y variable, if $y_i \geq K_0$ for some positive constant

K_0 and $\omega_{0i} = n/N$ the basic sampling weight. The PS estimator has the form

$$\bar{y}_{PS} = \frac{1}{N} \sum_{i=1}^n \omega_i^* y_i \text{ where } \omega_i^* = \begin{cases} T/t & \text{if } y_i \geq K_0, \\ (N-T)/(n-t) & \text{otherwise.} \end{cases} \quad (1.1)$$

This estimator which is constructed by replacing the basic weight ω_{0i} by the new weight ω_i^* has very interesting conditional properties; the condition being that there is at least one outlier in the sample. This estimator seems to perform well when the number of population outliers is large enough. The main problem of this estimator is its dependence upon the number of outliers in the population as well as in the sample. In practice, the number of population outliers is never known. Moreover it is not possible to know whether this estimator will be used or not prior to sample selection. These two problems limit the usefulness of this estimator.

The variance of this estimator conditionally on the number t of sample outliers is given by,

$$V(\bar{y}_{PS}|t) = (T/N)^2 \left(1/t - 1/T\right) S_\mu^2 + (1 - T/N)^2 \left(\frac{1}{n-t} - \frac{1}{N-T}\right) S_\nu^2.$$

where S_μ^2 and S_ν^2 are given by,

$$S_\mu^2 = \frac{1}{T-1} \sum_{i=1}^T (y_i - \bar{Y}_\mu)^2 \text{ and } S_\nu^2 = \frac{1}{N-T-1} \sum_{i=1}^{N-T} (y_i - \bar{Y}_\nu)^2,$$

$\bar{Y}_\mu$ is the population mean of the outlying values while $\bar{Y}_\nu$ is the mean of the non-outlying values. For this poststratified estimator to be used, there must be at least two outliers in the population. Moreover, it is not clear how the above conditional variance can be estimated. Not only T is always unknown, it is common in practice

that there is only one outlier in the sample. in which case S_{μ}^2 cannot be estimated easily.

Other estimators studied by Hidioglou and Srinath (1981) have the form (1.1) with

$$\omega_i^* = \begin{cases} r & \text{if } y_i \geq K_0, \\ (N - rt)/(n - t) & \text{otherwise,} \end{cases}$$

where r is the weight given to large units. In general $r = 1$ but Hidioglou and Srinath studied the case where r is chosen so as to minimize the MSE. This optimal weight depends on some population parameters that are unknown and must be estimated. These authors have not discussed the way large units are determined. But no matter how this is done, identifying outliers amounts to setting a threshold K_0 above which a unit is taken to be outlying. That is the reason for which we can use K_0 . The simulation results presented by these authors indicate some gain in precision compared to the sample mean. But it appears also clearly that when the number of population outliers goes beyond a certain value, these estimators become very inefficient. The maximum number of population outliers tolerated by these methods is not known in practice.

It is important to stress that the modified weights are usually based on the variable of interest y . If the survey data are collected for other variables, it might be inappropriate to use the same modified weights. Therefore calculating new weights for each variable of interest seems to be the only way to treat outliers efficiently.

Other popular methods for dealing with outlying units are referred to as value-

modification methods. The values of outliers are changed by Winsorization. Numerous versions of Winsorized estimators are available in the literature. Only some of them will be mentioned here. A general form of the Winsorized estimator under the SRS design is given by

$$\bar{y}_w = \sum_{i \in s} \omega_{0i} y_i^* / N,$$

where $\omega_{0i} = N/n$, $\forall i \in s$ and y_i^* is the new value. A general form of y_i^* often used in practice is given by

$$y_i^* = \begin{cases} f y_i + (1 - f) K_0 & \text{if } y_i \geq K_0, \\ y_i & \text{otherwise,} \end{cases}$$

where $0 \leq f \leq 1$ and K_0 a predetermined cut-off point. Using the terminology of Gross et al.(1986), $\bar{y}_w$ will be called the *Winsorized Type I Estimator (W1)* when $f = 0$ and the *Winsorized Type II estimator (W2)* when $f = n/N$. We have not mentioned the value-modification methods as one possible approach in the beginning of this section simply because it can be seen as a weight-modification method. In fact even weight-modification methods can be seen as value-modification methods.

Note that the Winsorized type II estimator $\bar{y}_{w2}$ can be rewritten as follows:

$$\bar{y}_{w2} = \sum_{i \in s} \omega_i^* y_i / N, \text{ where } \omega_i^* = \begin{cases} N/n & \text{if } y_i < K_0, \\ 1 + (N/n - 1) \frac{K}{y_i} & \text{if } y_i \geq K_0. \end{cases}$$

For the Winsorized type I estimator, we have

$$\omega_i^* = \begin{cases} N/n & \text{if } y_i < K_0, \\ (N/n) \frac{K}{y_i} & \text{if } y_i \geq K_0. \end{cases}$$

This modified weight is not necessarily more dependent on the y variable than the modified weights of the weight-modification methods. Unless the modified weights is

applied to all variables of interest (which is in general inappropriate) as the sampling weights, changing the weights in sample surveys will have almost the same effect as changing the values. The weight attached to an observed value is equally important for estimation purposes as the observed value itself.

Rivest (1993) studied Searls' optimal Winsorized estimator in the context of finite populations. The optimality criterion is satisfied if the Winsorizing cutoff K_0 minimizes the MSE of the Winsorized estimator. Rivest studied the optimal Winsorized type I estimator as well as a number of versions thereof. He expressed the means squared error of $\bar{y}_{w1}$ as,

$$\text{MSE}(\bar{y}_{w1}) = \frac{1-f}{n} \left(S_y^2 - 2 \int_K^\infty (y - \bar{Y})(1 - F(y)) dy - B^2(\bar{y}_{w1}) \right) + B^2(\bar{y}_{w1}).$$

where $B^2(\bar{y}_{w1})$ is given by,

$$B^2(\bar{y}_{w1}) = \int_K^\infty (1 - F(y)) dy.$$

Differentiating with respect to K , it was established that the optimal cut-off point K_0 is implicitly defined by the following equation,

$$\frac{K_0 - \bar{Y}}{n/(1-f) - 1} = \frac{1}{N} \sum_{i=1}^N \max(y_i - K_0, 0).$$

The optimal Winsorized type I estimator can perform very well provided a reliable auxiliary information is available to estimate the Winsorizing cutoff. If such information is unavailable, estimating the optimal cutoff from the sample will result in a substantial loss in efficiency. Rivest (1993) also discussed the extension

of this optimal Winsorized type I estimator to the case of stratified sampling design. Estimating the optimal cutoff still poses a problem. Furthermore, the author acknowledges that “*the biases of Winsorized estimators are large. In repeated surveys a systematic bias of more than 5% on individual population estimates is not acceptable*”.

More recently Kokic & Bell (1994) studied the optimal Winsorized type II estimator for estimating a stratified finite population mean $\bar{Y}$. It is assumed that the population is divided into H non-overlapping strata and the type II Winsorized estimator of $\bar{Y}$ is given by,

$$\bar{y}_{w2} = \frac{1}{N} \sum_{h=1}^H (N_h/n_h) \sum_{i=1}^{n_h} y_{hi}^*, \text{ where } y_{hi}^* = \begin{cases} f_h y_{hi} + (1-f_h)K_h & \text{if } y_{hi} \geq K_h, \\ y_{hi} & \text{otherwise.} \end{cases}$$

The finite population is also assumed to follow a superpopulation model (ξ) given by,

$$(\xi) \quad E_{\xi}(y_{hi}) = \mu_h \text{ and } V_{\xi}(y_{hi}) = \sigma_h^2, \quad \forall h = 1, \dots, H.$$

The problem taken up by Kokic and Bell (1994) was to find an optimal set of cutoff points $(K_1, \dots, K_H)$ that minimizes the mean squared error of the Winsorized estimator. Although this problem is close that of Rivest (1993), they propose a different algorithm for finding the optimal Winsorizing cutoffs K_{0h} for strata $h = 1, \dots, H$. They expressed each cutoff as a function of a population parameter that can be estimated more accurately. The optimal cutoffs K_{0h} 's are given by,

$$K_{0h} = (N_h/n_h - 1)^{-1}L + \mu_h, \text{ where } L = \sum_{h=1}^H N_h(1-f_h) \times [E_{\xi}(y_h J_h) - K_h E_{\xi}(J_h)].$$

and $J_{hi} = I(y_{hi} \geq K_h)$. The dependence of the optimal K_{oh} 's on a single constant L seems to be an advantage over Rivest's approach where each cutoff depends on the empirical cumulative distribution function (CDF) within the corresponding stratum. Estimating the CDF in small strata may be difficult. But Kokic & Bell (1994) also need auxiliary information to estimate the optimal cutoff. The simulation results they presented indicate some improvement over the standard unbiased estimator for stratified samples. The bias of the proposed estimator is not too high in the simulation study. This stems from the dependency of the Winsorized type II estimator on the sampling fraction, which provides some control over the bias for large samples.

Chambers (1982) extended the cutoff idea to the case where the variable of interest is related to the auxiliary variables according to a linear regression model. However even if this is easily done in the model-based inference framework, it is not easy to see how Winsorized estimators can be extended to complex designs such as the PPS sampling design or the two-phase sampling. The Winsorized estimators are very simple to understand and to explain to the users of survey data. But this technique does not rely on a strong theoretical basis and that might explain why the extension to complex designs is still difficult.

As stated before, all methods presented so far were mainly developed for the SRS design. Many sample surveys are conducted under complex designs. In order to solve

the outlier problem in this context, Dalén (1987) proposed the following estimator:

$$\bar{y}_D = \frac{1}{N} \sum_{i \in s} \omega_i^* y_i, \text{ where } \omega_i^* = \begin{cases} w_{0i} & \text{if } \omega_{0i} y_i < K_0. \\ 1 + (\omega_{0i} - 1)(K_0 / \omega_{0i} y_i) & \text{if } \omega_{0i} y_i \geq K_0 \end{cases}$$

and ω_{0i} is the original sampling weight. Dalén reported some gain in efficiency by this method in a simulation study. However the theoretical properties of this estimator have not been studied thoroughly yet. Also in Dalén approach, the outlier problem is not treated independently from the design. This has the disadvantage of relating the outlyingness of sample units to the way they are selected.

In the next section, we will review the methods that are based on the modern theory of robust estimation.

1.2.3. M-estimation Approach

The approach based on M-estimation, can be seen as a version of the weight-modification approach as will be seen later. Because it is the focus of this thesis, it was deemed more convenient to treat M-estimators in a different section. Many ideas underlying this approach, are borrowed from the theory of robust statistics originally developed within the parametric statistical inference framework by Huber (1964). Hampel et al.(1986) provide a comprehensive development of the theory of robust statistics.

Most applications of the theory of robust statistics to survey problems are confined to M-statistics. Chambers (1986) presented one of the first application of the theory to survey problems. His approach is model-based. Therefore, the reliability of the

resulting inference heavily relies upon the validity of the assumed model. Chambers' robust estimation procedure was proposed in the context of the following ratio superpopulation model,

$$(\xi_0) \quad \begin{cases} y_i = \beta x_i + \epsilon_i, \\ E_{\xi_0}(\epsilon_i \epsilon_j) = 0 \quad \forall i \neq j, \\ E_{\xi_0}(\epsilon_i) = 0 \text{ and } V_{\xi_0}(\epsilon_i) = \sigma_i^2 = \sigma^2 v(x_i). \end{cases}$$

Let s be the sample of size n , $\mathcal{U}$ the population of size N and $f = n/N$ the sampling fraction. Chambers' M-estimator of the population mean $\bar{Y}$ is given by,

$$\hat{\bar{Y}}_{CH} = f\bar{y} + (1 - f)\hat{\beta}_{CH}\bar{X}_2$$

where

$$\bar{X}_2 = (N - n)^{-1} \sum_{i \in \mathcal{U} \setminus s} x_i$$

and

$$\hat{\beta}_{CH} = \hat{\beta}_M + \frac{\sum_{i \in s} (x_i/\sigma_i) \psi_H((y_i - \hat{\beta}_M x_i)/\sigma_i)}{\sum_{i \in s} x_i^2/\sigma_i^2},$$

$\bar{y} = (y_1 + \dots + y_n)/n$ being the sample mean. Here, $\hat{\beta}_M$ is the M-estimator of β as proposed by Beaton and Tukey (1974). Chamber's (1986) also derived the asymptotic variances as well as the asymptotic bias of that estimator under an outlier-prone superpopulation model. These results are somewhat complicated and it is not known how sensitive they are to the model. Moreover, no variance or mean squared error(MSE) estimator was proposed. This is a serious limitation to the practical use of this estimator.

Using a model-assisted approach, Gwet and Rivest (1992) developed a robust alternative to the ratio estimator under the simple random sampling (SRS) design. Suppose that y is the variable of interest and assume that there exists an auxiliary variable x known for all population units. It is also assumed that the following superpopulation model provides a reasonable description of the population:

$$(\xi) \begin{cases} y_i = \beta x_i + \epsilon_i, \\ \forall i = 1, \dots, N, \quad E_\xi(\epsilon_i) = 0 \text{ and } V_\xi(\epsilon_i) = \sigma^2 x_i. \end{cases}$$

A simple random sample s is selected without replacement from the population $\mathcal{U}$. In this context, the population mean $\bar{Y} = \sum_{i \in \mathcal{U}} y_i$ is usually estimated by the ratio estimator $\bar{y}_R = (\bar{X}/\bar{x})\bar{y}$ where $\bar{X}$ is the population mean of the auxiliary variable. It is a well-known fact that random part of this estimator $(\bar{y}/\bar{x})$ can be obtained as the weighted least squares estimator of β . The Gwet-Rivest estimator is obtained by replacing the least squares principle by the bounded influence principle as discussed in Hampel et al. (1981). It is given by

$$\bar{y}_{GR} = \hat{\beta}_{GR} \bar{X},$$

where $\hat{\beta}_{GR}$ is implicitly defined by the following estimating equation

$$\sum_{i \in s} h_i \sqrt{x_i} \psi \left(\frac{y_i - \hat{\beta}_{GR} x_i}{\hat{\sigma} h_i \sqrt{x_i}} \right) = 0, \quad (1.2)$$

and

- ψ is the influence function which is usually odd, close to the identity in the neighbourhood of 0 and bounded. All these features are necessary to achieve robustness and efficiency.

- $h_i = h(x_i)$ is a function of x_i that must be chosen so as to dampen the impact of high leverage points.
- $\hat{\sigma}$ is an estimator of the superpopulation parameter σ that must not be too sensitive to outlying units.

It is important to stress here that in the univariate case discussed previously, the concepts of “outlying unit” and “large unit” are identical. The situation is more complex when there is an auxiliary variable. A unit is outlying when there is no indication that it might have been generated by the superpopulation model. Such units will have large residuals. If the least squares principle is used to estimate β there will be no way of identifying aberrant values. The impact of such units on an estimator based on the least squares principle can be dramatic. In equation 1.2, the h_i 's are needed as the bounded ψ function cannot protect the estimator against high leverage points.

The estimator $\hat{\beta}_{GR}$ is usually referred to as the GM-estimator of β in the robustness literature. GM-estimators are Generalized M-estimators and the generalization comes from the introduction of the h_i factors which do not exist in the definition of an M-estimator. Gwet & Rivest suggested to take $h_i = (\bar{X}/x_i)^{1/2}$ and have used the Huber type influence function ψ_H given by

$$\psi_H(t) = \text{sgn}(t) \min(K_H, |t|),$$

where K_H is the robustness constant that is usually chosen so as to achieve a 95% efficiency under normality of the errors in model (ξ) and x follows a χ_1^2 distribution

($K_H = 2.56$). The estimator $\hat{\sigma}$ of the superpopulation parameter is given by

$$\hat{\sigma} = \frac{1}{\Phi^{-1}(0.75)} \text{median}_{1 \leq i \leq n} |r_i| \text{ where } r_i = y_i - \hat{\beta}_{GR},$$

and Φ is the cumulative distribution function of the standard normal distribution. Note that if the superpopulation errors ϵ_i 's are normally distributed then under some regularity conditions, $\hat{\sigma}$ is a consistent estimator of σ with respect to the model ξ and the design p . The simulation results presented by Gwet and Rivest (1992) indicate some improvement over the ordinary ratio estimator although the study was limited to small samples. The gain in efficiency over the ratio estimator might decrease quickly as the sample size increases. This is partly due to the fact the ratio estimator has already some interesting robustness properties as pointed out by Rao (1978).

Hulliger (1995) suggested the use of outlier robust Horvitz-Thompson estimators (RHT estimators). He uses the same superpopulation model (ξ) as Gwet and Rivest but assumes that the sample is drawn according to the PPS sampling design where the size variable is x_i . His problem was slightly different as the auxiliary information is used also at the design stage. If n is the sample size, then the probability of selecting unit i is given by:

$$\pi_i = \frac{nx_i}{\sum_{i \in \mathcal{U}} x_i}.$$

After proving that the Horvitz-Thompson estimator is the weighted least squares estimator of the superpopulation parameter β , he derived a new robust estimator by replacing the least squares principle by the bounded influence principle. Hulliger's

estimator is given by

$$\bar{y}_{HU} = \hat{\beta}_{HU} \bar{X},$$

where $\hat{\beta}_{HU}$ is implicitly defined as solution to the following estimating equation:

$$\sum_{i \in s} h_i \sqrt{x_i} \psi \left(\frac{y_i - \hat{\beta}_{HU} x_i}{\sqrt{x_i}} \right) = 0,$$

h_i being defined by,

$$h_i = \min \left(1, K_x \frac{\text{median}_{1 \leq i \leq N} \sqrt{x_i}}{\sqrt{x_i}} \right).$$

The robustness constants K_H and K_x are both given value 2. Hulliger developed the bias as well as the variance of the robust Horvitz-Thompson estimator. These results as well as those of Gwet and Rivest (1992) are special cases of general results derived in chapter 3. The scale parameter is not estimated in Hulliger proposal. If it is much greater than 1, the Hulliger estimator will be inefficient.

Hulliger (1995) proposed another robust estimator called MER-estimator where MER stands for minimum estimated risk. It is similar to the outlier robust Horvitz-Thompson estimator except that the robustness constant is chosen so as to minimize the estimated MSE. This robustness constant is likely to be very unstable. But the simulation results presented in his paper indicate some gain in precision for both robust estimators although the sample size was again quite small and the populations not too skewed. Furthermore it would be interesting to compare the performance of the RHT estimator with the sample mean suggested by Rao (1966) when y and x are poorly correlated.

The Robust estimators based on M-statistics seem to perform very well when the sample is not very large and the population not too skewed. But it is clear that they are sensitive to the superpopulation model they are built upon and this a limitation of the M-estimation technique. Protection can be expected against model contamination and not against model breakdowns.

In the past few sections we have reviewed a number of robust procedures for estimating a finite population mean. However, each procedure is designed within a very narrow framework where the sampling design is fully specified and the estimation method is simple. Most of these methods are not applicable when multiple regression models are involved and the sampling design is quite complex. This thesis is an attempt to provide a general methodology of deriving robust estimators under a wide variety of situations. This will be done within the framework of Generalized REGression models (GREG). These models play an important role in the choice of estimators in model-assisted survey sampling. See Särndal et al. (1992) for a comprehensive development of model-assisted survey sampling. In the next section, we will review briefly the theory of GREG estimators.

1.3. The Generalized Regression Estimator

This section aims at providing the general set-up within which our investigation will be carried out. It seems to be more convenient to use a common structure that encompasses various well-known estimators as special cases. The robust estimation

methodology for survey problems will be developed around such a common structure avoiding thereby the disparate approaches found in the literature. As we have seen in the previous section, many authors have developed robust estimators to be used only in some specific situations. The Gwet-Rivest estimator (Gwet & Rivest, 1992) as well as the Ghangurde estimator (Ghangurde, 1989a) are alternative to the ratio estimator under the simple random sampling (SRS) design. Chambers (1986) proposed a robust alternative to the ratio estimator within the model-based framework. Hidioglou and Srinath (1981) as well as Fuller (1991) and Rivest (1994) propose robust alternatives to the sample mean under the SRS design. The Hulliger estimator (Hulliger, 1995) is a robust alternative to the Horvitz-Thompson estimator under the PPS sampling design. Kokic and Bell (1994) developed a robust estimator to be used as an alternative to the standard unbiased estimator under the stratified random sampling design. Therefore if one deals with a specific design, a suitable robust estimator might just not be available.

It was found that the generalized regression (GREG) estimation techniques under an arbitrary design, offers the common structure we need. Depending on the underlying superpopulation model and the sampling design, many well-known estimators can be seen as specific cases of the GREG estimator. Furthermore, the use of regression models makes easier the transition between the theory of robustness developed in the parametric statistical inference framework (see Huber, 1981 or Hampel et al., 1986) and the survey problems.

It is assumed that $\mathbf{x}_k$ is a q -dimensional vector of auxiliary variables associated with unit k belonging to the population $\mathcal{U}$ of size N . A sample s of size n is drawn following a given design $p(\cdot)$ and the variable of interest is denoted y . The inclusion probability of individual $i \in \mathcal{U}$ are given by

$$\pi_i = \sum_{s: i \in s} p(s),$$

and the join inclusion probability of individuals i and $j > i$ is given by:

$$\pi_{ij} = \sum_{s: i, j \in s} p(s).$$

The population $\mathcal{U}$ is supposed to be a simple random sample of size N drawn from the superpopulation described by the following regression model:

$$(\xi) \begin{cases} y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \quad i = 1, \dots, N, \\ E_\xi(\epsilon_i) = 0 \text{ and } V_\xi(\epsilon_i) = v_i, \quad \forall i \in \mathcal{U} \end{cases}$$

Let $\mathbf{B}$ be the weighted least squares estimator of $\boldsymbol{\beta}$ using the whole population. $\mathbf{B}$ is given by:

$$\mathbf{B} = \left(\sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i^\top / v_i \right)^{-1} \sum_{i=1}^N \mathbf{x}_i y_i / v_i.$$

Noting that the population $\bar{Y}$ can be written as

$$\bar{Y} = \frac{1}{N} \left(\sum_{i=1}^N y_{0i} + \sum_{i=1}^N E_i \right) \text{ where } y_{0i} = \mathbf{x}_i^\top \mathbf{B} \text{ and } E_i = y_i - y_{0i}.$$

the GREG estimator of the population mean is obtained by replacing the y_{0i} 's by predictors $\hat{y}_i$'s of the y_i 's and the total prediction errors $N\bar{E}$ by a suitable estimator.

We have that

$$\hat{\bar{Y}}_{\text{GREG}} = \sum_{i=1}^n \hat{y}_i / N + \sum_{i \in s} e_i / (N\pi_i) \text{ where } \hat{y}_i = \mathbf{x}_i^\top \hat{\mathbf{B}} \text{ and } e_i = y_i - \hat{y}_i.$$

The statistic $\hat{\mathbf{B}}$ is simply the estimator of $\mathbf{B}$ given by

$$\hat{\mathbf{B}} = \left(\sum_{i \in s} \mathbf{x}_i \mathbf{x}_i^\top / (v_i \pi_i) \right)^{-1} \sum_{i \in s} \mathbf{x}_i y_i / (v_i \pi_i).$$

It is usually assumed that the variable of interest y is well explained by the superpopulation model so that the residuals e_i 's be fairly small. But there is no guarantee that this will always be the case. As mentioned by Särndal et al. (1989), it is even possible for large residuals to occur simultaneously with large sampling weights, a situation which can be dramatic to the GREG estimator. Särndal et al. (1992, page 239) even recognize that "*For an ill-fitting model, the residuals E_i can be substantial and the variance ... can be large.*" In some specific cases however, the total prediction errors is known to be 0. That will lead to an estimator with some interesting robustness properties. In general, this is the case when $v_i = \boldsymbol{\lambda}^\top \mathbf{x}_i$, $i \in \mathcal{U}$ for some scalar vector $\boldsymbol{\lambda}$. The standard ratio estimator falls into this category.

Under some moment conditions on the y_i 's and on the $\mathbf{x}_i$'s as well as some mild regularity conditions on the inclusion probabilities, Robinson & Särndal (1983) have proved the design consistency of the GREG estimator. They also demonstrated that its asymptotic variance V_a is given by,

$$\begin{aligned} V_a(\hat{\bar{Y}}_{\text{GREG}}) &= V \left(\sum_{k=1}^n \frac{E_k}{N \pi_k} \right), \\ &= \sum_{\mathcal{U}} \sum_{\mathcal{U}} \left(\frac{\pi_{kl}}{\pi_k \pi_l} - 1 \right) E_k E_l. \end{aligned}$$

The above equations show that large residuals can have a dramatically high impact on the variance of the GREG estimator. Clearly, all these problems call for the

development of efficiency-robust estimation procedures with strong theoretical basis and a means to assess their performance. That will provide survey practitioners with a statistically sound method of handling aberrant values.

Another area where robust estimation is important is that of small domain estimation. When the number of sample units in a given domain is very small, direct estimators are usually very unstable. Since robust estimators, as mentioned previously perform well on small samples, it is natural to think that they can be efficient if applied to the estimation of small domain means. The use of M-statistics in small domain estimation is very limited and will be explored in this thesis. In the next section we review briefly the literature of small domain estimation.

1.4. Robustness in Small Area Estimation

The small area estimation problem can be seen by its very nature as a problem of robust estimation. The number of sample units falling into a given domain is usually very small to provide reliable estimates. This problem is more critical when the domain itself is small. To solve this problem, it is usually indicated to relate the small domain to other domains via modeling so as to compensate for the lack of data. This is the main approach advocated in the literature of small area statistics. However when the model fails, the performance of these estimators can be poor. Therefore a good estimator for small domains must be robust against model misspecification as the sample size is usually very small. An overview of

the existing methods can be found in Rao and Choudhry (1995) in the context of business surveys or in Ghosh and Rao (1994). In this section the focus will be on the estimators that will be dealt with during the study of robust estimation for small domains. In fact a substantial part of this thesis will be devoted to the efficiency robustness in small area estimation. The literature on the topic is not extensive.

It is assumed that the population $\mathcal{U}$ is made up with a finite number of domains $D, \mathcal{U}_1, \dots, \mathcal{U}_D$. A sample s of size n is drawn without replacement according to a given design $p(\cdot)$ independently of the domains. Afterwards estimates of the domain means $\bar{Y}_i, i = 1, \dots, D$ are requested, where

$$\bar{Y}_i = \frac{1}{N_i} \sum_{k=1}^{N_i} y_{ik}.$$

The characteristic of interest y takes the value y_{ik} on unit j of the i th domain. The sample s contains n_i units belonging to domain i . Some of the n_i 's might be 0. The small domain estimation problem stems from the fact that many of the n_i 's might be very small making the standard estimators (so called direct estimators) extremely unstable. Ghosh and Rao (1994) presents a comprehensive overview of the problem and some recent developments on the topic. To solve this problem, three main approaches have been advocated: (i) the direct methods that exploit the domain data only, (ii) methods that relate the small domain to other domains making implicit use of models and (iii) the model-based approach that relies upon a model such as that of Fay and Herriot (1979). These three approaches will be reviewed in the next three sections. We assume at the outset that a unit k is selected

with probability π_k and that units k and l will be selected with probability π_{kl} .

1.4.1. The Direct Methods

Many estimators are known in this context. The *expansion estimator* $\widehat{Y}_{E,i}$ as used for estimation at the population level is given by

$$\widehat{Y}_{E,i} = \sum_{k=1}^{n_i} \frac{y_{ik}}{N_i \pi_k} \Delta(n_i - 1),$$

where Δ is the indicator function defined for any real number a by $\Delta(a) = 1$ if $a \geq 0$ and $\Delta(a) = 0$ otherwise. This estimator is unbiased and performs very well for large samples but can offer a poor performance for small samples. A first level improvement was obtained by the *poststratified estimator* $\widehat{Y}_{P,i}$ given by

$$\widehat{Y}_{P,i} = \sum_{k=1}^{n_i} \frac{y_{ik}}{N_i \pi_k} \Delta(n_i - 1),$$

where $\widehat{N}_i$ is given by

$$\widehat{N}_i = \sum_{k=1}^{n_i} \frac{1}{\pi_k}.$$

This estimator possesses some interesting conditional properties when the inference is performed conditionally on the domain size $n_i \geq 1$. But it is a well-known fact that if the sample is very small, this estimator is likely to be very inefficient. These two estimators are seldom used in practice unless n_i is large enough. In this case, the properties of these estimators are similar to those of their counterparts at the population level.

1.4.2. Estimators Making Implicit Use of Models.

The estimators falling in this category are usually referred to as the one “borrowing strength” from related small areas. One such estimator is the *synthetic estimator*. We will give a general definition of the synthetic estimator within the Särndal’s(1984) generalized regression framework. Such a definition will encompass many versions of the synthetic estimator found in the literature.

We assume that for each population unit k , there exists a vector $\mathbf{x}_k$ of auxiliary variables which is linked to the variable of interest y through the following superpopulation model:

$$(\xi) \begin{cases} y_k = \mathbf{x}_k^\top \boldsymbol{\beta} + \epsilon_k, \text{ where} \\ \forall k \in \mathcal{U}, E_\xi(\epsilon_k) = 0 \text{ and } V_\xi(\epsilon_k) = v_k. \end{cases}$$

If the population under study is assumed to be a simple random sample of size N drawn from the superpopulation, the weighted least squares estimator of $\boldsymbol{\beta}$ is given by:

$$\mathbf{B} = \left(\sum_{k=1}^N \mathbf{x}_k \mathbf{x}_k^\top / v_k \right)^{-1} \sum_{k=1}^N \mathbf{x}_k y_k / v_k.$$

This parameter is usually estimated from a sample s by the following estimator:

$$\hat{\mathbf{B}} = \left(\sum_{k \in s} \mathbf{x}_k \mathbf{x}_k^\top / (v_k \pi_k) \right)^{-1} \sum_{k \in s} \mathbf{x}_k y_k / (v_k \pi_k).$$

Under some mild regularity conditions on the inclusion probabilities and on the

moments of the variable of interest (see Isaki and Fuller, 1982) it is possible to show that $\hat{\mathbf{B}}$ is a consistent estimator of $\mathbf{B}$.

The synthetic estimator of the i – th domain mean $\bar{Y}_i$ is given by

$$\hat{\bar{Y}}_{s,i} = \frac{1}{N_i} \sum_{k=1}^{N_i} \mathbf{x}_{ik}^\top \hat{\mathbf{B}}. \quad (1.3)$$

If we assume the ratio model, that is the existence of a single auxiliary variable x , a variance term v_k that is proportional to x_k and the simple random sampling design without replacement, then the above estimator takes the more familiar form that is usually found in the literature of small area estimation:

$$\hat{\bar{Y}}_{s,i} = \bar{X}_i(\bar{y}/\bar{x}).$$

This synthetic estimator will perform remarkably well if the ratio model applies and the area ratios $\bar{Y}_i/\bar{X}_i$ is approximately equal to the population ratio $\bar{Y}/\bar{X}$. The between-strata homogeneity without the ratio model will lead to a consistent estimator which can yield good performance only for moderate to large samples.

More generally, the general synthetic estimator 1.3 will perform very well if the following two conditions are satisfied:

- (i) the model m with $v_k = \boldsymbol{\lambda}' \mathbf{x}_k \ \forall k \in \mathcal{U}$ holds.
- (ii) For each small domain i , $\mathbf{B}_i$ is approximately equal to $\mathbf{B}$ where

$$\mathbf{B}_i = \left(\sum_{k \in \mathcal{U}_i} \mathbf{x}_{ik} \mathbf{x}_{ik}^\top / v_{ik} \right)^{-1} \sum_{k \in \mathcal{U}_i} \mathbf{x}_{ik} y_{ik} / v_{ik}.$$

If these two conditions are not satisfied, the synthetic estimator can be heavily biased losing thereby most of the efficiency induced by a small variance.

In order to be protected against the potential bias of the synthetic estimator, Särndal (1984) proposes to adjust the synthetic estimator for the bias. Such an adjustment leads to the following approximately unbiased regression-type estimator:

$$\widehat{Y}_{Ri} = \sum_{k=1}^{N_i} \mathbf{x}_{ik}^T \widehat{\mathbf{B}} / N_i + \sum_{k \in s_i} \frac{e_{ik}}{N_i \pi_k} \Delta(n_i - 1),$$

where $e_{ik} = y_{ik} - \mathbf{x}_{ik}^T \widehat{\mathbf{B}}$ is the residual. However, it is a well-known fact that the adjustment term $\bar{e}_i$ can be very unstable for small samples. In fact it estimates the parameter

$$\bar{E}_i = \frac{1}{N_i} \sum_{k=1}^{N_i} E_{ik},$$

where $E_{ik} = y_{ik} - \mathbf{x}_{ik}^T \mathbf{B}$, without borrowing strength from other small areas. Hidioglou and Särndal (1985) suggest to estimate the total of prediction errors by a post-stratified estimator to obtain

$$\widehat{Y}_{Rpi} = \sum_{k=1}^{N_i} \mathbf{x}_{ik}^T \widehat{\mathbf{B}} / N_i + \sum_{k \in s_i} \frac{e_{ik}}{\widehat{N}_i \pi_k} \Delta(n_i - 1).$$

This estimator is likely to yield some improvement over $\widehat{Y}_{Ri}$. However, the use of other areas is so limited that the gain will probably be small. Särndal and Hidioglou (1989) propose the “*dampened regression estimator*” given by

$$\widehat{Y}_{RDi} = \sum_{k=1}^{N_i} \mathbf{x}_{ik}^T \widehat{\mathbf{B}} / N_i + \left(\frac{\widehat{N}_i}{N_i} \right)^H \sum_{k \in s_i} \frac{e_{ik}}{\widehat{N}_i \pi_k} \Delta(n_i - 1),$$

where $H = 0$ if $\hat{N} \geq N$ and $H = 2$ otherwise. Note that this estimator can be written as

$$\hat{Y}_{RDi} = \left(\frac{\hat{N}_i}{N_i} \right)^H \hat{Y}_{Rpi} + \left[1 - \left(\frac{\hat{N}_i}{N_i} \right)^H \right] \hat{Y}_{si}.$$

As pointed out by Rao and Choudhry (1995), one disadvantage of this estimator is that it does not exploit the stability of the synthetic estimator when $\hat{N}_i \geq N_i$ even if the domain of interest is very small. This seems to be the price to be paid for conditional design-unbiasedness.

Other authors have also proposed estimators by taking a weighted average of the synthetic estimator and an unbiased or approximately unbiased estimator. These estimators are called *composite estimators*. Schaible et al.(1977) suggested to determine the weights so as to minimize the mean squared error. These optimal weights are usually very unstable.

Given the lack of data in small domains, some authors advocate the pure model-based approach. These estimators are based on realistic domain models and aim at achieving high efficiency in terms of the mean squared error. Some of the robustness problems incidental to the use of these domain models will be reviewed in the next section.

1.4.3. Model-Based Methods

Linear models are widely used in small area methodology as a means to obtain reliable estimates in small areas. This problem stems from the fact that the sample.

usually drawn from a population covering many small areas cannot provide enough units in a specific small area to produce reliable direct estimates. However, model-based methods always raise the problem as to the adequacy of the assumed models. This problem is mentioned in the context of small area estimation by Ghosh and Rao (1994). More generally, the problem of outlying sample units in the context of mixed linear models is discussed by Christensen, Pearson and Johnson (1992) as well as by Fellner (1986).

Model-checking in standard regression analysis is usually performed using at least one of the many diagnostic tools available in the literature. The book of Cook and Weisberg (1982) gives a good account of the most widely-used diagnostic tools. However, very little has been done for linear models involving random effects. In order to fill this gap, Christensen, Pearson and Johnson (1992) propose some extensions of the well-known diagnostics to mixed linear models. Their approach will be discussed later.

Another approach, as mentioned above in a different context, is based on M-statistics. The theory of M-statistics was originally developed as an alternative to the least squares principle. Such an alternative is needed as the least squares principle besides simplicity can be rigorously justified only under normality. Our intention to apply M-statistics to small area estimation is motivated by the use of the least squares principle in deriving the Best Linear Unbiased Predictor of a

small area mean. An attempt to apply the theory of robust statistics to small area estimation is presented by Fellner (1986).

Let us assume the following general mixed linear model:

$$(\xi) \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{e} \quad (1.4)$$

where $\mathbf{y}$ is the $n \times 1$ -vector of sample observations, $\mathbf{X}$ and $\mathbf{Z}$ are respectively $n \times p$ and $n \times q$ known matrices. It is further assumed that $\boldsymbol{\gamma}$ and $\mathbf{e}$ are distributed independently with means $\mathbf{0}$ and covariance matrices $\mathbf{G}$ and $\mathbf{R}$ respectively.

Let r be the number of random effects in model 1.4 . It is also possible to express model 1.4 as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \sum_{j=1}^r \mathbf{Z}_j \boldsymbol{\gamma}_j + \mathbf{e} \quad (1.5)$$

where $\mathbf{Z}_j$ is an $n \times q_j$ -matrix and $\boldsymbol{\gamma}_j$ is a $q_j \times 1$ -vector. For simplicity, the following two assumptions are made:

$$\mathbf{G} = \text{diag}[\sigma_1^2 \mathbf{I}_{q_1}, \dots, \sigma_m^2 \mathbf{I}_{q_m}],$$

$$\mathbf{R} = \sigma_e^2 \mathbf{I}_n$$

Most small area models are special cases of this general model. A comprehensive review of these models is presented in Ghosh and Rao (1994).

1.4.4. Case-Deletion Diagnostics

Using model (1.5), Christensen, Pearson and Johnson (1992) extend to mixed linear models some diagnostic tools based on case-deletion and originally developed

in standard regression analysis. Their ultimate goal was to develop the equivalent of the Cook's distance to measure the influence of a single unit on $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$, the optimal estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ respectively. For the fixed effects, they derived the following Cook's distance:

$$C_i = p^{-1}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)})^\top (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}) \quad (1.6)$$

where

$$\mathbf{V} = \sigma_e^2 \mathbf{I}_n + \sum_{j=1}^m \sigma_j^2 \mathbf{Z}_j \mathbf{Z}_j^\top,$$

and $\hat{\boldsymbol{\beta}}$ is the Generalized Least Squares estimator (GLS) of $\boldsymbol{\beta}$ given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$$

For computational purposes, the authors proposed for the Cook's distance, a much more simplified formula. Let $\mathbf{V}_{(i)}$ be the matrix obtained after removing the i -th row and the i -th column and $\mathbf{v}_i$, the i -th column of $\mathbf{V}$ without the element v_{ii} . The simplified Cook's distance is then given by:

$$C_i = (y_i^* - \mathbf{x}_i^{*\top} \hat{\boldsymbol{\beta}})^2 \cdot \frac{h_i^*}{p(s_i - h_i^*)^2}$$

where

$$\begin{cases} \mathbf{x}_i^* = \mathbf{x}_i - \mathbf{X}_{(i)}^\top \mathbf{V}_{(i)}^{-1} \mathbf{v}_i \\ h_i^* = \mathbf{x}_i^{*\top} (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{x}_i^* \\ y_i^* = y_i - \mathbf{y}_{(i)}^\top \mathbf{V}_{(i)}^{-1} \mathbf{v}_i \\ s_i = v_{ii} - \mathbf{v}_i' \mathbf{V}_{(i)}^{-1} \mathbf{v}_i \end{cases}$$

These authors also proposed another Cook's distance for the variance components similar to the previous one by replacing β by vector $\sigma^{2\top} = (\widehat{\sigma_e^2}, \widehat{\sigma_1^2}, \dots, \widehat{\sigma_m^2})$ where the $\widehat{\sigma_i^2}$'s are the restricted maximum likelihood estimator the variance components. These distances have the disadvantage of detecting only one outlier at a time. Other disadvantages will be mentioned in the next section.

The next section provides a review of the use of M-statistics with mixed linear models.

1.4.5. Robust Methods

The case-deletion diagnostics suggested for mixed linear models suffer from the same drawbacks as those suggested in standard regression analysis. These diagnostic tools are vulnerable to the so-called *masking and swamping effects*. If two outliers are close to each other, removing one from the sample will not affect the model parameters significantly (masking effect), the outlier in the sample has masked the removed one preventing it from being detected. The swamping effect will occur if the impact of an outlier on the adjustment is big enough to make some "normal units" look outlying. All these problems have been mentioned very often in the robustness literature.

In an attempt to develop more robust procedures for mixed linear models, Fellner (1986) proposed M-estimation-based alternatives to the Best Linear Unbiased Prediction estimators of the model parameters proposed by Henderson (1963). Fellner

also proposed a robust alternative to the restricted maximum likelihood estimation technique for variance components estimation. Although Fellner's main concern was the estimation of the variance components, his approach can be adapted to provide robust estimators of small area means.

Fellner's contribution to robust estimation in mixed linear models although carried out within the framework of the general mixed linear model given in 1.5, is described here in the context of small area estimation and under a special case of (1.5). Suppose that there are m small areas of interest. Each small area i is made up of N_i population units. A sample of size n is selected according to a certain sampling design. Let n_i be the number of sample units that fell into small area i and $\mathbf{x}_{ij}$ a p -dimensional vector of auxiliary variables known for each unit j in small area i . The variable of interest is y and it is assumed that

$$(\xi) \quad y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \gamma_i z_i + e_{ij},$$

where the γ_i 's are independent and identically distributed random variables with mean 0 and variance σ_γ^2 , representing the area effect and the z_i 's are known constants. The e_{ij} 's are also independent and identically distributed with mean 0 and common variance σ_e^2 . Moreover the γ_i 's and the e_{ij} 's are assumed to be independent. This small area model which was suggested by Battese, Harter and Fuller (1988) is a special case of 1.5 and will provide the framework for the development of our proposed robust estimators in chapter 5.

Note that the Battese-Harter-Fuller model can be obtained from the general model

(1.5) by taking $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_m)^\top$ where $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})^\top$.

$$\mathbf{X} = \text{col}_{1 \leq i \leq m} (\mathbf{X}_i)$$

where

$$\mathbf{X}_i = \text{col}_{1 \leq j \leq n_i} (\mathbf{x}_{ij}^\top).$$

Since there is only one random effect ($r = 1$), $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_m)^\top$ and $\mathbf{Z}_1$ is given by

$$\mathbf{Z}_1 = \begin{pmatrix} z_1 \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \dots & \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} & \ddots & & \\ \vdots & & \ddots & \\ \mathbf{0}_{n_m} & \dots & \mathbf{0}_{n_m} & z_m \mathbf{1}_{n_m} \end{pmatrix}$$

Usually the quantities to be estimated are $\boldsymbol{\beta}$, the realized value of $\boldsymbol{\gamma}$ and the vector of variance components $\boldsymbol{\sigma} = (\sigma_\gamma, \sigma_e)$. For now, we assume that the variance components are known. We only have to estimate $\boldsymbol{\beta}$ and the realized value of $\boldsymbol{\gamma}$. The BLUP estimation technique suggested by Henderson (1963) provides two linear unbiased estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ as solution to the following system of normal equations:

$$\begin{cases} \sum_{i=1}^m \sum_{j=1}^{n_i} \mathbf{x}_{ij} \mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_H + \sum_{i=1}^m \sum_{j=1}^{n_i} \mathbf{x}_{ij} z_i \hat{\gamma}_{Hi} = \sum_{i=1}^m \sum_{j=1}^{n_i} \mathbf{x}_{ij} y_{ij}, \\ \alpha_i \bar{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}_H + z_i \hat{\gamma}_{Hi} = \alpha_i \bar{y}_i \end{cases} \quad (1.7)$$

where $\alpha_i = \sigma_\gamma^2 z_i^2 (\sigma_\gamma^2 z_i^2 + \sigma_e^2 / n_i)^{-1}$. The first step towards the development of robust estimators for the variance components is find suitable robust alternatives to $\hat{\boldsymbol{\beta}}_H$ and $\hat{\gamma}_{Hi}$. To this end, Fellner exploited the concept of “pseudo-observations” introduced by Bickel (1976, p. 176). Bickel (1976) observed that it is possible to calculate the M-estimators of Huber using standard least squares programs simply

by changing the original observations y_{ij} in equation (1.7) by the so-called “pseudo-observations”. To define the pseudo-values, let ψ be the influence function to be used. The function ψ can be of Huber type ψ_H defined for any real number t by

$$\psi_H(t) = \text{sgn}(t) \min(K_H, |t|),$$

K_H being the robustness constant. The pseudo-value associated with y_{ij} is given by

$$y_{ij}^* = \mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_F + z_i \hat{\gamma}_{Fi} + \sigma_e \psi(\hat{e}_{ij}/\sigma_e),$$

where $\hat{e}_{ij} = y_{ij} - (\mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_F + z_i \hat{\gamma}_{Fi})$ and F stands for Fellner.

Fellner (1986) identified another outlier problem that does not exist in standard regression analysis and that is important to mention. The parameters γ_i 's are assumed to have mean 0 and variance σ_γ^2 . It is then natural for a robust procedure to prevent $\hat{\gamma}_{Fi}/\sigma_\gamma$ from being much greater than 0. Since the right-hand side of the second equation in (1.7) can be read as $\alpha_i \bar{y}_i + 0$, Fellner suggested to replace it with the pseudo-value $\alpha_i \bar{y}_i^* + 0_i^*$ where $\bar{y}_i^*$ and 0_i^* are respectively given by

$$\bar{y}_i^* = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}^*$$

and $0_i^* = \hat{\gamma}_{Fi} - \sigma_\gamma \psi(\hat{\gamma}_{Fi}/\sigma_\gamma)$. Therefore Fellner robust estimators $\hat{\boldsymbol{\beta}}_F$ and $\hat{\gamma}_{Fi}$ are solution to the following system of equations:

$$\begin{cases} \sum_{i=1}^m \sum_{j=1}^{n_i} \mathbf{x}_{ij} \mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_F + \sum_{i=1}^m \sum_{j=1}^{n_i} \mathbf{x}_{ij} z_i \hat{\gamma}_{Fi} = \sum_{i=1}^m \sum_{j=1}^{n_i} \mathbf{x}_{ij} y_{ij}^*, \\ \alpha_i \bar{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}_H + z_i \hat{\gamma}_{Hi} = \alpha_i \bar{y}_i^* + z_i 0_i^* \end{cases} \quad (1.8)$$

Fellner has proposed an iterative algorithm for calculating these robust estimators. If the variance components are assumed to be unknown and must be estimated, then Fellner uses $\hat{\boldsymbol{\beta}}_F$ and $\hat{\gamma}_{Fi}$ in a two-stage procedure which is a modification of the restricted maximum likelihood approach. A similar approach was taken by Richardson and Welsh (1995). Some simulation results presented by Fellner (1986) indicate a substantial improvement over the classical restricted maximum likelihood method. However his method is vulnerable to high leverage points as he acknowledged. It is possible to solve this problem by applying the generalized M-estimation technique to mixed linear models. This technique was used successfully by Gwet and Rivest (1992) in the context of ratio estimators.

In this section, we have mentioned a number of robustness problems incidental to the use of small area models. Fellner (1986) has proposed a robust estimation methodology for estimating the variance components in general mixed linear models that is promising. Our intention is to follow up these ideas and propose robust procedures for estimating small area means.

1.5. Thesis Outline

In chapter 2, general robust procedures are proposed for estimating finite population means under the common mean superpopulation model. It is then assumed that no auxiliary information is available. A general theory will be presented under an arbitrary sampling design and some specific cases will be explored as well. An emphasis is on the conditions under which the proposed robust procedures are

consistent for estimating the parameter of interest.

The traditional M-estimator of a location parameter will be studied first under a general sampling design and under the assumption that the superpopulation (not the finite population) has a symmetric structure. It will be shown the M-estimator is a consistent estimator for estimating the finite population mean under some regularity conditions and some restrictions on the superpopulation. This will make more precise the conditions under which the M-estimator can be used in sample surveys.

Afterwards the symmetry assumption on the superpopulation is relaxed. We then propose a consistent M-estimator or CM-estimator for estimating the population mean. The consistency of the CM-estimator unlike that of the M-estimator does not rely upon any restriction on the superpopulation model. The price to be paid for this model-free consistency will be more sensitivity to large sample units. But for skewed populations, the gain in efficiency on standard procedures could still be substantial. This estimator is safer than the M-estimator when there is no evidence supporting the symmetry assumption on the superpopulation.

Another robust consistent estimator called LM-estimator is proposed in chapter 2. It uses an M-estimator as a first approximation to the population mean and then performs an adjustment with an estimation of the mean of the prediction errors via an L-estimator such as the trimmed mean. The use of M- and L-statistics led to naming LM-estimator. This estimator is introduced here but it will be most

useful when auxiliary variables are used at the estimation stage. Because the CM-estimation approach will not work under any superpopulation model.

Chapter 3 is essentially devoted to the extension of the theory to more complex superpopulation models. Many theoretical results mentioned in chapter 2 are proved in chapter 3 in a more general setting. Some practical considerations are also examined here. Among others we study the calibration properties of the proposed robust procedures. We will prove that the M-estimator, the CM-estimator and the LM-estimator are all calibration estimators in the sense of Deville & Särndal (1992). The general technique for deriving calibration estimators suggested by Deville & Särndal (1992) is adapted here to generate robust calibration estimators of the form of the LM-estimator. Another practical issue addressed in this chapter is the use of robust procedures in large-scale multipurpose surveys. A proposal for using common weights for many variables of interest is discussed.

We will study the proposed general estimators under some well-known sampling designs and superpopulation models. It will be seen that well-known robust estimators such as the Gwet-Rivest M-estimator under the ratio model or the robust Horvitz-Thompson estimator of Hulliger (1995) can be derived from our general theory.

In chapters 2 and 3 the approximate variance of the proposed estimator is obtained by the linearization method. The derivation of the variance formula is tedious and the simulation results in chapter 2 will indicate that the resulting variance

estimators underestimate the true variance especially of the CM-estimator. In an attempt to improve on the linearization method, we propose in this chapter. the jackknife variance estimator for each of our robust procedures. It is shown that the jackknife variance estimator is a consistent estimator of the asymptotic variance. A form of jackknife estimator of the squared bias is proposed as well and can be used to estimate the mean squared error for small samples.

The simulation results presented in this chapter indicate a substantial gain in efficiency of the jackknife estimator of the variance of the CM-estimator. The jackknife does not lead to a striking improvement for the LM-estimator.

Chapter 5 is entirely devoted to the estimation of small area means within the model-based framework. A robust alternative to the Best Linear Unbiased Prediction estimator of small area means is proposed under the assumption of known variance components. A first-order approximation to the mean squared error of the estimator is derived. It is shown theoretically as well as empirically that the use of the GM-estimation technique in small area is promising. The gain in efficiency of the proposed GM-estimator over the BLUP estimator is substantial when the number of small areas is small. For a large number of small areas the BLUP estimator perform very well.

In the cases where the variance components are unknown, we propose a robust alternative to the method of fitting constants. The large sample properties of these estimators are not studied here. But when the variance components are unknown

we expect the proposed GM-estimator to perform even better.

In chapter 6, a summary of the thesis is provided as well as some suggestions for further research.

CHAPTER 2

Robust Estimation under the Common Mean Superpopulation Model

2.1. Introduction

The problem addressed here is to estimate the population mean of a given characteristic y exploiting the simplest superpopulation model possible. That model is the one where all population units share a common theoretical mean β . The regression estimator is available in this context and can be used. However this estimator is inefficient for skewed populations because of its sensitivity to large units. Our goal is to develop estimators that will be more stable for skewed populations while sharing the consistency property of the regression estimator. The price to be paid for the robustness will be a bias that goes more slowly to 0 for large samples. But even in that case, the mean square error of the robust estimators will in general be

much smaller than the variance of the regression estimator.

The population $\mathcal{U}$ under study is finite and made up of N distinct units labeled as $k = 1, \dots, N$. The problem to be solved is that of estimating the mean $\bar{Y}$ of the variable of interest y , given by

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N y_i.$$

Let $\mathcal{S}_n$ be the set of all possible samples of size n that can be drawn from $\mathcal{U}$. The estimation of $\bar{Y}$ is based on a sample s of fixed size n selected from $\mathcal{U}$ according to an arbitrary sampling design $p(\cdot)$ which is a function from $\mathcal{S}_n$ to $[0, 1]$. The probability of selecting a specified sample $s \in \mathcal{S}_n$ is $p(s)$. Under a given sampling design, a population unit k will be included in the sample s with probability π_k given by

$$\pi_k = \sum_{s \in \mathcal{S}_n : s \ni k} p(s).$$

Two population units k and l will be included in the sample s with probability π_{kl} given by,

$$\pi_{kl} = \sum_{s \in \mathcal{S}_n : s \ni k, l} p(s).$$

The π_k 's and π_{kl} 's are the first-order and second-order inclusion probabilities respectively.

Using the model-assisted approach, the population is assumed to have been gener-

ated by the following superpopulation model:

$$(\xi_0) \quad \begin{cases} y_i = \beta + \epsilon_i, \\ E_{\xi_0}(\epsilon_i \epsilon_j) = 0, \text{ if } i \neq j, \\ E_{\xi_0}(\epsilon_i^2) = \sigma^2. \end{cases}$$

This common mean model is appropriate when there is no auxiliary variable that is related to the variable of interest y . In such a case, the natural sampling design is Simple Random Sampling (SRS) where all samples have the same probability of selection. However, in multipurpose surveys a few key variables of interest might be highly correlated with some auxiliary variables justifying thereby, the use of complex designs under the common mean superpopulation model.

In section 2.2, we present an overview of the existing procedures for the estimation of the finite population mean. The regression estimator under the common mean superpopulation model is discussed in section 2.2.1 and some of its most important asymptotic properties reminded. In section 2.2.2, the Gwet-Rivest M-estimator is discussed in the present set-up although originally developed in the context of ratio models (Gwet & Rivest, 1992). More recently Hulliger (1995) proposed a robust alternative to the Horvitz-Thompson estimator that is also mentioned in section 2.2.3.

In section 2.3, we propose an extension of M-estimators to complex samples under the assumption that the finite population is a simple random sample of size N from a **symmetric** superpopulation. In sub-section 2.3.1, we define the complex

M-estimator. Sub-section 2.3.2 is devoted to large-sample properties of this M-estimator.

In section 2.4 two additional robust alternatives to the GREG estimator to be used for asymmetric populations are proposed. Our goal is to improve on the Gwet-Rivest approach in the sense of reducing the influence of large units on the estimator while ensuring consistency. We will call the resulting estimator, the CM-estimator where “C” stands for consistency. The CM-estimator is only applicable to the case where the GREG estimator takes its simple (or cosmetic) form. It is obtained by replacing the least squares principle, not by the bounded influence principle as in the case of the M-estimator, but by an hybrid principle defined by an influence function that is obtained as a weighted linear combination of the identity function and the bounded robustness function.

In order to deal with the cases where the GREG estimator does not take the simple form, we propose another robust estimator in sub-section 2.4.2. It is obtained in two stages: (i) robust predictors of the y_i 's using the M-estimation technique are obtained first and (ii) the total prediction error is estimated by means of a trimmed mean (which is an L-estimator). The use of L-statistics in survey sampling was discussed recently by Shao (1994). Some of his results will be used in the course of our theoretical investigation in sub-sections 2.4.3 and 2.4.4. The combination of M and L estimation techniques has led us to call the proposed estimator, LM-estimator and not ML as the latter usually refers to maximum likelihood estimators.

Some simulation results are presented in section 2.4

2.2. Review of Some Existing Procedures

As mentioned in the previous chapter, a number of estimators that could be used for estimating the finite population mean have been proposed in the literature. Some of these estimators are related to some extent to our proposals and will be briefly discussed here within the framework of the common mean superpopulation model.

2.2.1. The GREG Estimator

The GREG estimator derived from the model (ξ_0) is given by

$$\hat{\bar{Y}}_G = \hat{B}_G + \sum_{i=1}^n \frac{e_{Gi}}{\hat{N}\pi_i} \text{ where } \begin{cases} e_{Gi} = y_i - \hat{B}_G, \\ \hat{B}_G = \sum_{i=1}^n \frac{y_i}{\hat{N}\pi_i} \text{ and } \hat{N} = \sum_{i=1}^n 1/\pi_i. \end{cases}$$

In this particular case, this estimator can be written as

$$\hat{\bar{Y}}_G = \sum_{i=1}^n \frac{y_i}{\hat{N}\pi_i}.$$

Under general conditions on the inclusion probabilities, it is possible to show that the asymptotic variance of the GREG estimator is given by

$$V(\hat{\bar{Y}}_G) = \sum_{i=1}^N \sum_{j=1}^N \frac{\pi_{ij} - \pi_i\pi_j}{\pi_i\pi_j} E_{Gi}E_{Gj}, \text{ where } E_{Gi} = y_i - \bar{Y}.$$

This estimator is asymptotically unbiased and the validity of the inference is quite robust to any model misspecification. This is one of the most important arguments in favor of this model-assisted procedure. However, the price to be paid for achieving

this model-free validity can be dramatically high. If the population contains a few outlying units, some of the E_{Gi} 's will be large increasing thereby the asymptotic variance as will be seen in the simulation study presented in section 5.

In an attempt to solve this problem, Gwet and Rivest (1992) have used the theory of robust statistics for estimating the finite population mean. The Gwet-Rivest framework is that of the ratio model under the simple random sampling design. The Gwet-Rivest approach is essentially the design-based interpretation of Huber's M-estimator (see Huber, 1964).

2.2.2. The Gwet-Rivest M-Estimator

The M-estimation approach relies upon the following idea: if the estimator to be "robustified" is the least squares or the weighted least squares estimator of a model parameter, then its robust counterpart can be obtained by replacing the least squares principle with the bounded influence principle. M-estimation approach can also be motivated as follows. The contribution of the large units to the value of the parameter of interest $\bar{Y}$ is usually substantial. It seems intuitively difficult to estimate a changing parameter by a pretty stable estimator. For that reason, one approach might be to define a new population parameter B_M on which the effect of large units will be less important and which hopefully will not be too far from the parameter of interest. An asymptotically unbiased estimator of B_M will be a good estimator of $\bar{Y}$ if the two parameters are not too different.

Royall (1970) shows that under the ratio model, the ratio estimator is the best linear unbiased estimator of the population mean. His finding was used by Gwet and Rivest (1992) to derived a resistant alternative to the ratio estimator under the simple random sampling (SRS) design.

Let us assume at the outset that the ϵ_i 's are symmetrically distributed. This is a customary assumption that underlies the theory of robust statistics as developed by Huber (1964). Under the common mean superpopulation model, the Gwet-Rivest approach yields the following estimator of $\bar{Y}$, $\hat{Y}_{GR} = \hat{B}_M$, where $\hat{B}_M$ is implicitly defined by the following equation.

$$\sum_{i=1}^n \psi \left(\frac{y_i - \hat{B}_M}{\sigma} \right) = 0.$$

For the sake of simplicity, it is assumed here that the scale parameter σ is known. Robustness considerations require that the ψ function, sometimes called the influence function, satisfy some standard conditions that are well known (see Huber. 1964).

Many influence functions have been proposed in the literature and address various needs. One popular influence function was proposed by Huber (1964) and is defined by,

$$\psi_H(t) = \text{sgn}(t) \min(K, |t|),$$

for a robustness constant K to be specified. The popularity of this influence function stems essentially from its simplicity. Another interesting influence function.

sometimes called the “FAIR” function, is defined by

$$\psi_F(t) = \frac{t}{1 + |t|/K}.$$

A good feature of this function is that it has everywhere continuous derivatives. Unlike the Huber’s proposal which is not everywhere differentiable. Moreover, it is a well-known fact that ψ_F is not very sensitive to the choice of the scale parameter σ . This property is important in practice, as σ will in general be replaced by a robust estimator that is not usually easy to justify.

The fact that the Huber’s function is not differentiable everywhere is somewhat troublesome in simulation studies using the Newton-Raphson algorithm for example. For that reason, it is sometimes more convenient to use its modified version proposed by Rey (1983):

$$\psi_R(t) = \begin{cases} K \sin \frac{t}{K}, & \text{if } |t| \leq \frac{K\pi}{2}, \\ \text{sgn}(t)K, & \text{if } |t| \geq \frac{K\pi}{2}. \end{cases}$$

Gwet and Rivest (1992) have studied the design-based properties of their estimator unconditionally as well as conditionally on the number of outliers in the sample. The conclusion is that this estimator can yield substantial gain in efficiency over the standard GREG estimator if the superpopulation model provides a reasonable description of the population under study and the true distribution of the error term is outlier-prone (has long tails).

In the next chapter, we propose a general approach for deriving robust alternatives to standard estimators that belong to the class of regression estimators advocated

by Särndal et al. (1992). Only estimation problems will be addressed. The design problems are not taken up and the proposed procedures will only be resistant to outliers generated by the superpopulation model, not by the design.

2.3. M-estimation for Complex Samples: the Symmetric Case

The applications of the theory of robust statistics to survey problems proposed to date are limited to the simple random sampling design. One exception is the “robustified” Horvitz-Thompson estimator proposed by Hulliger (1995). But Hulliger extension was limited to the probability-proportional-to-size (PPS) sampling design and the underlying superpopulation model was the ratio model, the same model considered by Gwet and Rivest. In this section, we propose an extension of M-estimators to more general sampling designs. Since the form of the finite population is relevant for proving some asymptotic properties of the M-estimator, the type of inference used here is the one that takes into account both the model ξ_0 and the sampling design $p(\cdot)$.

2.3.1. The M-estimator

The GREG estimator under model ξ_0 when the sample is selected according to an arbitrary design was defined in the previous section. It can be seen as the least squares estimator $\hat{B}_G$ of β ; that is as solution to the following equation:

$$\int \left(\frac{y - \hat{B}_G}{\sigma} \right) d\hat{F}(y) = 0, \text{ where } \hat{F}(t) = \sum_{i=1}^n \frac{I(y_i \leq t)}{\hat{N}\pi_i}.$$

This method of deriving the GREG estimator provides a natural way of extending

the Gwet-Rivest M-estimator to a general design. Such an extension has not been studied by Gwet & Rivest (1992). A robust alternative $\widehat{Y}_M$ of $\widehat{Y}_G$ is then given for a known σ , by $\widehat{Y}_M = \widehat{B}_M(\sigma)$ where $\widehat{B}_M(\sigma)$ is obtained as solution to the equation

$$\int \psi_H\left(\frac{y - \widehat{B}_M(\sigma)}{\sigma}\right) d\widehat{F}(y) = 0,$$

This equation can also be written as

$$\sum_{i=1}^n \psi_H\left(\frac{y_i - \widehat{B}_M(\sigma)}{\sigma}\right) / \pi_i = 0.$$

We recall that $\psi_H(\cdot)$ is defined by,

$$\psi_H(t) = \text{sgn}(t) \min(K, |t|), \quad \forall t \in \mathbb{R}.$$

In the robustness literature, the robustness constant K is usually determined so as to achieve a 95% efficiency on good data (under normality) with respect to the standard procedure. In this context, we suggest a robustness constant K so that if the distribution of the ϵ_i 's is normal and N is large,

$$V_{\xi_0}(\overline{Y}) / V_{\xi_0}(B_M) \approx 95\%,$$

where B_M is the population parameter given by (2.1). Let K_{95} be such a constant. When the distribution of errors in the superpopulation model is symmetric with long tails or is a contaminated normal distribution then the M-estimator can yield a substantial gain in efficiency.

Very often, the nuisance parameter σ is unknown and has to be estimated by means of an estimator $\widehat{\sigma}_M$. The solution is usually to calculate $\widehat{B}_M$ and $\widehat{\sigma}_M$ simultaneously

as follows: if the population is assumed to be a simple random sample drawn from the superpopulation model ξ_0 then σ can be estimated by:

$$\sigma_M = C_\Phi \text{ med}_{1 \leq i \leq N} |R_i|, \text{ where } R_i = y_i - B_M \text{ and } C_\Phi = 1/\Phi^{-1}(0.75).$$

Note that Φ is the cumulative distribution function of the standard normal distribution and the parameter B_M is implicitly defined by the following estimating equation.

$$\sum_{i=1}^N \psi_H \left(\frac{y_i - B_M}{\sigma_M} \right) = 0. \quad (2.1)$$

The finite population parameter σ_M will then be estimated from a sample by a procedure such as the one described below. This parameter is the Median Absolute Deviation (MAD) of the y_i 's multiplied by the adjustment factor to achieve consistency under normality. It was suggested in the framework of parametric statistical inference by many authors (see Hampel et al. 1986, p.105).

Definition 2.1. Let $\hat{\sigma}_M$ be a consistent estimator of σ_M and the statistic $\hat{B}_M$ be implicitly defined by the following estimating equation,

$$\sum_{i=1}^n \psi_H \left(\frac{y_i - \hat{B}_M}{\hat{\sigma}_M} \right) / \pi_i = 0,$$

The M-estimator $\hat{\bar{Y}}_M$ of $\bar{Y}$ is given by,

$$\hat{\bar{Y}}_M = \hat{B}_M.$$

The M-estimator can easily be obtained by performing a few iterations of

$$\widehat{B}_M^{(t+1)} = \sum_{i=1}^n \frac{w_i^{(t)} y_i}{\pi_i} \bigg/ \sum_{i=1}^n \frac{w_i^{(t)}}{\pi_i},$$

where

$$w_i^{(t)} = \psi_H \left(\frac{y_i - \widehat{B}_M^{(t)}}{\widehat{\sigma}_M^{(t)}} \right) \bigg/ \left(\frac{y_i - \widehat{B}_M^{(t)}}{\widehat{\sigma}_M^{(t)}} \right), \quad \widehat{\sigma}_M^{(t)} = C_\Phi \widehat{M}_R^{(t)},$$

and $\widehat{M}_R$ is an estimator of M_R given by.

$$M_R = \text{med}_{1 \leq i \leq N} |y_i - B_M|.$$

To define the estimator $\widehat{M}_R$, let us introduce some notations. The empirical robust residual r_i of the i -th sample unit is given by $r_i = y_i - \widehat{B}_M$. The empirical distribution function $\widehat{F}_{|R|}$ of the absolute values of the r_i 's is defined for any real number r by

$$\widehat{F}_{|R|}(r) = \sum_{i=1}^n \frac{\Delta(r - |r_i|)}{\widehat{N} \pi_i},$$

where $\Delta(a) = 0$ if $a \leq 0$ and $\Delta(a) = 1$ otherwise. If $y_{(i)}$ is the i -th order statistic of the sample $(y_1, \dots, y_n)$ and $\pi_{(i)}$ the corresponding inclusion probability, then the estimator $\widehat{M}_R$ is defined by

$$\begin{aligned} \widehat{M}_R &= \text{Inf}\{r : \widehat{F}_R(r) \geq 0.5\}, \\ &= |y_{(k)} - \widehat{B}_M|, \end{aligned}$$

where k is the smallest value such that

$$\sum_{i=1}^k \frac{1}{\pi_{(i)}} \geq \sum_{i=1}^n \frac{1}{2\pi_{(i)}}.$$

The starting values for $\widehat{B}_M^{(t)}$ and $\widehat{\sigma}_M^{(t)}$ can be for example

$$\widehat{B}_M^{(0)} = \widehat{M}_Y \text{ and } \widehat{\sigma}_M^{(0)} = C_\Phi \widehat{M}_R^{(0)},$$

where $\widehat{M}_Y$ is an estimator of the population median of the y_i 's.

Note that the iterative procedure described here is in fact the modified-weights procedure suggested by Huber (1981, p.146). It will be shown later theoretically as well as empirically that this estimator has very good robustness properties provided the errors in the superpopulation model are symmetrically distributed.

The following example deals with the M-estimator under the stratified random sampling design.

EXAMPLE

Suppose that the population has been stratified into H strata and that each strata h is of size N_h . Within each stratum, a simple random sample of size n_h will be selected. The M-estimator in this context is given by,

$$\widehat{Y}_{M,ST} = \widehat{B}_{M,ST},$$

where $\widehat{B}_{M,ST}$ is given by the following equation,

$$\sum_{h=1}^H \frac{N_h}{n_h} \sum_{i=1}^{n_h} \psi_H \left(\frac{y_{hi} - \widehat{B}_{M,ST}}{\widehat{\sigma}_{M,ST}} \right) = 0.$$

Some large sample results for the symmetric case are given in section 2.3.2. Essentially, we prove that the M-estimator under a general sampling design is asymptotically unbiased and consistent for estimating the finite population mean $\overline{Y}$. Afterwards, the asymptotic variance of that estimator is derived.

2.3.2. Large Sample Properties of the M-Estimator

The asymptotics will be developed within the well-known framework of embedded finite populations $\{\mathcal{U}_t\}_{t=1}^{+\infty}$. Therefore the population under study is assumed to be of size N_t and the samples are of fixed size n_t . For any value t , $\mathcal{U}_t \subset \mathcal{U}_{t+1}$. $N_t \leq N_{t+1}$ and $n_t \leq n_{t+1}$. In general, it is also assumed that

$$\lim_{t \rightarrow \infty} N_t = \lim_{t \rightarrow \infty} n_t = \infty.$$

However, the sampling fraction n_t/N_t is assumed to go to a positive constant f_0 as t goes to ∞ .

Definition 2.2. *An estimator $\widehat{\bar{Y}}_t$ of $\bar{Y}_t$ is said to be asymptotically unbiased if the following relation is satisfied:*

$$\lim_{t \rightarrow \infty} [E(\widehat{\bar{Y}}_t) - \bar{Y}_t] = 0.$$

In this section, the expectation is taken with respect to both the model and the sampling design. This means that $E(\widehat{\bar{Y}}) \equiv E_{\xi_0} E_d(\widehat{\bar{Y}})$ where E_d is the expectation with respect to the sampling design.

Definition 2.3. *An estimator $\widehat{\bar{Y}}_t$ of $\bar{Y}_t$ is consistent if,*

$$\lim_{t \rightarrow \infty} P(|\widehat{\bar{Y}}_t - \bar{Y}_t| > \epsilon) = 0, \quad \forall \epsilon > 0.$$

Again, probabilities are taken here with respect to both the model and the sampling design. The finite population values are assumed to be the realizations of N random variables, $Y_1, \dots, Y_N$ that are independently and identically distributed. Let $F_{\xi_0}(\cdot)$ be the common distribution function of the Y_i 's. We have that.

$$P(|\widehat{Y}_t - \bar{Y}_t| > \epsilon) = \int P_d(|\widehat{Y}_t - \bar{Y}_t| > \epsilon | Y_k = y_k, k = 1, \dots, N_t) \prod_{i=1}^{N_t} dF_{\xi_0}(y_i),$$

where P_d is the probability with respect to the sampling design.

One of our objectives is to study the conditions under which the M-estimator is a consistent estimator for estimating the population mean $\bar{Y}$. Its asymptotic unbiasedness and asymptotic variance will be studied as well. Let the functions $g_0(\cdot)$, $g_t(\cdot)$ and $\widehat{g}_t(\cdot)$ be defined for any real number b and any integer $t \geq 1$ by

$$g_0(b) = \int \psi_H[(y - b)/\sigma] dF_{\xi_0}(y).$$

$$g_t(b) = \frac{1}{N_t} \sum_{i=1}^{N_t} \psi_H[(y_i - b)/\sigma_M] \text{ and } \widehat{g}_t(b) = \sum_{i=1}^{n_t} \frac{\psi_H[(y_i - b)/\widehat{\sigma}_M]}{\widehat{N}_t \pi_i}.$$

where $\widehat{N}_t = \sum_{i=1}^{n_t} 1/\pi_i$.

The following conditions on the inclusion probabilities will be used very often;

$$\begin{cases} N_t \min_{1 \leq i \leq N_t} \pi_i \rightarrow \infty \text{ as } t \rightarrow \infty. \\ \max_{1 \leq i \neq j \leq N_t} \left| \frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right| \rightarrow 0 \text{ as } t \rightarrow \infty. \end{cases} \quad (2.2)$$

The first condition simply means that as t increases, each population unit will have an inclusion probability that remains reasonably large compared to $1/N_t$. Note that this condition also implies that the sample size n_t goes to infinity as t increases.

The second condition of 2.2 means that the sample membership indicators I_k , $k = 1, \dots, N_t$ are approximately pairwise independent for large samples. The k -th sample membership indicator I_k is defined by

$$I_k = \begin{cases} 1 & \text{if } k \text{ belongs to the sample,} \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 2.1. *Assume that the conditions 2.2 as well as the following conditions are satisfied,*

(i) $g_0(b) = 0$ has a unique solution β .

(ii) $\sup_{t \geq 1} |dg_t(\beta)/d\beta| \leq \sigma^{-1}$

Then $\widehat{Y}_{mt}$ is asymptotically unbiased and a consistent estimator for estimating $\bar{Y}_t$.

Assumption (i) merely means that the y_i 's in model ξ_0 are symmetrically distributed. Moreover, for the sake of simplicity, we will assume that $\widehat{\sigma}_m$ is a consistent estimator of σ but the theorem still holds even when this is not true. Assumption (ii) is usually satisfied for all standard robustness functions advocated in the literature.

Proof.

Note that for all integer $t \geq 1$, we have that

$$\begin{aligned} \left| E\left(\widehat{Y}_{mt} - \bar{Y}_t\right) \right| &\leq E\left|\widehat{Y}_{mt} - \bar{Y}_t\right|, \\ &= \int_0^\infty \left[1 - P\left(\left|\widehat{Y}_{mt} - \bar{Y}_t\right| \leq t\right) \right] dt, \\ &= \int_0^\infty P\left(\left|\widehat{Y}_{mt} - \bar{Y}_t\right| > t\right) dt. \end{aligned}$$

Thus, it is sufficient to prove that for all $\epsilon > 0$, we have that

$$\lim_{t \rightarrow \infty} P\left(\left|\widehat{Y}_{Mt} - \bar{Y}_t\right| > \epsilon\right) = 0.$$

Since $\widehat{\sigma}_M$ is a consistent estimator of σ , we can assume that by Slutsky's theorem that σ is known.

Let $\epsilon_0 > 0$. Normally, we have to prove that

$$\lim_{t \rightarrow \infty} P\left(\widehat{Y}_{Mt} > \bar{Y}_t + \epsilon_0\right) = 0$$

and

$$\lim_{t \rightarrow \infty} P\left(\widehat{Y}_{Mt} < \bar{Y}_t - \epsilon_0\right) = 0.$$

We will only prove that $\lim_{t \rightarrow \infty} P\left(\widehat{Y}_{Mt} > \bar{Y}_t + \epsilon_0\right) = 0$, the other part being similar.

Since $\widehat{g}_t(\cdot)$ is non-decreasing, $P\left(\widehat{Y}_{Mt} > \bar{Y}_t + \epsilon_0\right) \leq P\left(\widehat{g}_t(\bar{Y}_t + \epsilon_0) \geq 0\right)$. It follows from Chebychev's inequality that

$$\begin{aligned} P\left(\widehat{g}_t(\bar{Y}_t + \epsilon_0) \geq 0\right) &= P\left(\widehat{g}_t(\bar{Y}_t + \epsilon_0) - g_0(\beta + \epsilon_0) \geq -g_0(\beta + \epsilon_0)\right), \\ &\leq \frac{E_{\xi_0} E_d \left(\widehat{g}_t(\bar{Y}_t + \epsilon_0) - g_0(\beta + \epsilon_0)\right)^2}{g_0^2(\beta + \epsilon_0)}. \end{aligned}$$

It follows from (i) that $-g_0(\beta + \epsilon_0) > 0$. We only have to prove that the equation

$$E_{\xi_0} E_d \left(\widehat{g}_t(\bar{Y}_t + \epsilon_0) - g_0(\beta + \epsilon_0)\right)^2 = o(1) \text{ holds.}$$

$$\begin{aligned} E_{\xi_0} E_d \left(\widehat{g}_t(\bar{Y}_t + \epsilon_0) - g_0(\beta + \epsilon_0)\right)^2 &= E_{\xi_0} E_d \left[\left(\widehat{g}_t(\bar{Y}_t + \epsilon_0) - \widehat{g}_t(\beta + \epsilon_0)\right) + \right. \\ &\quad \left. + \left(\widehat{g}_t(\beta + \epsilon_0) - g_t(\beta + \epsilon_0)\right) + \left(g_t(\beta + \epsilon_0) - g_0(\beta + \epsilon_0)\right) \right]^2, \\ &= E_{\xi_0} E_d \left(\widehat{g}_t(\bar{Y}_t + \epsilon_0) - \widehat{g}_t(\beta + \epsilon_0)\right)^2 + E_{\xi_0} E_d \left(\widehat{g}_t(\beta + \epsilon_0) - g_t(\beta + \epsilon_0)\right)^2 + \end{aligned}$$

$$\begin{aligned}
& + E_{\xi_0} E_d (g_t(\beta + \epsilon_0) - g_0(\beta + \epsilon_0))^2 + 2E_{\xi_0} E_d (\widehat{g}_t(\bar{Y}_t + \epsilon_0) - \widehat{g}_t(\beta + \epsilon_0)) \times \\
& \times (\widehat{g}_t(\beta + \epsilon_0) - g_t(\beta + \epsilon_0)) + 2E_{\xi_0} E_d (\widehat{g}_t(\bar{Y}_t + \epsilon_0) - \widehat{g}_t(\beta + \epsilon_0)) \times \\
& \times (g_t(\beta + \epsilon_0) - g_0(\beta + \epsilon_0))
\end{aligned}$$

Note that one cross-product is missing. It is because, $E_p(\widehat{g}_t(\beta + \epsilon_0) - g_t(\beta + \epsilon_0)) = 0$.

Moreover, it follows from the mean-value theorem that for some value β_t^* such that

$|\beta_t^* - \beta| \leq |\beta - \bar{Y}_t|$ we have that

$$g_t(\bar{Y}_t + \epsilon_0) = g_t(\beta + \epsilon_0) + (\bar{Y}_t - \beta)g'_t(\beta_t^*).$$

This relation together with assumption (ii) implies that

$$|g_t(\bar{Y}_t + \epsilon_0) - g_t(\beta + \epsilon_0)| \leq \sigma^{-1} |\bar{Y}_t - \beta|, \quad \forall t \geq 1.$$

Therefore,

$$\begin{aligned}
& E_{\xi_0} E_d [\widehat{g}_t(\bar{Y}_t + \epsilon_0) - \widehat{g}_t(\beta + \epsilon_0)] [g_t(\beta + \epsilon_0) - g_0(\beta + \epsilon_0)] = \\
& = E_{\xi_0} [g_t(\bar{Y}_t + \epsilon_0) - g_t(\beta + \epsilon_0)] [g_t(\beta + \epsilon_0) - g_0(\beta + \epsilon_0)], \\
& \leq (1/\sigma) E_{\xi_0} |(\bar{Y}_t - \beta)(g_t(\beta + \epsilon_0) - g_0(\beta + \epsilon_0))|, \\
& \leq (1/\sigma) E_{\xi_0}^{1/2} (\bar{Y}_t - \beta)^2 \times E_{\xi_0}^{1/2} [g_t(\beta + \epsilon_0) - g_0(\beta + \epsilon_0)]^2, \\
& = N_t^{-1/2} \sqrt{V_{\xi_0}(g_t(\beta + \epsilon_0))}.
\end{aligned}$$

Similarly, it is quite straightforward to show that

$$E_{\xi_0} E_d [\widehat{g}_t(\bar{Y}_t + \epsilon_0) - \widehat{g}_t(\beta + \epsilon_0)] [\widehat{g}_t(\beta + \epsilon_0) - g_t(\beta + \epsilon_0)] = N_t^{-1/2} \sqrt{E_{\xi_0} V_d(\widehat{g}_t(\beta + \epsilon_0))}.$$

Again, the mean value theorem leads to the conclusion that

$$E_{\xi_0} E_d [\hat{g}_t(\bar{Y}_t + \epsilon_0) - \hat{g}_t(\beta + \epsilon_0)]^2 \leq \sigma^2 / N_t.$$

Since, $E_{\xi_0} E_d [\hat{g}_t(\beta + \epsilon_0) - g_t(\beta + \epsilon_0)]^2 = E_{\xi_0} V_d(\hat{g}_t(\beta + \epsilon_0))$ and $E_{\xi_0} E_d [g_t(\beta + \epsilon_0) - g_0(\beta + \epsilon_0)]^2 = V_{\xi_0}(g_t(\beta + \epsilon_0))$, we have that

$$\begin{aligned} E_{\xi_0} E_d (\hat{g}_t(\bar{Y}_t + \epsilon_0) - g_0(\beta + \epsilon_0))^2 &\leq 1/N_t + E_{\xi_0} V_d(\hat{g}_t(\beta + \epsilon_0)) + V_{\xi_0}(g_t(\beta + \epsilon_0)) + \\ &+ 2N_t^{-1/2} \sqrt{V_{\xi_0}(g_t(\beta + \epsilon_0))} + 2N_t^{-1/2} \sqrt{E_{\xi_0} V_d(\hat{g}_t(\beta + \epsilon_0))} \end{aligned}$$

It follows from the boundedness of ψ_H and from assumption 2.2 that,

$$\begin{aligned} E_{\xi_0} V_d(\hat{g}_t(\beta + \epsilon_0)) &\leq \left[\left(N_t \min_{1 \leq i \leq N_t} \pi_i \right)^{-1} + \max_{1 \leq i \neq j \leq N_t} \left| \frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right| \right] \times \\ &\times \int \psi_H^2 \{ [y - (\beta + \epsilon_0)] \} dF_{\xi_0}(y) = o(1), \\ V_{\xi_0}(g_t(\beta + \epsilon_0)) &\leq \frac{1}{N_t} \int \psi_H^2 \{ [y - (\beta + \epsilon_0)] \} dF_{\xi_0}(y) = O(1/N_t); \end{aligned}$$

which completes the proof.

■

Note that this result can only be obtained by using the randomness from the superpopulation model as well as that of the design. For, consistency is guaranteed by the symmetrical nature of the distribution of errors in the superpopulation model.

The next theorem provides us with a linear representation of an M-estimator, that is critical to deriving its asymptotic distributional properties. The proof of the validity of this linear representation is omitted as such a proof in a more general setting is given later for the CM-estimator.

Theorem 2.2. Assume conditions 2.2 as well as the following ones are satisfied:

- (i) $\hat{\sigma}_M$ is a consistent estimator of σ_M .
- (ii) We have that $\hat{B}_{Mt} - B_{Mt} = O_p(n_t^{-1/2})$.

Then the following linear representation of the M-estimator is valid,

$$\hat{Y}_{Mt} = B_{Mt} + \sum_{i=1}^{n_t} \frac{Q_{Mi}}{N_t \pi_i} + O_p(1/n_t),$$

where $\begin{cases} Q_{Mi} = -\psi_H(E_{Mi}^*)/g'_t(B_{Mt}), \\ E_{Mi}^* = (y_i - B_M)/\sigma_M \text{ and } g'_t(B_{Mt}) = -\sigma_{Mt}^{-1} \times \frac{1}{N_t} \sum_{i=1}^{N_t} \psi'_H(E_{Mi}^*). \end{cases}$

Since the conditions 2.2 imply that, uniformly in $b \in \mathbb{R}$, $\hat{g}_t(b)$ is a consistent estimator of $g_t(b)$, it is intuitively clear that assumption (i) is quite reasonable. As for assumption (ii), it is a well-known condition in the parametric framework. The regularity conditions needed for (ii) to be valid are studied in chapter 3 in a more general setting (see theorem 3.3).

Theorem 2.2 indicates that the M-estimator has the same asymptotic distribution as the pseudo-random variable, $\bar{Y}_M$ given by

$$\bar{Y}_M = B_M + \sum_{i=1}^n \frac{Q_{Mi}}{N \pi_i}.$$

Therefore, the asymptotic (or approximate) variance of the M-estimator is given by

$$AV(\hat{Y}_M) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) Q_{Mi} Q_{Mj}.$$

A consistent estimator of this approximate variance when N is known, is given by

$$\widehat{V}(\widehat{Y}_M) = \frac{1}{N^2} \sum_{i=1}^n \sum_{j=1}^n \frac{1}{\pi_{ij}} \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) \widehat{Q}_{Mi} \widehat{Q}_{Mj},$$

$$\text{where } \begin{cases} \widehat{Q}_{Mi} = -\psi_H(e_{Mi}^*) / \widehat{g}'_t(\widehat{B}_{Mt}), \\ e_{Mi}^* = (y_i - \widehat{B}_M) / \widehat{\sigma}_M \text{ and } \widehat{g}'_t(\widehat{B}_{Mt}) = -\widehat{\sigma}_{Mt}^{-1} \times \sum_{i=1}^{n_t} \frac{\psi'_H(e_{Mi}^*)}{N\pi_i}. \end{cases}$$

Note that, for the simple random sampling without replacement, the asymptotic variance becomes

$$AV(\widehat{Y}_M) = \frac{1-f}{n} \frac{1}{N-1} \sum_{i=1}^N Q_{Mi}^2 \text{ as } \overline{Q}_M = \frac{1}{N} \sum_{i=1}^N Q_{Mi} = 0.$$

and the corresponding estimator of variance is

$$v(\widehat{Y}_M) = \frac{1-f}{n} \frac{1}{n-1} \sum_{i=1}^n \widehat{Q}_{Mi}^2.$$

It is fairly straightforward to check that if the ψ_H -function is replaced with the identity function then one recovers the standard formulae for the sample mean.

It is also important to note that the form of the formula for the variance of the M-estimator indicates clearly why this estimator will be better than the GREG estimator when the observations are symmetrically distributed with heavy tails. Some simulation results presented at the end of this chapter will confirm this.

The M-estimator works well when the superpopulation exhibits some symmetry. However some natural populations are skewed in nature on one side. For example, if the variable of interest is the annual income of working people, then no symmetry can be expected under the common mean superpopulation model. In

such cases, M-estimators can be heavily biased and other alternatives are needed. In the next section we will propose two such alternatives: the CM-estimator and the LM-estimator. Our goal is to provide a general methodology for deriving robust alternatives to all estimators belonging to the class of GREG estimators. The CM-estimators which are also M-estimators in some sense, will provide robust alternatives to the GREG estimators that have the simple (or cosmetic) form. For GREG estimators with a general form, robust alternatives are provided by the LM-estimators which combine an M- and an L-estimator.

2.4. M-estimation for Complex Samples: the Asymmetric Case

In this section we will propose a design-consistent version of the M-estimator studied in the last section. It will be called consistent M-estimator (CM-estimator). Another alternative proposed in this section, is not an M-estimator as such. It is obtained by adjusting the M-estimator for the bias by a consistent L-estimator of the mean of predicted errors and is given the name LM-estimator. It will be seen in chapter 4 that the LM-estimator is the only robust alternative that will work when the GREG estimator does not take its simple form.

2.4.1. The Consistent M-Estimator

It is a well-known fact that for skewed populations the M-estimator can be heavily biased. Consequently, as the sample size increases, the GREG estimator will be getting better while the M-estimator will be getting worse. In general however, the

M-estimator has a mean square error that is lower than the variance of the GREG for small samples. But for large samples, the bias of the M-estimator will dominate its MSE that will become greater than the variance of the GREG estimator. The problem here is that it is never possible to know how small the sample must be for the M-estimator to be more efficient.

The problem mentioned above, is essentially a consistency problem. In order to solve it, we propose a modification of the M-estimator so as to achieve consistency. This modification is obtained by replacing the robustness function ψ_H with a compromise $\psi_{\bullet H}$ between the identity function and a bounded robustness function $\psi_H(\cdot, K_n)$ having the feature that the robustness constant K_n goes to ∞ with n .

More formally, the proposed robustness function is defined for all real number t by:

$$\psi_{\bullet H}(t) = w_n t + (1 - w_n) \psi_H(t, K_n), \quad (0 \leq w_n < 1)$$

where w_n is a weight that might depend on the sample size n . Note that the Winsorized Type II estimator of a population mean $\bar{X} = (1/N) \sum_{i=1}^N X_i$, discussed by Kokic and Bell (1994) under the simple random sampling design, uses such a linear combination and is given by,

$$\hat{\bar{Y}}_w = \frac{1}{n} \sum_{i=1}^n \psi_{\bullet H}(X_i) \text{ where } \begin{cases} \psi_{\bullet H}(X_i) = (n/N)X_i + (1 - n/N)\psi_H(X_i, K), \\ \psi_H(X_i) = \min(X_i, K). \end{cases}$$

Unlike Kokic and Bell, we will use the composite robust function to define not a Winsorized estimator but an M-estimator. Another difference with the Kokic-Bell approach is that we will let the winsorizing cutoff K_n (robustness constant K_n in

our framework) go to infinity so that for large samples the CM-estimator will get closer and closer to the population mean in probability.

No sound theory exists for determining the robustness constant K_n and the way it goes to infinity as the sample size increases. It is known however that a trade-off is needed between the bias and the variance. If K_n goes to infinity too fast, the bias will be small but the variance will be large. On the other hand, if K_n goes to infinity too slowly, the bias might dominate the variance. We will see later that some existing results on the optimal choice of the winsorization constant by Rivest and Hurtubise (1995) might guide the choice of K_n .

The proposed consistent M-estimator $\widehat{Y}_{CM} = \widehat{B}_{CM}$ (CM-estimator) is implicitly defined as solution to the following equation:

$$\sum_{i=1}^n \frac{1}{\pi_i} \psi_{\bullet_H} \left(\frac{y_i - \widehat{B}_{CM}}{\widehat{\sigma}_{CM}} \right) = 0. \quad (2.3)$$

$\widehat{\sigma}_{CM}$ is calculated as $\widehat{\sigma}_M$ for the Gwet-Rivest estimator after replacing $\widehat{B}_M$ by $\widehat{B}_{CM}$.

Note that if the $\psi_{\bullet_H}$ function is replaced with the identity function, then the CM-Estimator becomes equivalent to the ordinary GREG estimator. The idea behind this estimator is to have more stability by introducing some bias which must remain reasonable.

The CM-Estimator can be obtained by performing a few iterations of

$$\widehat{B}_{CM}^{(t+1)} = \frac{1}{\widehat{W}^{(t)}} \sum_{i=1}^n \frac{W_i^{(t)}}{\pi_i} y_i, \text{ where } \begin{cases} \widehat{W}^{(t)} = \sum_{i=1}^n \frac{W_i^{(t)}}{\pi_i}, W_i^{(t)} = w_n + (1 - w_n) \omega_i^{(t)}, \\ \omega_i^{(t)} = \psi_F(e_{CMi}^{*(t)}, K_n) / e_{CMi}^{*(t)}, \\ e_{CMi}^{*(t)} = (y_i - \widehat{B}_{CM}^{(t)}) / \widehat{\sigma}_{CM}^{(t)}. \end{cases}$$

The initial value $\widehat{B}_{\text{CM}}^{(0)}$ can be the median of the y_i 's. The simulation results reported in section 2.4 indicate a substantial gain in efficiency over the M-estimator as the sample size increases.

It is shown later that $(\widehat{B}_{\text{CM}}, \widehat{\sigma}_{\text{CM}})$ is design-consistent for estimating $(B_{\text{CM}}, \sigma_{\text{CM}})$. The latter being implicitly defined by

$$\sum_{i=1}^N \psi_{\bullet H} \left(\frac{y_i - B_{\text{CM}}}{\sigma_{\text{CM}}} \right) = 0.$$

Note that $\widehat{B}_{\text{CM}}$ and $\widehat{\sigma}_{\text{CM}}$ are sample size dependent.

The best choice for K_n , at least asymptotically depends on the limiting distribution of the population under study. Let $H(\cdot)$ be the population distribution function given by

$$H(r) = \frac{1}{N} \sum_{i=1}^N I(y_i \leq r),$$

where I is the indicator function.

The limiting distribution is never known in practice as only a fixed finite population is available at a given point in time. We suggest a conservative approach which assumes that the limiting distribution has heavy tails. We confine ourselves to one broad class of distributions with heavy tails where

$$1 - H(x) \approx C/x^\alpha \quad (\alpha > 0) \tag{2.4}$$

if x in the neighborhood of infinity and C a scale parameter. For such distributions. Rivest and Hurtubise (1995) proved that the optimal winsorization constant R to

be used with a Winsorized mean is defined by the following equation:

$$1 - H(R) \approx (\alpha - 1)/n. \quad (2.5)$$

It will be seen later in the next section that for large samples, and for populations that are positively skewed, the CM-estimator behaves like a Winsorized mean where the winsorization constant is given by

$$R = \sigma_{\text{CM}} K_n + B_{\text{CM}}.$$

Under 2.4, a solution to 2.5 is obtained with,

$$K_n = \sigma_{\text{CM}}^{-1} \left\{ [Cn/(\alpha - 1)]^{1/\alpha} - B_{\text{CM}} \right\}.$$

For samples of moderate size, K_n will be dominated by $[n/(\alpha - 1)]^{1/\alpha}$. Based on the above considerations, we suggest the following form for K_n :

$$K_n = K_{95} \left\{ 1 + [n/(\alpha - 1)]^{1/\alpha} \right\} \quad (\alpha > 1).$$

where K_{95} is the customary constant available in the literature, giving an efficiency of 95% of the M-estimator relative to the sample mean under normality.

Based on a previous census, one can try a few values $\alpha = 2, 3, 4$ and retain the one that minimizes the approximate mean squared error given in the next section. In our simulations reported in the end of the chapter, we have used $\alpha = 2$.

Note that no matter how the weight w_n is chosen, the resulting M-estimator will be consistent.

One way of choosing w_n is to use the optimality criterion according to which the optimal weight is the value of w_n (w_0 say), that minimizes the mean square error of the estimator. A drawback of such a strategy is that it requires data from a previous census. If a previous census is available, w_0 can be obtained in an iterative process as $w_0 = w_0(\alpha_0)$ is a function of α_0 , the optimal value of α . This strategy can be further explored.

Under the simple random sampling design, one can simply use $w_n = n/N$ as Kocio and Bell (1994). This choice which was used in our simulations, will lead to a CM-estimator that is also consistent in Cochran's sense, the Fisher's consistency.

2.4.2. The LM-Estimator

The CM-estimator like the M-estimator is only justifiable when the GREG estimator can be seen as the least squares estimator of some model parameter. Therefore, we will be able in general to justify the use of the CM-estimator when the GREG residuals sum to 0. But it is a well-known fact that sometimes this will not be the case, especially when the variance of the errors in the superpopulation model has a complex form. For these superpopulations, we need another approach to achieving robustness that will take into account the sum of prediction errors. This new approach, although unnecessary under model (ξ_0) (the sum of the GREG residuals always sum to 0 under ξ_0) is introduced here to be compared to the other robust strategies.

The population mean can be written as

$$\bar{Y} = B_M + \bar{E}_M \text{ where } \begin{cases} \bar{E}_M = \frac{1}{N} \sum_{i=1}^N E_{Mi}, \\ E_{Mi} = y_i - B_M, \end{cases}$$

and B_M is defined by (2.1). An estimator of the population mean can then be developed by estimating B_M and $\bar{E}_M$. The first parameter can easily be estimated by $\hat{B}_M$. We suggest to estimate the second term by a weighted combination of the π -estimator $\hat{\bar{E}}_M$ and the α_n -trimmed mean $\hat{\bar{E}}_{M(\alpha_n)}$ with the constraint that $\text{plim}_{n \rightarrow \infty} \alpha_n = 0$.

Let I be the indicator function. For each sample unit i , the sample M-residual e_{Mi} is given by $e_{Mi} = y_i - \hat{B}_M$. The empirical distribution function $\hat{F}$ of the sample M-residuals e_{Mi} 's is given for all real numbers r by

$$\hat{F}(r) = \sum_{i=1}^n \frac{I(e_{Mi} \leq r)}{\hat{N}\pi_i}.$$

The following notations are needed in the definition of the α_n -trimmed mean of the M-residuals. Let

$$e_{Mi}^* = \hat{J}_{ni} e_{Mi},$$

$$\text{where } \hat{J}_{ni} = [1 - (\lambda + 1)\alpha_n]^{-1} I\{\lambda\alpha_n < \hat{F}(e_{Mi}) \leq 1 - \alpha_n\}$$

$$\text{and } e_{Mi}^{**} = w_n e_{Mi} + (1 - w_n) e_{Mi}^*,$$

for some weight function w_n to be specified and λ is an integer that determines the number of units to be trimmed from the left tail. The mean of the M-residuals is then estimated by

$$\begin{aligned} \hat{\bar{E}}_{MR} &= w_n \hat{\bar{E}}_M + (1 - w_n) \hat{\bar{E}}_{M(\alpha_n)}, \\ &= \sum_{i=1}^n \frac{e_{Mi}^{**}}{\hat{N}\pi_i}. \end{aligned}$$

where $\widehat{E}_M$ and $\widehat{E}_{M(\alpha_n)}$ are respectively given by

$$\widehat{E}_M = \sum_{i=1}^n \frac{e_{Mi}}{\widehat{N}\pi_i}, \quad \widehat{E}_{M(\alpha_n)} = \sum_{i=1}^n \frac{e_{Mi}^*}{\widehat{N}\pi_i}$$

The general form of the LM-estimator is given by

$$\begin{aligned} \widehat{Y}_{LM} &= \widehat{B}_M + \widehat{E}_{MR}, \\ &= \widehat{B}_M + \sum_{i=1}^n \frac{e_{Mi}^{**}}{\widehat{N}\pi_i} \end{aligned} \quad (2.6)$$

Concerning α_n , we propose the following choice:

$$\alpha_n = 1/\widehat{n} \quad \text{where} \quad \widehat{n} = \widehat{N} \min_{1 \leq i \leq N} \pi_i.$$

For most known fixed-sample-size sampling designs, α_n will be fixed. However α_n might well be a random variable. The actual number of observations that will be trimmed depends on the sampling design. If some inclusion probabilities are small, the resulting weights will be substantial and more than one large M-residual is trimmed. It is safer in such a case to trim more large M-residuals so that the risk that a large residual and a large sampling weight will team up is minimized.

As an example, let us assume the stratified sampling design where the units are drawn within strata according to the SRS design. The population is divided into L strata of size N_h each, and a number of n_h units are drawn from stratum h . We have for all real numbers r that

$$\widehat{F}(r) = \sum_{h=1}^L W_h \widehat{F}_h(r) \quad \text{where} \quad \widehat{F}_h(r) = \frac{1}{n_h} \sum_{i=1}^{n_h} \Delta(e_{Mi} \leq r),$$

and $W_h = N_h/N$. Let $f_h = n_h/N_h$ be the sampling fraction in stratum h . We have that

$$\alpha_n = \frac{1}{\bar{n}} \text{ where } \bar{n} = N \min_{1 \leq h \leq L} f_h.$$

For simplicity, we assume that $L = 2$, that the two strata are identified as $\{h_1, h_2\}$ where $f_{h_1} = 2f_{h_2}$ and that stratum h_1 contains the two smallest robust residuals $e_{M(1)}$ and $e_{M(2)}$. In this case, $\alpha_n = 1/(Nf_{h_2})$. The two smallest residuals as well as the largest one will be trimmed. In fact,

$$\widehat{F}(e_{M(1)}) = \frac{1}{Nf_{h_1}} \text{ and } \widehat{F}(e_{M(2)}) = \frac{2}{Nf_{h_1}} \leq \alpha_n.$$

Similarly, if $f_{h_2} = 2f_{h_1}$ then the two largest residuals will be trimmed as well as the smallest one.

We also like to make the following remarks,

- In the robust statistics literature, the use of the trimmed mean in the univariate case is advocated by some authors as having some interesting robustness properties. The trimming parameter α is usually selected so as to achieve a given efficiency with respect to the sample mean under normality. For the 95% efficiency we have $\alpha = \alpha_{95} = 0.0893$. However, in the context of finite population sampling, it is reasonable to have more control over the number of sample units that will be trimmed. That justifies our choice of α_n .
- Under an arbitrary design, the smallest robust residual will always be trimmed. To see this, let $\pi_{(1)}$ be the inclusion probability associated with the smallest

residual $e_{M(1)}$. We have,

$$\widehat{\widehat{F}}(e_{M(1)}) = \frac{1}{\widehat{N}\pi_{(1)}} \leq \frac{1}{\widehat{N} \min_{1 \leq i \leq N} \pi_i} = \alpha_n.$$

The number of large units that will be trimmed depends upon the their inclusions probabilities. Suppose that there is $m \geq 0$ such that

$$\sum_{k=0}^m \frac{1}{\pi_{(n-k)}} < \frac{1}{\min_{1 \leq i \leq N} \pi_i}.$$

Then the $m+2$ largest residuals will be trimmed. If there is no $m \geq 0$ satisfying that condition, only the largest unit will be trimmed.

Concerning the choice of w_n , two options are possible. (i) if a previous census is available, then one can calculate a near optimal value w_n^* given by

$$w_n^* = \frac{\text{MSE}(\widehat{\widehat{E}}_{M(\alpha_n)})}{V(\widehat{\widehat{E}}_M) + \text{MSE}(\widehat{\widehat{E}}_{M(\alpha_n)})}$$

together with the optimal value of λ in an iterative process starting with $\lambda = 1$. (ii) if no previous census is available, then under the simple random sampling design. one can take $w_n = f$ (the sampling fraction) and $\lambda = 1$ or $\lambda = 2$ if the distribution is known to be skewed.

2.4.3. Asymptotic Properties of the CM-Estimator.

The asymptotics are developed within the customary framework of a sequence of nested finite populations $\{\mathcal{U}\}_{t \geq 1}$. For all t , $\mathcal{U}_{t+1} \supseteq \mathcal{U}_t$ and $N_{t+1} \geq N_t$, where N_t is the size of population $\mathcal{U}_t$.

Let $\sigma_{\text{CM}t}$ be the “population parameter” defined by:

$$\sigma_{\text{CM}t} = [1/\Phi^{-1}(3/4)] \text{med}_{1 \leq i \leq N_t} |E_{\text{CM}i}|,$$

where $E_{\text{CM}i} = y_i - B_{\text{CM}t}$, $B_{\text{CM}t}$ being another “population parameter” defined by:

$$\sum_{i=1}^{N_t} \psi_{\bullet H} \left(\frac{y_i - B_{\text{CM}t}}{\sigma_{\text{CM}t}} \right) = 0.$$

In fact $\sigma_{\text{CM}t}$ and $B_{\text{CM}t}$ are not population parameters in a strict sense as both quantities will in general be functions of the sample size n_t . In order to make the asymptotics more tractable, we will assume throughout this section, that the parameter $\sigma_{\text{CM}t}$ is known and that the statistic $\hat{\sigma}_{\text{CM}t}$ defined in (2.3) satisfies the condition $\hat{\sigma}_{\text{CM}t} - \sigma_{\text{CM}t} = o_p(1)$.

Before stating the theorems, let us introduce the following two notations

$$g_t(\theta) = \frac{1}{N_t} \sum_{i=1}^{N_t} \psi_{\bullet H} \left(\frac{y_i - \theta}{\sigma_{\text{CM}t}} \right), \text{ and } \hat{g}_t(\theta) = \sum_{i=1}^{n_t} \frac{1}{\hat{N}_t \pi_i} \psi_{\bullet H} \left(\frac{y_i - \theta}{\sigma_{\text{CM}t}} \right).$$

Theorem 2.3. Assume that 2.2 as well as the following conditions are satisfied:

- (i) $\lim_{t \rightarrow \infty} \bar{Y}_t = \mu_0$, $\lim_{t \rightarrow \infty} \sigma_{\text{CM}t} = \sigma_1 > 0$, and $\lim_{t \rightarrow \infty} w_n = w_0$ ($0 \leq w_0 \leq 1$).
 - (ii) For any fixed real value θ , we have $\lim_{t \rightarrow \infty} g_t(\theta) = g_0(\theta)$ where the function g_0 is such that $g_0(\mu_0) = 0$
 - (iii) $|\psi_{\bullet H}(r) - r| \leq r^2/K_n \forall r \in \mathbb{R}$.
 - (iv) $\sup_{t \geq 1} \frac{1}{N_t} \sum_{i=1}^{N_t} (y_i - \bar{Y}_t)^2 = O(1)$.
- Then $\hat{\bar{Y}}_{\text{CM}t}$ is a consistent estimator of $\bar{Y}_t$.

Lemma 2.1. *Under the condition (iv) of theorem 2.3 , we have that*

$$\lim_{t \rightarrow \infty} |\bar{Y}_t - B_{\text{CMT}}| = 0.$$

Proof: We have to show that

$$\forall \epsilon > 0, \exists t_\epsilon > 0 / t > t_\epsilon \rightarrow |\bar{Y}_t - B_{\text{CMT}}| < \epsilon.$$

As g_t is strictly increasing, this amounts to showing that

$$\forall \epsilon > 0, \exists t_\epsilon > 0 / t > t_\epsilon \Rightarrow g_t(\bar{Y}_t + \epsilon) < g_t(B_{\text{CMT}}) < g_t(\bar{Y}_t - \epsilon).$$

$$\Rightarrow g_t(\bar{Y}_t + \epsilon) < 0. \text{ and } g_t(\bar{Y}_t - \epsilon) > 0.$$

as $g_t(B_{\text{CMT}}) = 0$. We will only show that $g_t(\bar{Y}_t + \epsilon) < 0$ whenever $t > t_\epsilon$ for some $t_\epsilon > 0$. The proof of the inequality $g_t(\bar{Y}_t - \epsilon) > 0$ is similar.

Let $e_{\epsilon i} = (y_i - \bar{Y}_t) - \epsilon$ and $e_{\epsilon i}^* = e_{\epsilon i} / \sigma_{\text{CMT}}$. Using the definition of the function $\psi_{\bullet H}$ in section 2.4.1 , we have that

$$g_t(\bar{Y}_t + \epsilon) = -w_n \sigma_{\text{CMT}}^{-1} \epsilon + (1 - w_n) N_t^{-1} \sum_{i=1}^{N_t} \psi_H(e_{\epsilon i}^*),$$

It follows from this that,

$$\begin{aligned} g_t(\bar{Y}_t + \epsilon) + \epsilon \sigma_{\text{CMT}}^{-1} &= (1 - w_n) \left(\frac{1}{N_t} \sum_{i=1}^{N_t} [\psi_H(e_{\epsilon i}^*) - e_{\epsilon i}^*] \right), \\ &\leq \frac{1}{K_n} \left(\frac{1}{N_t} \sum_{i=1}^{N_t} e_{\epsilon i}^{*2} \right) = \frac{1}{K_n} \left(\sigma_{\text{CMT}}^{-2} \frac{1}{N_t} \sum_{i=1}^{N_t} (y_i - \bar{Y}_t)^2 + (\sigma_{\text{CMT}}^{-1} \epsilon)^2 \right). \end{aligned}$$

Since K_n goes to infinity, it follows from assumptions (i) and (iv) of theorem 2.3 that there exists t_0 such that if $t > t_0$ then the right-hand side of the above inequality is smaller than $\sigma_1^{-1} \epsilon / 2$. This implies that

$$t > t_0 \implies g_t(\bar{Y}_t + \epsilon) \leq -\sigma_1^{-1} \epsilon / 2 < 0.$$

■

Proof of theorem 2.3 :

We have to show that for all $\epsilon > 0$, $\lim_{n \rightarrow \infty} P(|\bar{Y}_t - \hat{B}_{Cmt}| > \epsilon) = 0$.

Let $\epsilon_0 > 0$. We have that

$$P(|\hat{B}_{Cmt} - \bar{Y}_t| > \epsilon_0) = P(\hat{B}_{Cmt} > \bar{Y}_t + \epsilon_0) + P(\hat{B}_{Cmt} < \bar{Y}_t - \epsilon_0).$$

Therefore, it suffices to show that both members of the right-hand side of the above equation go to 0 as t goes to ∞ . We will only show that $\lim_{t \rightarrow \infty} P(\hat{B}_{Cmt} > \bar{Y}_t + \epsilon_0) = 0$ as proving that the other limit is 0 is similar.

Note that $\hat{g}_t$ is a decreasing function of $\hat{B}_{Cmt}$. This implies that

$$\begin{aligned} P(\hat{B}_{Cmt} > \bar{Y}_t + \epsilon_0) &= P[\hat{g}_t(\hat{B}_{Cmt}) < \hat{g}_t(\bar{Y}_t + \epsilon_0)], \\ &= P[\hat{g}_t(\bar{Y}_t + \epsilon_0) > 0] \quad \text{as } \hat{g}_t(\hat{B}_{Cmt}) = 0. \end{aligned}$$

But,

$$P[\hat{g}_t(\bar{Y}_t + \epsilon_0) > 0] = P[\hat{g}_t(\bar{Y}_t + \epsilon_0) - g_t(\bar{Y}_t + \epsilon_0) > -g_t(\bar{Y}_t + \epsilon_0)].$$

It follows from lemma 2.1 , that $\exists t_{\epsilon_0}$ such that

$$t > t_{\epsilon_0} \Rightarrow \bar{Y}_t - B_{Cmt} > -\epsilon_0/2.$$

Therefore, if $t > t_{\epsilon_0}$ then

$$\begin{aligned} g_t(\bar{Y}_t + \epsilon_0) &= g_t[(\epsilon_0 + B_{Cmt}) + (\bar{Y}_t - B_{Cmt})], \\ &< g_t(\epsilon_0/2 + B_{Cmt}) < 0 \quad (\text{as } g_t(B_{Cmt}) = 0) \end{aligned}$$

It follows from the Chebychev's inequality that

$$P\left[\hat{g}_t(\bar{Y}_t + \epsilon_0) - g_t(\bar{Y}_t + \epsilon_0) > -g_t(\bar{Y}_t + \epsilon_0)\right] \leq \frac{V\left[\hat{g}_t(\bar{Y}_t + \epsilon_0)\right]}{g_t^2(\bar{Y}_t + \epsilon_0)}.$$

But,

$$\hat{g}_t(\bar{Y}_t + \epsilon_0) = \sum_{i=1}^{n_t} (N_t \pi_i)^{-1} \psi_{\bullet H} \left(\frac{(y_i - \bar{Y}_t) - \epsilon_0}{\sigma_{\text{CMT}}} \right).$$

It is not difficult to see that

$$V\left[\hat{g}_t(\bar{Y}_t + \epsilon_0)\right] \leq \left[\left(N_t \min_{1 \leq i \leq N_t} \pi_i \right)^{-1} + \max_{1 \leq i \neq j \leq N_t} \left| \frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right| \right] \frac{1}{N_t} \sum_{i=1}^{N_t} \psi_{\bullet H}^2 \left(\frac{(y_i - \bar{Y}_t) - \epsilon_0}{\sigma_{\text{CMT}}} \right).$$

It follows from the Jensen's inequality that

$$\begin{aligned} \psi_{\bullet H}^2 \left(\frac{(y_i - \bar{Y}_t) - \epsilon_0}{\sigma_{\text{CMT}}} \right) &= \\ &= \left\{ w_n \left(\frac{(y_i - \bar{Y}_t) - \epsilon_0}{\sigma_{\text{CMT}}} \right) + (1 - w_n) \psi_H \left(\frac{(y_i - \bar{Y}_t) - \epsilon_0}{\sigma_{\text{CMT}}}, K_n \right) \right\}^2 \\ &\leq w_n \left(\frac{(y_i - \bar{Y}_t) - \epsilon_0}{\sigma_{\text{CMT}}} \right)^2 + (1 - w_n) \psi_H^2 \left(\frac{(y_i - \bar{Y}_t) - \epsilon_0}{\sigma_{\text{CMT}}}, K_n \right), \\ &\leq \sigma_{\text{CMT}}^{-2} [(y_i - \bar{Y}_t) - \epsilon_0]^2. \end{aligned}$$

Therefore,

$$N_t^{-1} \sum_{i=1}^{N_t} \psi_{\bullet H}^2 \left(\frac{(y_i - \bar{Y}_t) - \epsilon_0}{\sigma_{\text{CMT}}} \right) \leq \sigma_{\text{CMT}}^{-2} \left(N_t^{-1} \sum_{i=1}^{N_t} (y_i - \bar{Y}_t)^2 + \epsilon_0^2 \right).$$

The term on the right is bounded by $\sigma_{\text{CMT}}^{-2} \left[\left(\max_{1 \leq i \leq N_t} |y_i - \bar{Y}_t| \right)^2 + \epsilon_0^2 \right]$ which goes to $\sigma_{\text{CMT}}^{-2} (C_0 + \epsilon_0^2)$ for some positive constant C_0 (this follows from assumptions (i) and (v)). It follows from assumptions (iii) and (iv) that the variance above goes to 0 as t goes to ∞ . Since $g_t^2(\bar{Y}_t + \epsilon_0)$ goes to $g_0^2(\mu_0 + \epsilon_0) > 0$ (assumption ii) as t goes to ∞ the proof follows. ■

The next result is essential in the derivation of the approximate MSE of the CM-estimator. It aims to develop a linear representation of the estimator $\widehat{Y}_{\text{CM}}$. Such a representation was developed in the context of M-estimators under a general regression model, by C.R. Rao & L.C. Zhao (1992). They obtained their results within the framework of parametric statistical inference using almost sure convergence.

We like to obtain a similar expansion in the context of design-based inference using convergence in probability. This type of convergence more popular in survey sampling

Theorem 2.4.

Assume that the conditions 2.2 as well as the following conditions are satisfied.

(i) *The function $\widehat{g}_t$ possesses a second-order derivative at each point*

$$\theta \in (\widehat{B}_{\text{CM}t}, B_{\text{CM}t})$$

(ii) $\widehat{B}_{\text{CM}t} - B_{\text{CM}t} = O_p(n_t^{-1/2})$

(iii) $\lim_{t \rightarrow \infty} w_{n_t} = w_0$ ($0 \leq w_0 < 1$) and $\lim_{t \rightarrow \infty} \sigma_{\text{CM}t} = \sigma_0 > 0$.

(iv) ψ_H is twice differentiable in the neighborhood of 0 and $|\psi_H''(t, K)| \leq 2/K$

for all t in that neighborhood.

(v) $1 - H(x) \sim C/x^\beta$ ($\beta \leq \alpha$)

Then the following representation holds

$$\widehat{Y}_{\text{CM}t} = B_{\text{CM}t} + \sum_{i=1}^{n_t} \frac{Q_{\text{CM}i}}{N_t \pi_i} + O_p(n_t^{-1}),$$

where,

$$Q_{\text{CM}i} = -[g_t'(B_{\text{CM}t})]^{-1} \psi_{\bullet H} \left(\frac{y_i - B_{\text{CM}t}}{\sigma_{\text{CM}t}} \right)$$

and

$$g'_t(B_{\text{CMT}}) = -\sigma_{\text{CMT}}^{-1} \left\{ w_n + (1 - w_n) N_t^{-1} \sum_{i=1}^{N_t} \psi'_H [\sigma_{\text{CMT}}^{-1} (y_i - B_{\text{CMT}}), K_n] \right\}.$$

The regularity conditions needed for the assumption (ii) to be valid are discussed in theorem 3.3 of chapter 3 in a more general setting. Assumption (v) will guarantee that the leading term in the above linear representation is not identically equal to the sample mean.

Let $\widehat{Q}_{\text{CM}}$ be given by

$$\widehat{Q}_{\text{CM}} = \sum_{i=1}^{n_t} \frac{Q_{\text{CM}i}}{N_t \pi_i}.$$

The CM-estimator will have an asymptotic variance different from that of the GREG estimator $\widehat{Y}_{\text{G}}$ if

$$\sqrt{n_t} [\widehat{Q}_{\text{CM}} - (\widehat{Y}_{\text{G}} - B_{\text{CM}})] = O_p(1).$$

The limiting distribution of $\sqrt{n_t} \widehat{Q}_{\text{CM}}$ is an outstanding research problem.

We will need the following lemma.

Lemma 2.2. (*Bhat (1981), lemma 5.2, p. 85*). *Let $\{X_n\}_{n \geq 1}$ be a sequence of random variables. Then the following relation holds:*

$$\sup_{n \geq 1} E|X_n| < \infty \Rightarrow X_n = O_p(1).$$

Proof of theorem 2.4 :

Since the function $\psi_{\bullet H}$ is continuous, differentiable everywhere, condition (i) makes

use of the mean value theorem possible. Therefore the following expansion holds:

$$\widehat{g}_t(\widehat{B}_{\text{CMT}}) = \widehat{g}_t(B_{\text{CMT}}) + (\widehat{B}_{\text{CMT}} - B_{\text{CMT}})\widehat{g}'_t(B_{\text{CMT}}) + (1/2)(\widehat{B}_{\text{CMT}} - B_{\text{CMT}})^2\widehat{g}''_t(B_{\text{CMT}}^*).$$

where B_{CMT}^* satisfies $|B_{\text{CMT}}^* - \widehat{B}_{\text{CMT}}| < |\widehat{B}_{\text{CMT}} - B_{\text{CMT}}|$.

It follows from the conditions “empty” and (iii) that

$$\begin{aligned}\widehat{g}'_t(B_{\text{CMT}}) &= -\sigma_{\text{CMT}}^{-1} \left\{ w_n + (1 - w_n)N_t^{-1} \sum_{i=1}^{N_t} \psi'_H \left(\frac{y_i - B_{\text{CMT}}}{\sigma_{\text{CMT}}}, K_n \right) \right\} + O_p(n_t^{-1/2}). \\ &= g'_t(B_{\text{CMT}}) + O_p(n_t^{-1/2}).\end{aligned}$$

It follows from the above expansion of $\widehat{g}_t(\widehat{B}_{\text{CMT}})$ that

$$0 = \widehat{g}_t(B_{\text{CMT}}) + (\widehat{B}_{\text{CMT}} - B_{\text{CMT}})g'_t(B_{\text{CMT}}) + (1/2)(\widehat{B}_{\text{CMT}} - B_{\text{CMT}})^2\widehat{g}''_t(B_{\text{CMT}}^*) + O_p(n_t^{-1}).$$

as $\widehat{g}_t(\widehat{B}_{\text{CMT}}) = 0$. Now, we have to show that

$$(\widehat{B}_{\text{CMT}} - B_{\text{CMT}})^2\widehat{g}''_t(B_{\text{CMT}}^*) = O_p(n_t^{-1})$$

Assumption (ii) implies that $(\widehat{B}_{\text{CMT}} - B_{\text{CMT}})^2 = O_p(n_t^{-1})$. Therefore, it remains to prove that $\widehat{g}''_t(B_{\text{CMT}}^*) = O_p(1)$. It follows from lemma 2.2 that it suffices to show that

$$\sup_{t \geq 1} E|\widehat{g}''_t(B_{\text{CMT}}^*)| < \infty.$$

Using the definition of $\psi_{\bullet H}$, it is possible to obtain the following expression for the second-order derivative of $\widehat{g}_t$,

$$\widehat{g}''_t(\theta) = (1 - w_n)\sigma_{\text{CMT}}^{-2} \sum_{i=1}^{n_t} (N_t\pi_i)^{-1} \psi''_H[\sigma_{\text{CMT}}^{-1}(y_i - \theta), K_n], \text{ for any } \theta.$$

Appealing to the Slutsky's theorem, we get

$$\begin{aligned} E|\hat{g}_t''(B_{\text{CM}t}^*)| &\leq (1 - w_n)\sigma_{\text{CM}t}^{-2}E\left\{\sum_{i=1}^{n_t}(N_t\pi_i)^{-1}\left|\psi_H''[\sigma_{\text{CM}t}^{-1}(y_i - B_{\text{CM}t}^*), K_n]\right|\right\}, \\ &= (1 - w_n)\sigma_{\text{CM}t}^{-2}N_t^{-1}\sum_{i=1}^{N_t}\left|\psi_H''[\sigma_{\text{CM}t}^{-1}(y_i - B_{\text{CM}t}), K_n]\right| + o(1). \end{aligned}$$

It follows from condition (iv) that $|\psi_H''(t, K_n)| \leq 2/K_n$. Hence.

$$E|\hat{g}_t''(B_{\text{CM}t}^*)| \leq 2(1 - w_n)\sigma_{\text{CM}t}^{-2}K_n^{-1} + o(1)$$

and $E|\hat{g}_t''(B_{\text{CM}t}^*)| = o(1)$.

■

We now turn to the problem of estimating the variance of the CM-estimator. If the conditions of theorem 2.4 are satisfied, then the CM-estimator is asymptotically equivalent to the following statistics:

$$\bar{Y}_{\text{CM}t}^* = B_{\text{CM}t} + \sum_{i=1}^n \frac{Q_{\text{CM},i}}{N\pi_i}.$$

Therefore, the asymptotic variance of $\hat{\bar{Y}}_{\text{CM}}$ is equal to the variance of $\bar{Y}_{\text{CM}t}^*$. Thus it is not difficult to see that,

$$AV(\hat{\bar{Y}}_{\text{CM}}) = N^{-2} \sum_{i=1}^N \sum_{j=1}^N \left(\frac{\pi_{ij}}{\pi_i\pi_j} - 1 \right) Q_{\text{CM},i} Q_{\text{CM},j}. \quad (2.7)$$

In practice, this variance has to be estimated from the sample at hand. Estimating the asymptotic variance essentially amounts to estimating the $Q_{\text{CM},i}$'s. The $Q_{\text{CM},i}$

depends on the parameters $\sigma_{\text{CM}t}$ and $B_{\text{CM}t}$ that must be estimated. The corresponding estimators, $\hat{\sigma}_{\text{CM}t}$ and $\hat{B}_{\text{CM}t}$ are defined in section 2.4.1. We then propose the following variance estimator:

$$\hat{V}(\hat{Y}_{\text{CM}}) = N^{-2} \sum_{i=1}^n \sum_{j=1}^n \frac{1}{\pi_{ij}} \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) \hat{Q}_{\text{CM},i} \hat{Q}_{\text{CM},j},$$

where $\hat{Q}_{\text{CM},i} = [\hat{g}_t^{*'}(\hat{B}_{\text{CM}t})]^{-1} \psi_{\bullet H} \left(\frac{y_i - \hat{B}_{\text{CM}t}}{\hat{\sigma}_{\text{CM}t}} \right)$, and

$$\hat{g}_t^{*'}(\hat{B}_{\text{CM}t}) = \hat{\sigma}_{\text{CM}t}^{-1} \left\{ w_n + (1 - w_n) \sum_{i=1}^{n_t} (\hat{N}_t \pi_i)^{-1} \psi_H' \left(\frac{y_i - \hat{B}_{\text{CM}t}}{\hat{\sigma}_{\text{CM}t}}, K_n \right) \right\}.$$

This variance estimator is consistent for estimating the asymptotic variance of the CM-estimator.

It is important to point out that, although the CM-estimator is consistent for estimating $\bar{Y}$, its bias might not be negligible for small samples except maybe in some particular cases. It is difficult to study the bias problem for very samples as the linear representation in theorem 2.4 might not hold. However, if the sample size is large enough for the linear representation to hold, then the bias will be approximately given by $B_a(\hat{Y}_{\text{CM}t}) = B_{\text{CM}t} - \bar{Y}_t$. For many distributions, it is at most of order $O(n_t^{-1/\alpha})$. In such a case, instead of adjusting the CM-estimator for the bias, which might lead to a highly unstable estimator, we propose to assess the magnitude of the squared bias. This assessment could be used to estimate the mean square error of the CM-estimator.

A natural estimator of $B_a(\hat{Y}_{\text{CM}t})$ is given by

$$\hat{B}_a(\hat{Y}_{\text{CM}t}) = \hat{B}_{\text{CM}t} - \hat{Y}_G.$$

However,

$$E[\widehat{B}_a^2(\widehat{Y}_{\text{CMt}})] = \left[E\widehat{B}_a(\widehat{Y}_{\text{CMt}}) \right]^2 + V\left[\widehat{B}_a(\widehat{Y}_{\text{CMt}}) \right],$$

which suggests the following estimator for the squared bias:

$$\widehat{B}_{2a}(\widehat{Y}_{\text{CM}}) = \widehat{B}_a^2(\widehat{Y}_{\text{CM}}) - \widehat{V}\left[\widehat{B}_a(\widehat{Y}_{\text{CM}}) \right].$$

The asymptotic variance of $\widehat{B}_a(\widehat{Y}_{\text{CM}})$ is given by

$$AV\left[\widehat{B}_a(\widehat{Y}_{\text{CM}}) \right] = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) U_i U_j,$$

where

$$U_i = Q_{\text{CM},i} - E_{\text{CM},i}.$$

Moreover,

$$E_{\text{CM},i} = y_i - B_{\text{CM}}$$

and

$$Q_{\text{CM},i} = \frac{\sigma_{\text{CM}} \psi_{\bullet H}(E_{\text{CM},i}/\sigma_{\text{CM}})}{(1/N) \sum_{i=1}^N \psi'_{\bullet H}(E_{\text{CM},i}/\sigma_{\text{CM}})}$$

This suggests the following estimator,

$$\widehat{V}\left[\widehat{B}_a(\widehat{Y}_{\text{CM}}) \right] = \sum_{i=1}^n \sum_{j=1}^n \frac{1}{\widehat{N} \pi_{ij}} \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) u_i u_j,$$

where,

$$u_i = \widehat{Q}_{\text{CM},i} - e_{\text{CM},i}, \quad e_{\text{CM},i} = y_i - \widehat{B}_{\text{CM}}$$

and

$$\hat{Q}_{\text{CM},i} = \frac{\hat{\sigma}_{\text{CM}} \psi_{\bullet H}(e_{\text{CM},i}/\hat{\sigma}_{\text{CM}})}{(1/n_t) \sum_{i=1}^{n_t} \psi'_{\bullet H}(e_{\text{CM},i}/\hat{\sigma}_{\text{CM}})}$$

Another problem remains to be solved. It is possible for some samples that $\hat{B}_a^2(\hat{Y}_{\text{CM}})$ be greater than $\hat{V}[\hat{B}_a(\hat{Y}_{\text{CM}})]$. For that reason the estimator of the squared bias should be modified as follows:

$$\hat{B}_{2a}(\hat{Y}_{\text{CM}}) = \max\left\{0, \hat{B}_a^2(\hat{Y}_{\text{CM}}) - \hat{V}[\hat{B}_a(\hat{Y}_{\text{CM}})]\right\}.$$

The MSE estimator of the CM-estimator is then given by,

$$\widehat{MSE}(\hat{Y}_{\text{CM}}) = \hat{V}(\hat{Y}_{\text{CM}}) + \hat{B}_{2a}(\hat{Y}_{\text{CM}}).$$

2.4.4. Asymptotic Properties of the LM-estimator.

We start this section by introducing the following notations :

- F_t : the population distribution of the theoretical residuals E_{Mi} where $E_{\text{Mi}} = y_i - B_{\text{M}}$.
- $\hat{F}_t$: the empirical distribution of the theoretical residuals E_{Mi} 's.
- $\hat{\hat{F}}_t$: the empirical distribution of the empirical residuals e_{Mi} 's where $e_{\text{Mi}} = y_i - \hat{B}_{\text{M}}$.

These distribution functions are defined as follows: $\forall r \in \mathbb{R}$,

$$F_t(r) = \sum_{i=1}^{N_t} \frac{\Delta(r - E_{\text{Mi}})}{N_t}, \quad \hat{F}_t(r) = \sum_{i=1}^{n_t} \frac{\Delta(r - E_{\text{Mi}})}{\hat{N}_t \pi_i}, \quad \hat{\hat{F}}_t(r) = \sum_{i=1}^{n_t} \frac{\Delta(r - e_{\text{Mi}})}{\hat{N}_t \pi_i}.$$

It follows from (2.6) that the LM-estimator can be written as:

$$\widehat{Y}_{LM} = \widehat{B}_M + \Theta(\widehat{F}),$$

where $\Theta(\cdot)$ is a functional given for any distribution function G of the residuals. by

$$\Theta(G) = \int r \left\{ w_n + (1 - w_n) J_n [G(r)] \right\} dG(r).$$

In general the J_n function we will be using takes the following form:

$$\forall t \in \mathbb{R}, \quad J_n(t) = [1 - (\lambda + 1)\alpha_{n_t}]^{-1} I(\lambda\alpha_{n_t} < t \leq 1 - \alpha_{n_t}), \quad (2.8)$$

where $0 \leq \alpha_{n_t} < 0.50$. Note that this function is bounded but discontinuous. If

smoothness is required, the following form can be used:

$$\begin{aligned} J_n(t) = & \frac{t - \lambda\alpha_{n_t}}{\lambda\alpha_{n_t} [1 - 3\alpha_{n_t}(\lambda + 1)/2]} I(\lambda\alpha_{n_t} \leq t < 2\lambda\alpha_{n_t}) + \\ & + \frac{1}{[1 - 3\alpha_{n_t}(\lambda + 1)/2]} I(2\lambda\alpha_{n_t} \leq t < 1 - 2\alpha_{n_t}) + \\ & + \frac{1 - (t + \alpha_{n_t})}{\alpha_{n_t} [1 - 3\alpha_{n_t}(\lambda + 1)/2]} I(1 - 2\alpha_{n_t} \leq t < 1 - \alpha_{n_t}), \end{aligned} \quad (2.9)$$

for $0 \leq \alpha_{n_t} < 1/4$.

In order to make the proofs less complicated we will assume throughout this section that

$$\alpha_{n_t} \simeq 1/n_0 \quad \text{where } n_0 = N \min_{1 \leq i \leq N} \pi_i,$$

so that α_{n_t} is no longer a random variable.

Definition 2.4. (*Tightness*) A sequence of distribution functions $\{F_t\}_{t \geq 1}$ is said to be tight if for each $\epsilon > 0$ there exists x_ϵ and y_ϵ such that for all $t \geq 1$ $F_t(x_\epsilon) < \epsilon$ and $F_t(y_\epsilon) > 1 - \epsilon$.

Lemma 2.3. (*R. Durrett, 1991, page 74*) Let $\{F_t\}_{t \geq 1}$ be a sequence of distribution functions. If there is a $\varphi \geq 0$ so that $\varphi(x) \rightarrow 0$ as $|x| \rightarrow \infty$ and

$$C = \sup_{t \geq 1} \int \varphi(x) dF_t(x) < \infty,$$

then the sequence $\{F_t\}_{t \geq 1}$ is tight.

The proof of this lemma is trivial and given in Durrett (1991).

Lemma 2.4. Let $\{U_t\}_{t \geq 1}$ be a sequence of uniform discrete random variables where U_t can take any value in the set of numbers $\{1/N_t, 2/N_t, \dots, 1\}$ with the same probability $1/N_t$. Let F_{U_t} be the distribution function of U_t . Then the sequence $\{F_{U_t}\}_{t \geq 1}$ is tight.

Proof.

Let $\varphi(\cdot)$ be defined for all x by $\varphi(x) = |x|$. Then we have that,

$$\begin{aligned} C &= \sup_{t \geq 1} \int \varphi(x) dF_{U_t}(x), \\ &= \sup_{t \geq 1} \sum_{i=1}^{N_t} \frac{i}{N_t} \frac{1}{N_t} = \sup_{t \geq 1} \frac{1}{N_t^2} \sum_{i=1}^{N_t} i, \\ &= \sup_{t \geq 1} \frac{N_t(N_t + 1)}{2N_t^2} < 3/2. \end{aligned}$$

The result follows from lemma 2.3 .

■

The following two assumptions will be used regularly throughout this section.

A1• $N_t \min_{1 \leq i \leq N_t} \pi_i = O(n_t)$.

$$\text{A2} \bullet \max_{1 \leq i \neq j \leq N_t} \left| \frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right| = O(1/n_t).$$

The first result we will prove is the consistency of the LM-estimator.

Lemma 2.5. (*Cesàro averages, Billingsley, 1986, page 572*) Let $(x_n)_{n \geq 1}$ be sequence of real numbers.

$$\text{If } \lim_{n \rightarrow \infty} x_n = x \text{ then } \lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n x_k = x.$$

We introduce the following notations:

$$\hat{J}_{ni} = J_n[\hat{F}(e_{Mi})], \quad \hat{J}_{ni} = J_n[\hat{F}(E_{Mi})], \quad J_{ni} = J_n[F(E_{Mi})].$$

Lemma 2.6. Under assumptions A1 and A2, the following relation holds:

$$\sum_{i=1}^n \frac{(\hat{J}_{ni} - 1)^2}{\hat{N}_t \pi_i} = o_p(1).$$

Proof.

Note that for all $\epsilon > 0$,

$$\begin{aligned} P\left(\left|\sum_{i=1}^{n_t} \frac{(\hat{J}_{ni} - 1)^2}{\hat{N}_t \pi_i}\right| > \epsilon\right) &\leq \\ &\leq P\left(\left|\sum_{i=1}^{n_t} \frac{(\hat{J}_{ni} - 1)^2}{N_t \pi_i}\right| + \left|\sum_{i=1}^{n_t} \frac{(\hat{J}_{ni} - 1)^2}{\pi_i} \left(\frac{1}{N_t} - \frac{1}{\hat{N}_t}\right)\right| > \epsilon\right), \\ &\leq P\left(\left|\sum_{i=1}^{n_t} \frac{(\hat{J}_{ni} - 1)^2}{N_t \pi_i}\right| > \epsilon/2\right) + P\left(\left|\sum_{i=1}^{n_t} \frac{(\hat{J}_{ni} - 1)^2}{\pi_i} \left(\frac{1}{N_t} - \frac{1}{\hat{N}_t}\right)\right| > \epsilon/2\right), \\ &\leq P\left(\left|\sum_{i=1}^{n_t} \frac{(\hat{J}_{ni} - 1)^2}{N_t \pi_i}\right| > \epsilon/2\right) + P\left(\frac{4}{[1 - (\lambda + 1)\alpha_{n_t}]^2 N_t} |\hat{N}_t - N_t| > \epsilon/2\right) \end{aligned}$$

But $\lim_{t \rightarrow \infty} \left\{ [1 - (\lambda + 1)\alpha_{n_t}]^2 N_t \right\}^{-1} = 0$. Therefore, $\exists t_0 > 0$ such that if $t > t_0$ then

$$P\left(\frac{1}{[1 - (\lambda + 1)\alpha_{n_t}]^2 N_t} |\hat{N}_t - N_t| > \epsilon/2\right) \leq P(|\hat{N}_t - N_t| > \epsilon/2).$$

It follows from Chebychev's inequality and from assumptions A1 and A2 that this probability goes to 0 as t goes to ∞ .

It remains to show that

$$\lim_{t \rightarrow \infty} P\left(\left|\sum_{i=1}^{n_t} \frac{(\hat{J}_{ni} - 1)^2}{N_t \pi_i}\right| > \epsilon/2\right) = 0.$$

We have that,

$$\left|\sum_{i=1}^{n_t} \frac{(\hat{J}_{ni} - 1)^2}{N_t \pi_i}\right| = \left(\frac{(\lambda + 1)\alpha_{n_t}}{1 - (\lambda + 1)\alpha_{n_t}}\right)^2 \sum_{i=1}^{n_t} \frac{\hat{I}_{+it}}{N_t \pi_i} + \sum_{i=1}^{n_t} \frac{\hat{I}_{-it}}{N_t \pi_i},$$

where

$$\hat{I}_{+it} = I(\alpha_{n_t} \leq \hat{F}_t(E_{Mi}) \leq 1 - \alpha_{n_t})$$

and

$$\hat{I}_{-it} = 1 - \hat{I}_{+it}.$$

Note that $\lim_{t \rightarrow \infty} \alpha_{n_t} = 0$ and that

$$\left|\sum_{i=1}^{n_t} \frac{\hat{I}_{+it}}{N_t \pi_i}\right| \leq \frac{\hat{N}_t}{N_t} = O_p(1).$$

Therefore,

$$\left(\frac{(\lambda + 1)\alpha_{n_t}}{1 - (\lambda + 1)\alpha_{n_t}}\right)^2 \sum_{i=1}^{n_t} \frac{\hat{I}_{+it}}{N_t \pi_i} = o_p(1).$$

Moreover,

$$\begin{aligned} \sum_{i=1}^{n_t} \frac{\hat{I}_{-it}}{N_t \pi_i} &= \sum_{i=1}^{N_t} \frac{\xi_{it} \hat{I}_{-it}}{N_t \pi_i}, \\ &= \sum_{i=1}^{N_t} \frac{\xi_{it} \hat{I}_{-it} I_{it}^+(\epsilon_0)}{N_t \pi_i} + \sum_{i=1}^{N_t} \frac{\xi_{it} \hat{I}_{-it} I_{it}^-(\epsilon_0)}{N_t \pi_i}, \end{aligned}$$

where $\xi_{it} = \mathbf{I}(i \in s_t)$,

$$I_{it}^+(\epsilon_0) = \mathbf{I}(|\hat{F}_t(E_{Mi}) - F_t(E_{Mi})| > \epsilon_0) \text{ and } I_{it}^-(\epsilon_0) = \mathbf{I}(|\hat{F}_t(E_{Mi}) - F_t(E_{Mi})| \leq \epsilon_0).$$

for an arbitrary strictly positive real number ϵ_0 . We have that $\forall \epsilon > 0$,

$$\begin{aligned} P\left(\left|\sum_{i=1}^{n_t} \frac{\hat{I}_{-it}}{N_t \pi_i}\right| > \epsilon\right) &\leq P\left(\left|\sum_{i=1}^{N_t} \frac{\xi_{it} \hat{I}_{-it} I_{it}^+(\epsilon_0)}{N_t \pi_i}\right| > \epsilon/2\right) + \\ &+ P\left(\left|\sum_{i=1}^{N_t} \frac{\xi_{it} \hat{I}_{-it} I_{it}^-(\epsilon_0)}{N_t \pi_i}\right| > \epsilon/2\right). \end{aligned}$$

It follows from Chebychev's inequality that,

$$P\left(\left|\sum_{i=1}^{N_t} \frac{\xi_{it} \hat{I}_{-it} I_{it}^+(\epsilon_0)}{N_t \pi_i}\right| > \epsilon/2\right) \leq \frac{2}{\epsilon} \frac{1}{N_t} \sum_{i=1}^{N_t} P(|\hat{F}_t(E_{Mi}) - F_t(E_{Mi})| > \epsilon_0). \quad (2.10)$$

But,

$$\lim_{t \rightarrow \infty} \sup_{1 \leq i \leq N_t} P(|\hat{F}_t(E_{Mi}) - F_t(E_{Mi})| > \epsilon_0) = 0.$$

as a direct consequence of assumptions A1 and A2. Now we can use lemma 2.5 to conclude that the left hand side of 2.10 goes to 0.

It remains to show that,

$$\lim_{t \rightarrow \infty} P\left(\left|\sum_{i=1}^{N_t} \frac{\xi_{it} \hat{I}_{-it} I_{it}^-(\epsilon_0)}{N_t \pi_i}\right| > \epsilon/2\right) = 0.$$

Appealing to Chebychev's inequality again, we have that

$$P\left(\left|\sum_{i=1}^{N_t} \frac{\xi_{it} \hat{I}_{-it} I_{it}^-(\epsilon_0)}{N_t \pi_i}\right| > \epsilon/2\right) \leq 2\epsilon^{-1} \sum_{i=1}^{N_t} \frac{E \hat{I}_{-it} I_{it}^-(\epsilon_0)}{N_t}.$$

But $\hat{I}_{-it} I_{it}^-(\epsilon_0) = I_{it}^-(\epsilon_0) - \hat{I}_{+it} I_{it}^-(\epsilon_0)$. This implies that,

$$\begin{aligned} P\left(\left|\sum_{i=1}^{N_t} \frac{\xi_{it} \hat{I}_{-it} I_{it}^-(\epsilon_0)}{N_t \pi_i}\right| > \epsilon/2\right) &\leq 2\epsilon^{-1} \left\{ \frac{1}{N_t} \sum_{i=1}^{N_t} P\left(|\hat{F}_t(E_{Mi}) - F_t(E_{Mi})| \leq \epsilon_0\right) - \right. \\ &\quad \left. - \frac{1}{N_t} \sum_{i=1}^{N_t} P(\hat{I}_{+it} I_{it}^-(\epsilon_0) = 1) \right\}. \end{aligned} \quad (2.11)$$

Note that

$$\left. \begin{aligned} \alpha_{n_t} + \epsilon_0 \leq F_t(E_{Mi}) \leq 1 - (\alpha_{n_t} + \epsilon_0) \\ |\hat{F}_t(E_{Mi}) - F_t(E_{Mi})| \leq \epsilon_0 \end{aligned} \right\} \implies \hat{I}_{+it} I_{it}^-(\epsilon_0) = 1 \quad (2.12)$$

and let $\delta_{it}(\epsilon_0)$ be defined by

$$\delta_{it}(\epsilon_0) = \begin{cases} 1 & \text{if } \alpha_{n_t} + \epsilon_0 \leq F_t(E_{Mi}) \leq 1 - (\alpha_{n_t} + \epsilon_0), \\ 0 & \text{otherwise.} \end{cases}$$

Now, we have that,

$$\begin{aligned} \frac{1}{N_t} \sum_{i=1}^{N_t} P(\hat{I}_{+it} I_{it}^-(\epsilon_0) = 1) &= \frac{1}{N_t} \sum_{i=1}^{N_t} \delta_{it}(\epsilon_0) P(\hat{I}_{+it} I_{it}^-(\epsilon_0) = 1) + \\ &\quad + \frac{1}{N_t} \sum_{i=1}^{N_t} (1 - \delta_{it}(\epsilon_0)) P(\hat{I}_{+it} I_{it}^-(\epsilon_0) = 1). \end{aligned}$$

It follows from 2.12 and 2.11 that,

$$\begin{aligned} P\left(\left|\sum_{i=1}^{N_t} \frac{\xi_{it} \hat{I}_{-it} I_{it}^-(\epsilon_0)}{N_t \pi_i}\right| > \epsilon/2\right) &\leq \frac{2}{\epsilon} \frac{1}{N_t} \sum_{i=1}^{N_t} (1 - \delta_{it}(\epsilon_0)) \left\{ P(|\hat{F}_t(E_{Mi}) - F_t(E_{Mi})| \leq \epsilon_0) - \right. \\ &\quad \left. - P(\hat{I}_{+it} I_{it}^-(\epsilon_0) = 1) \right\} \\ &\leq \frac{4}{\epsilon} \frac{1}{N_t} \sum_{i=1}^{N_t} (1 - \delta_{it}(\epsilon_0)). \end{aligned}$$

Note that the $F_t(E_{Mi})$'s are uniformly distributed in the set $\{1/N_t, 2/N_t, \dots, 1\}$.

Let F_{U_t} be defined for all x by,

$$F_{U_t}(x) = \frac{1}{N_t} \sum_{i=1}^{N_t} I(F_t(E_{Mi}) \leq x).$$

The result follows from the tightness of the sequence $\{F_{U_t}\}_{t \geq 1}$ (see lemma 2.4 .

■

Theorem 2.5. Assume that the conditions of lemma 2.6 are all satisfied as well as the following ones:

- (i) $\lim_{t \rightarrow \infty} B_{Mt} = \beta_0$ and $\lim_{t \rightarrow \infty} \bar{Y}_t = \mu_0$.
- (ii) $\hat{B}_M = B_M + o_p(1)$.
- (iii) $\frac{1}{N_t} \sum_{i=1}^{N_t} y_i^2 = O(1)$.

Then $\hat{Y}_{LM}$ is a consistent estimator of $\bar{Y}$.

Proof.

We have to prove that for all $\epsilon > 0$, $\lim_{t \rightarrow \infty} P(|\hat{Y}_{LM} - \bar{Y}| > \epsilon) = 0$. We recall that the LM-estimator of $\bar{Y}$ is given by:

$$\hat{Y}_{LM} = \hat{B}_M + \sum_{i=1}^n \frac{e_{Mi}^{**}}{\hat{N}\pi_i} \text{ where } e_{Mi}^{**} = [w_n + (1 - w_n)\hat{J}_{ni}]e_{Mi}.$$

We have that

$$\begin{aligned} \hat{Y}_{LM} - \bar{Y} &= (1 - w_n) \left[(\hat{B}_M - B_M) \left(1 - \sum_{i=1}^n \frac{\hat{J}_{ni}}{\hat{N}\pi_i} \right) + B_M \left(1 - \sum_{i=1}^n \frac{\hat{J}_{ni}}{\hat{N}\pi_i} \right) + \right. \\ &\quad \left. - \sum_{i=1}^n \frac{1 - \hat{J}_{ni}}{\hat{N}\pi_i} y_i \right] + \left(\sum_{i=1}^n \frac{y_i}{\hat{N}\pi_i} - \bar{Y} \right). \end{aligned}$$

Therefore, for all $\epsilon > 0$,

$$\begin{aligned}
 P(|\widehat{Y}_{LM} - \bar{Y}| > \epsilon) &= \\
 &= P\left(|\widehat{B}_M - B_M| \left|1 - \sum_{i=1}^n \frac{\hat{J}_{ni}}{\widehat{N}\pi_i}\right| > \epsilon/4\right) + P\left(B_M \left|1 - \sum_{i=1}^n \frac{\hat{J}_{ni}}{\widehat{N}\pi_i}\right| > \epsilon/4\right) + \\
 &+ P\left(\left|\sum_{i=1}^n \frac{1 - \hat{J}_{ni}}{\widehat{N}\pi_i} y_i\right| > \epsilon/4\right) + P\left(\left|\sum_{i=1}^n \frac{y_i}{\widehat{N}\pi_i} - \bar{Y}\right| > \epsilon/4\right), \\
 &= p_{1t}(\epsilon) + p_{2t}(\epsilon) + p_{3t}(\epsilon) + p_{4t}(\epsilon).
 \end{aligned}$$

We have to show that $\lim_{t \rightarrow \infty} p_{1t}(\epsilon) = \lim_{t \rightarrow \infty} p_{2t}(\epsilon) = \lim_{t \rightarrow \infty} p_{3t}(\epsilon) = \lim_{t \rightarrow \infty} p_{4t}(\epsilon) = 0$. The fact that $\lim_{t \rightarrow \infty} p_{1t}(\epsilon) = 0$ follows from assumption (ii) and the lemma 2.6. The lemma 2.6 and the assumption (i) imply that $\lim_{t \rightarrow \infty} p_{2t}(\epsilon) = 0$. The assumptions A1 and A2 evoked in lemma 2.6 ensure that $\lim_{t \rightarrow \infty} p_{4t}(\epsilon) = 0$. Concerning $p_{3t}(\epsilon)$, note that.

$$\left|\sum_{i=1}^{n_t} \frac{1 - \hat{J}_{ni}}{\widehat{N}_t \pi_i} y_i\right| \leq \left(\sum_{i=1}^{N_t} \frac{(1 - \hat{J}_{ni})^2}{\widehat{N}_t \pi_i}\right)^{1/2} \times \left(\sum_{i=1}^{n_t} \frac{y_i^2}{\widehat{N}_t \pi_i}\right)^{1/2}$$

which goes to 0 as t goes to ∞ .

■

In order to study the asymptotic variance of the LM-estimator, it is essential that linear representations of $\widehat{B}_M$ and $\Theta(\widehat{\widehat{F}})$ be available. The linear representation of $\widehat{B}_M$ is known and is similar to that of $\widehat{B}_{CM}$. It remains to linearize $\Theta(\widehat{\widehat{F}})$. Our efforts now are directed towards proving the following relation.

$$\Theta(\widehat{\widehat{F}}) = (1 - w_n)\Lambda(F) + \sum_{i=1}^n \frac{E_{Mi}^{**}}{\widehat{N}\pi_i} - \sum_{i=1}^n \frac{Q_{Mi}}{\widehat{N}\pi_i} + o_p(n^{-1/2}),$$

where for $E_{Mi} = y_i - B_M$ we have $E_{Mi}^{**} = w_n E_{Mi} + (1 - w_n)Z_i$,

$$Q_{Mi} = \frac{\psi_H[\sigma_M^{-1} E_{Mi}]}{N^{-1} \sum_{i=1}^N \psi'_H[\sigma_M^{-1} E_{Mi}]}, \quad \Lambda(F) = \int r J_n[F(r)] dF(r),$$

and Z_i is defined as follows:

$$Z_i = - \int [\Delta(E_{Mi} \leq r) - F_t(r)] J_n[F_t(r)] dr.$$

It is important to stress that the population mean $\bar{Z}$ of the z -variable is 0. It is straightforward to demonstrate it. We start by stating and proving the following lemma:

Lemma 2.7. *Assume that conditions A1 and A2 are satisfied and that the sequence of distribution functions $\{F_t\}_{t \geq 1}$ is tight.*

Then,

$$\Lambda(\hat{F}_t) = \Lambda(F_t) + \bar{z} + o_p(n^{-1/2}), \text{ where } \bar{z} = \sum_{i=1}^{n_t} \frac{Z_i}{\hat{N}\pi_i}.$$

Proof:

Note that

$$\begin{aligned} \Lambda(\hat{F}_t) &= \int x J_n[\hat{F}_t(x)] d\hat{F}_t(x), \\ &= \int x dK_n[\hat{F}_t(x)], \end{aligned}$$

where

$$K_n(t) = \int_0^t J_n(s) ds.$$

Likewise,

$$\Lambda(F_t) = \int x dK_n[F_t(x)].$$

For the J_n functions we are using (see 2.8 and 2.9), K_n is usually bounded as J_n vanishes outside the interval $[\lambda\alpha_{n_t}, 1 - \alpha_{n_t}]$. For the J_n function given by 2.8 for

example. K_n is defined by:

$$K_n(t) = \frac{t - \lambda\alpha_{n_t}}{1 - (\lambda + 1)\alpha_{n_t}} \Delta(\lambda\alpha_{n_t} \leq t \leq 1 - \alpha_{n_t}) + \Delta(t > 1 - \alpha_{n_t}).$$

By integrating by parts one obtains

$$\begin{aligned} \Lambda(\widehat{F}_t) - \Lambda(F_t) &= \int x d\left\{ K_n[\widehat{F}_t(x)] - K_n[F_t(x)] \right\}, \\ &= \lim_{A \rightarrow +\infty} x \left\{ K_n[\widehat{F}_t(x)] - K_n[F_t(x)] \right\} \Big|_{-A}^A - \int \left\{ K_n[\widehat{F}_t(x)] - K_n[F_t(x)] \right\} dx, \\ &= - \int \left\{ K_n[\widehat{F}_t(x)] - K_n[F_t(x)] \right\} dx. \end{aligned}$$

Note that if A is large enough, we will have $\widehat{F}_t(A) > 1 - \alpha_{n_t}$, $F_t(A) > 1 - \alpha_{n_t}$.

$\widehat{F}_t(-A) < \lambda\alpha_{n_t}$ and $F_t(-A) < \lambda\alpha_{n_t}$. This implies that $K_n[\widehat{F}_t(A)] - K_n[F_t(A)] = 0$ and $K_n[\widehat{F}_t(-A)] - K_n[F_t(-A)] = 0$. Furthermore,

$$\begin{aligned} \bar{z} &= - \sum_{i=1}^{n_t} \frac{1}{\widehat{N}_t \pi_i} \int [\Delta(E_{Mi} \leq r) - F_t(r)] J_n[F_t(r)] dr, \\ &= - \int \left[\sum_{i=1}^{n_t} \frac{\Delta(E_{Mi} \leq r)}{\widehat{N}_t \pi_i} - F_t(r) \right] J_n[F_t(r)] dr, \\ &= - \int [\widehat{F}_t(r) - F_t(r)] J_n[F_t(r)] dr. \end{aligned}$$

Therefore,

$$\begin{aligned} \Lambda(\widehat{F}_t) - \Lambda(F_t) - \bar{z} &= \\ &= - \int \left\{ K_n[\widehat{F}_t(r)] - K_n[F_t(r)] - J_n[F_t(r)] [\widehat{F}_t(r) - F_t(r)] \right\} dr, \\ &= - \int W_t(r) [\widehat{F}_t(r) - F_t(r)] dr, \end{aligned}$$

where

$$W_t(r) = \begin{cases} \frac{K_n[\widehat{F}_t(r)] - K_n[F_t(r)]}{\widehat{F}_t(r) - F_t(r)} - J_n[F_t(r)], & \text{if } \widehat{F}_t(r) \neq F_t(r), \\ 0 & \text{otherwise.} \end{cases}$$

We recall that what we have to prove is that for every $\eta > 0$, there exist an integer $n(\eta)$ and a constant $K(\eta)$ such that if $n \geq n(\eta)$ then

$$P\left(n \left| \Lambda(\widehat{F}_t) - \Lambda(F_t) - \bar{z} \right| \leq K(\eta)\right) > 1 - \eta.$$

Let $\eta > 0$ be an arbitrary positive real number. It follows from the tightness of the sequence $\{F_t\}_{t \geq 1}$ that there exists $\lambda_\eta > 0$ such that,

$$\limsup_{t \rightarrow \infty} 1 - F_t(\lambda_\eta) + F_t(-\lambda_\eta) \leq \eta.$$

Define the following events,

$$\mathcal{B}_t(\eta) = \left\{ n \left| \Lambda(\widehat{F}_t) - \Lambda(F_t) - \bar{z} \right| \leq K(\eta) \right\},$$

$$\mathcal{A}_t = \{s_t \in \mathcal{S}_t : \widehat{F}_t(-\lambda_\eta) < \lambda \alpha_{n_t}, \widehat{F}_t(\lambda_\eta) > 1 - \alpha_{n_t} \text{ \& } \widehat{N}_t \min_{1 \leq i \leq N_t} \pi_i \leq N_t\}.$$

where λ_η is the constant defined in assumption (i) of this theorem and $\mathcal{S}_t$ is the set of all possible samples from population $\mathcal{U}_t$. Our problem is to show that if $n \geq n(\eta)$ then

$$P(\mathcal{B}_t|\mathcal{A}_t)P(\mathcal{A}_t) + P(\mathcal{B}_t|\mathcal{A}'_t)P(\mathcal{A}'_t) > 1 - \eta,$$

where $\mathcal{A}'_t$ is the complement of $\mathcal{A}_t$. Note that,

$$\begin{aligned} P(\mathcal{B}_t|\mathcal{A}_t)P(\mathcal{A}_t) + P(\mathcal{B}_t|\mathcal{A}'_t)P(\mathcal{A}'_t) &= P(\mathcal{B}_t|\mathcal{A}_t) - \left(P(\mathcal{B}_t|\mathcal{A}_t) - P(\mathcal{B}_t|\mathcal{A}'_t) \right) P(\mathcal{A}'_t). \\ &\geq P(\mathcal{B}_t|\mathcal{A}_t) - P(\mathcal{A}'_t). \end{aligned}$$

But $\lim_{t \rightarrow \infty} P(\mathcal{A}_t) = 1$. This implies that there exists an integer $n_0(\eta)$ such that if $n \geq n_0(\eta)$ then

$$P(\mathcal{A}'_t) < \eta/2 \text{ or } -P(\mathcal{A}'_t) > -\eta/2.$$

Therefore, it is enough to prove that $P(\mathcal{B}_t|\mathcal{A}_t) > 1 - \eta/2$. Under the event $\mathcal{A}_t$ we have that:

$$\Lambda(\widehat{F}_t) - \Lambda(F_t) - \bar{z} = \int_{-\lambda_\eta}^{\lambda_\eta} W_t(r)(\widehat{F}_t(r) - F_t(r))dr.$$

It follows from Schwartz inequality that

$$|\Lambda(\widehat{F}_t) - \Lambda(F_t) - \bar{z}| \leq \left(\int_{-\lambda_\eta}^{\lambda_\eta} W_t^2(r)dr \right)^{1/2} \left(\int_{-\lambda_\eta}^{\lambda_\eta} [\widehat{F}_t(r) - F_t(r)]^2 dr \right)^{1/2}.$$

To complete the proof, it is enough to prove the following two statements

$$(A) \quad \int_{-\lambda_\eta}^{\lambda_\eta} [\widehat{F}_t(r) - F_t(r)]^2 dr = O_p(n^{-1}),$$

$$(B) \quad \int_{-\lambda_\eta}^{\lambda_\eta} W_t^2(r)dr = o_p(1).$$

- To prove (A), note that by Markov's inequality

$$P\left(\int_{-\lambda_\eta}^{\lambda_\eta} [\widehat{F}_t(r) - F_t(r)]^2 dr \leq \alpha\right) \geq 1 - \frac{1}{\alpha} E\left(\int_{-\lambda_\eta}^{\lambda_\eta} [\widehat{F}_t(r) - F_t(r)]^2 dr\right).$$

By using Fubini's theorem one gets

$$E\left(\int_{-\lambda_\eta}^{\lambda_\eta} [\widehat{F}_t(r) - F_t(r)]^2 dr\right) = \int_{-\lambda_\eta}^{\lambda_\eta} V[\widehat{F}_t(r)]dr.$$

It follows from assumptions A1 and A2 that

$$\int_{-\lambda_\eta}^{\lambda_\eta} V(\widehat{F}_t(r))dr = O(n_t^{-1}).$$

This completes the proof of (A).

- It remains to prove (B). To this end, note that

$$\forall \epsilon > 0. \quad P\left(\int_{-\lambda_\eta}^{\lambda_\eta} W_t^2(r)dr > \epsilon\right) \leq \frac{1}{\epsilon} E\left(\int_{-\lambda_\eta}^{\lambda_\eta} W_t^2(r)dr\right).$$

It is possible to apply the Lebesgue dominated convergence theorem to the right side of the above inequality. First of all we like to check the boundedness of the integrand and to calculate its limit.

a)

$$\begin{aligned} \int_{-\lambda_\eta}^{\lambda_\eta} W_t^2(r) dr &\leq \int_{-\lambda_\eta}^{\lambda_\eta} \left[\frac{1}{|\widehat{F}_t(r) - F_t(r)|} \int_{F_t(r)}^{\widehat{F}_t(r)} J_n(t) dt + J_n[F_t(r)] \right] dr. \\ &\leq 4\lambda_\eta \|J_n\|_\infty. \end{aligned}$$

b)

$$\begin{aligned} \lim_{t \rightarrow \infty} \int_{-\lambda_\eta}^{\lambda_\eta} W_t^2(r) dr &= \int_{-\lambda_\eta}^{\lambda_\eta} \lim_{t \rightarrow \infty} W_t^2(r) dr. \\ &= \int_{-\lambda_\eta}^{\lambda_\eta} \left\{ K'_n[F_t(r)] - J_n[F_t(r)] \right\} dr = 0. \end{aligned}$$

$K'_n(\cdot)$ is the first-order derivative evaluated at the given point. Hence,

$$\lim_{t \rightarrow \infty} P\left(\int_{-\lambda_\eta}^{\lambda_\eta} W_t^2(r) dr > \epsilon\right) \leq \frac{1}{\epsilon} E \lim_{t \rightarrow \infty} \int_{-\lambda_\eta}^{\lambda_\eta} W_t^2(r) dr = 0.$$

This completes the proof.

■

Theorem 2.6. Assume that all the conditions of lemma 2.7 are satisfied as well as condition (i) of theorem 2.2 . If A1 and A2 are also satisfied then,

$$\begin{aligned} \Theta(\widehat{F}_t) &= (1 - w_n)\Lambda(F_t) + \sum_{i=1}^{n_t} \frac{E_{Mi}^{**}}{\widehat{N}_t \pi_i} - \sum_{i=1}^{n_t} \frac{Q_{Mi}}{\widehat{N}_t \pi_i} + o_p(n_t^{-1/2}), \\ \widehat{Y}_{LM} &= [B_M + (1 - w_n)\Lambda(F)] + \sum_{i=1}^{n_t} \frac{E_{Mi}^{**}}{\widehat{N}_t \pi_i} + o_p(n_t^{-1/2}), \end{aligned}$$

where

$$E_{Mi}^{**} = w_n E_{Mi} + (1 - w_n) Z_i.$$

Proof.

Note that $\hat{J}_{ni} = \hat{J}_{ni}$. In fact,

$$\begin{aligned}\hat{J}_{ni} &= J_n \left[\hat{F}_t(r_i) \right] = J_n \left(\sum_{k=1}^{n_t} \frac{\Delta(y_k - \hat{B}_M \leq y_i - \hat{B}_M)}{\hat{N}_t \pi_k} \right), \\ &= J_n \left(\sum_{k=1}^{n_t} \frac{\Delta(y_k - B_M \leq y_i - B_M)}{\hat{N}_t \pi_k} \right) = \hat{J}_{ni}.\end{aligned}$$

First of all, let us show that

$$\Lambda(\hat{\hat{F}}_t) = \Lambda(\hat{F}_t) + \sum_{i=1}^n \frac{Z_i}{\hat{N}_t \pi_i} - \sum_{i=1}^n \frac{Q_{Mi}}{\hat{N}_t \pi_i} + o_p(n^{-1/2}). \quad (2.13)$$

Since the conditions of lemma 2.7 hold, it is enough to prove that

$$\Lambda(\hat{\hat{F}}_t) = \Lambda(\hat{F}_t) - \sum_{i=1}^n \frac{Q_{Mi}}{\hat{N}_t \pi_i} + O_p(n^{-1}).$$

We have

$$\begin{aligned}\Lambda(\hat{\hat{F}}_t) &= \sum_{i=1}^{n_t} \frac{\hat{J}_{ni}}{\hat{N}_t \pi_i} (e_{Mi} - E_{Mi}) + \sum_{i=1}^{n_t} \frac{\hat{J}_{ni}}{\hat{N}_t \pi_i} E_{Mi}, \\ &= -(\hat{B}_M - B_M) \sum_{i=1}^{n_t} \frac{\hat{J}_{ni}}{\hat{N}_t \pi_i} + \sum_{i=1}^{n_t} \frac{\hat{J}_{ni}}{\hat{N}_t \pi_i} E_{Mi}.\end{aligned}$$

Since $\hat{J}_{ni} = \hat{J}_{ni}$, the second summation above equals $\Lambda(\hat{F}_t)$. But,

$$-(\hat{B}_M - B_M) \sum_{i=1}^{n_t} \frac{\hat{J}_{ni}}{\hat{N}_t \pi_i} = (\hat{B}_M - B_M) \left(1 - \sum_{i=1}^{n_t} \frac{\hat{J}_{ni}}{\hat{N}_t \pi_i} \right) - (\hat{B}_M - B_M).$$

Therefore,

$$\begin{aligned}\Lambda[\hat{\hat{F}}_t] &= \Lambda(\hat{F}_t) - (\hat{B}_M - B_M) + (\hat{B}_M - B_M) \left(1 - \sum_{i=1}^{n_t} \frac{\hat{J}_{ni}}{\hat{N}_t \pi_i} \right), \\ &= \Lambda(\hat{F}_t) - \sum_{i=1}^{n_t} \frac{Q_{Mi}}{\hat{N}_t \pi_i} + O_p(n_t^{-1/2}) o_p(1) + o_p(n_t^{-1/2}).\end{aligned}$$

This follows from lemma 2.6 and theorem 2.2 . This proves equation 2.13 .

Now note that,

$$\Theta[\widehat{\widehat{F}}_t] = w_n \mu(\widehat{\widehat{F}}_t) + (1 - w_n) \Lambda(\widehat{\widehat{F}}_t).$$

It follows from theorem 2.2 that

$$\mu(\widehat{\widehat{F}}_t) = \sum_{i=1}^n \frac{E_{Mi}}{\widehat{N}\pi_i} - \sum_{i=1}^n \frac{Q_{Mi}}{\widehat{N}\pi_i} + o_p(n^{-1/2}). \quad (2.14)$$

Combining equations 2.14 and 2.13 one gets the result.

■

We are now in a position to derive the asymptotic variance of the LM-estimator.

It follows straightforwardly from theorem 2.6 that the asymptotic variance of the LM-estimator is given by:

$$V_a(\widehat{\widehat{Y}}_{LM}) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) E_{Mi}^{**} E_{Mj}^{**}. \quad (2.15)$$

For large samples, the bias of the LM-estimator is approximately equal to

$$B_a(\widehat{\widehat{Y}}_{LM}) = -(1 - w_n)(\overline{E}_M - \overline{E}_{\alpha_{n_t}}).$$

An asymptotically unbiased estimator of this variance is given by:

$$\widehat{V}_a(\widehat{\widehat{Y}}_{LM}) = \frac{1}{N^2} \sum_{i=1}^n \sum_{j=1}^n \frac{1}{\pi_{ij}} \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) e_{Mi}^{**} e_{Mj}^{**},$$

where the $E_{Mi}^{**} = w_n E_{Mi} + (1 - w_n) Z_i$, $e_{Mi}^{**} = w_n e_{Mi} + (1 - w_n) z_i$ and the z_i 's are obtained from the Z_i 's by replacing the theoretical residuals E_{Mi} 's by the e_{Mi} 's and the population distribution function F by $\widehat{F}$.

It is important to stress that for small samples, the bias of the LM-estimator might not be negligible. In such a case, it is possible to estimate the asymptotic MSE of the LM-estimator as follows:

$$\text{amse}(\widehat{Y}_{\text{LM}}) = \widehat{V}_a(\widehat{Y}_{\text{LM}}) + b_2,$$

where $b_2 = (1 - w_n)^2 \max[0, (\widehat{E}_{\text{M}} - \widehat{E}_{\alpha_{n_t}})^2 - \widehat{V}(\widehat{E}_{\text{M}} - \widehat{E}_{\alpha_{n_t}})]$,

$$\widehat{E}_{\text{M}} = \sum_{i=1}^n \frac{e_{\text{Mi}}}{\widehat{N}\pi_i}, \quad \text{and} \quad \widehat{E}_{\alpha_{n_t}} = \sum_{i=1}^n \frac{\hat{j}_{ni} e_{\text{Mi}}}{\widehat{N}\pi_i}.$$

As for the variance of the difference between these two estimators, we can take

$$\widehat{V}(\widehat{E}_{\text{M}} - \widehat{E}_{\alpha_{n_t}}) \simeq \widehat{V}\left(\sum_{i=1}^n \frac{e_{\text{Mi}} - z_i}{\widehat{N}\pi_i}\right).$$

To fully specify the variances, the z_i 's must be expressed in a form that can easily be used for estimation purposes. If $J_n(t) = (1 - 2\alpha_{n_t})^{-1} \mathbf{I}(\alpha_{n_t} < t \leq 1 - \alpha_{n_t})$, then

$$Z_i = \begin{cases} -\overline{E}_{\alpha_{n_t}} - \frac{\alpha_{n_t} F^{-1}(1 - \alpha_{n_t}) - (1 - \alpha_{n_t}) F^{-1}(\alpha_{n_t}^*)}{(1 - 2\alpha_{n_t})} & \text{if } F(E_{\text{Mi}}) \leq \alpha_{n_t}, \\ -\overline{E}_{\alpha_{n_t}} + \frac{E_{\text{Mi}} - \alpha_{n_t} [F^{-1}(1 - \alpha_{n_t}) + F^{-1}(\alpha_{n_t}^*)]}{1 - 2\alpha_{n_t}} & \text{if } \alpha_{n_t} < F(E_{\text{Mi}}) \leq 1 - \alpha_{n_t}, \\ -\overline{E}_{\alpha_{n_t}} + \frac{(1 - \alpha_{n_t}) F^{-1}(1 - \alpha_{n_t}) - \alpha_{n_t} F^{-1}(\alpha_{n_t}^*)}{1 - 2\alpha_{n_t}} & \text{if } F(E_{\text{Mi}}) > 1 - \alpha_{n_t} \end{cases}$$

$\alpha_{n_t}^* = \alpha_{n_t} + 1/\widehat{n}$ and z_i can be obtained by replacing F by $\widehat{F}$.

2.5. Simulation Study

2.5.1. Objectives

The main objective of this Monte Carlo experiment is to study the robustness properties of the M-, the CM- and the LM- estimators for symmetric long-tailed populations and highly skewed populations and for various sample sizes. We will be looking at the efficiency robustness as well as the bias robustness relative to the GREG estimator. We have also evaluated the accuracy of the formulas for the asymptotic variances (see 2.7 and 2.15) and the properties of the corresponding variance estimators. The biases of the MSE estimators developed for the CM- and the LM-estimators are studied as well. In this study, we confine ourselves to the simple random sampling without replacement (SRSWOR) although it is possible to extend the experiment to more general sampling designs. Also no auxiliary information is used to improve on the estimators. Therefore estimators that require auxiliary information are not considered in this study.

In addition to the GREG, the CM and the LM estimators, we have included in this simulation study a number of other robust estimators that are proposed in the literature. One of these, is the bias-adjusted M-estimator proposed by Chambers (1986) within the model-based framework. It is given by

$$\widehat{Y}_{CH} = f\bar{y} + (1 - f)\left\{\widehat{B}_M + \widehat{S}_M\psi_x[(y_i - \widehat{B}_M)/\widehat{S}_M]\right\},$$

where $f = n/N$ is simply the sampling fraction.

Other estimators were proposed within the same context of model-based inference by Fuller (1991) and Rivest (1994). Fuller proposed a general class of r th Winsorized mean given by

$$\widehat{Y}_{w_r} = f\bar{y} + (1 - f)\widehat{W}_r,$$

where $\bar{y}$ is the sample mean, f the sampling fraction and

$$\widehat{W}_r = n^{-1} \left(\sum_{j=1}^{n-r} y_{(j)} + ry_{(n-r)} \right).$$

the $y_{(j)}$'s being the ordered statistics. In our simulation we have included only the once-Winsorized mean $\widehat{Y}_{w_1}$ and the twice-Winsorized mean $\widehat{Y}_{w_2}$. Another estimator proposed by Fuller follows a preliminary test of the shape parameter of the Weibull distribution. This estimator is not included in our simulation study.

Another estimator used in this simulation belongs to the general class of Winsorized means proposed by Rivest (1994). He proposed to replace the largest sample unit $y_{(n)}$ by a linear combination of the two largest ordered statistics $y_{(n-1)}$ and $y_{(n)}$. One estimator belonging to this class and suggested by Rivest (1994) is given by

$$\widehat{Y}_{w_\alpha} = f\bar{y} + (1 - f)\widehat{W}_\alpha \quad (0 < \alpha < 1)$$

where $\widehat{W}_\alpha$ is given by

$$\widehat{W}_\alpha = n^{-1} \left(\sum_{j=1}^{n-1} y_{(j)} + (1 - \alpha)y_{(n)} + \alpha y_{(n-1)} \right).$$

In this simulation study, we have used $\alpha = 0.75$.

Well-known estimators in the literature on outlier treatment are those proposed by Hidioglou and Srinath (1981). The key idea behind the Hidioglou-Srinath approach is to reduce the weights attached to the outliers. The outlier weight might be equal to 1 or can be determined by the minimizing the mean squared error of the estimator. The reason for which these estimators are not included in our Monte Carlo study is that they have two major weaknesses. (i) The outliers are assumed to be identified. Therefore, the Hidioglou-Srinath procedure overlooks an important part in the process of treating the aberrant values. (ii) The Hidioglou-Srinath estimators are only usable if at least one outlier is identified in the sample. Consequently, the performance of these estimators can only be studied conditionally on the number of sample outliers. Given that the conditioning is on a statistic whose impact on the variable of interest can be substantial, the performance of these estimators is still questionable.

2.5.2. Description of the experiment

The primary reason for developing the CM- and the LM- estimators was to improve on the standard M-estimator which is usually heavily biased. It is then natural to know how efficient are the proposed estimators when applied to ideal populations for which the M-estimator performs well. These are typically, populations where the outliers are symmetrically distributed. Simulation results reported below indicate that the M-estimator is very efficient for symmetric populations and that the lost

in efficiency of the CM and LM estimators with respect to the M-estimator is fairly small. Five symmetric populations each of size 1000, denoted by POP I, POP II, POP III, POP IV and POP V, have been generated according to five different distributions. These populations are described below.

Table 2.1. Description of the Symmetric Finite Populations Generated From the Superpopulation Model $y_i = \mu + \epsilon_i$.

Population	Distribution of ϵ_i	Parameters
POP I	$\mathcal{N}(0, \sigma^2)$	$\mu = 1,550 \sigma^2 = 22,500$
POP II	$\mathcal{D}(0, \sigma^2)$	$\mu = 1,550 \sigma^2 = 22,500$
POP III	$\mathcal{CN}(0, \sigma^2, \epsilon, \lambda)$	$\mu = 1,550 \sigma^2 = 625 \epsilon = 1\% \lambda = 100$
POP IV	$\mathcal{CN}(0, \sigma^2, \epsilon, \lambda)$	$\mu = 1,550 \sigma^2 = 625 \epsilon = 2\% \lambda = 100$
POP V	$\mathcal{CN}(0, \sigma^2, \epsilon, \lambda)$	$\mu = 1,550 \sigma^2 = 625 \epsilon = 4\% \lambda = 100$

In table 2.1. $\mathcal{N}(\mu, \sigma^2)$ stands for the normal distribution with mean μ and variance σ^2 . $\mathcal{D}(\mu, \sigma^2)$ is the double exponential distribution with mean μ and scale parameter σ^2 and $\mathcal{CN}(\mu, \sigma^2, \epsilon, \lambda) = (1 - \epsilon)\mathcal{N}(\mu, \sigma^2) + \epsilon\mathcal{N}(\mu, \lambda\sigma^2)$.

The skewed population used here is a natural population. It consists of the annual revenues of the 284 municipalities in Sweden (see Särndal et al., 1992, page 652). This population is referred to as the MU284 population. The revenues are from the 1985 municipal taxation. This population has the feature of being extremely skewed because of the presence of three large municipalities.

The population mean which is the parameter of interest is $\bar{Y} = 244.99$. The population standard deviation is $S_y = 596.31$ where

$$S_y^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{Y})^2.$$

The parameter B_M is equal to 151.71. The difference between B_M and $\bar{Y}$ confirms the skewness of the population.

Concerning the CM-estimator, we have taken $\alpha = 2$ in the definition of K_n for the MU284 data and $\alpha = 5$ for the symmetric populations. Normally, as mentioned in section 2.4.1, one can try a number of values and select the one that minimizes the MSE of the CM-estimator. Table 2.6. contains the gain in efficiency of the CM-estimator for various values of α . It follows from these results that $\alpha = 2$ is the best choice for all sample sizes except when $n = 20$, in which case the best choice would be $\alpha = 3$. If the MU284 is taken as a census, then these guidelines can be used for subsequent surveys.

Concerning the LM-estimator we have used $\lambda = 1$ for symmetric populations and $\lambda = 3$ for the skewed MU284 population.

Our results are based on 10,000 simple random samples. In selecting the samples, we have used the following implementation of the SRSWOR design for a given and fixed finite population of size N .

- For a population of size N we select N independent uniform random numbers $\varepsilon_1, \dots, \varepsilon_N$ in the interval $(0,1)$.
 - These numbers are then sorted in ascending order. If ε_k^* 's is the ordered numbers and r_k the rank of ε_k^* in the original sample then our simple random sample is given by the population units $(r_1, \dots, r_n)$.
-

2.5.3. Performance of the Robust Estimators on Symmetric Populations

The performance criterion used in this simulation study is the gain in efficiency over the GREG estimator. The gain in efficiency $\Delta\xi$ of a given estimator $\widehat{Y}$ over the GREG estimator is given by,

$$\Delta\xi = 1 - \frac{\text{MSE}(\widehat{Y})}{\text{MSE}(\widehat{Y}_G)},$$

where MSE stands for Mean Square Error. The true MSE's here are replaced by the Monte Carlo approximation. For example, if the set of selected replicates is $\mathcal{S}$, then the Monte Carlo MSE of an estimator $\widehat{Y}$ is given by,

$$\text{MSE}(\widehat{Y}) = \frac{1}{10,000} \sum_{s \in \mathcal{S}} (\widehat{Y}(s) - \bar{Y})^2.$$

In table 2.2, the M-estimator, the CM-estimator and LM-estimator are first compared on the five symmetric populations described in table 2.1. We have also included in the simulations the model-based estimator proposed by Chambers (1986). The comparison is made in terms of the gain in efficiency over the GREG estimator. Are also included in table 2.2, the Monte Carlo variance as well as the Monte Carlo bias of each of the estimator. As expected the bias of the robust estimators is very small due to the symmetric nature of the populations under study.

It follows from table 2.2 that the gain in efficiency over the GREG estimator obtained from the robust estimators can be substantial. The population units of POP I are normally distributed. In this case, the sample mean is the best estimator

and the robust estimators yield a lost in efficiency that ranges from -0.54% for Chambers' estimator to -4.20% for the M-estimator.

The POP II units are distributed according to the double exponential distribution. The gain in efficiency of the robust estimator is substantial. The best performance is offered by the M-estimator with a gain in efficiency of about 47.44% over the GREG estimator.

The populations POP III, POP IV and POP V are contaminated normal populations with contamination rates of 1%, 2% and 4% respectively. As expected the gain in efficiency obtained from the robust estimators increases substantially as the contamination rate increases. When the contamination rate is only 1%, the gain in efficiency of all robust estimators but Chambers' estimator is greater than 100%, which is quite substantial. For a contamination rate of 4%, the gain in efficiency of the M-estimator over the GREG estimator is about 172.37%. The smallest gain is obtained by Chambers' estimator which is 105.73%.

2.5.4. Performance of the Robust Estimators on the MU284 Population

The simulation results concerning the skew population MU284 are reported in table 2.3. We have included the two estimators $\widehat{Y}_{w_1}$ and $\widehat{Y}_{w_2}$ proposed by Fuller (1991) as well the estimator $\widehat{Y}_{w_{0.75}}$ advocated by Rivest (1994) to be used for skewed populations. This table provides the gain in efficiency of the estimators, their Mean Squared Errors (MSE), their variances, their biases and their relative biases. The

relative bias of and estimator $\widehat{Y}$ is given by

$$RB(\widehat{Y}) = BIAS(\widehat{Y})/\overline{Y}.$$

It follows from table 2.3 that when the sample size is small ($n = 20$) all robust estimators perform quite well, including the M-estimator. The best performance is offered by Chambers' estimator with a gain in efficiency of 156.94% followed by the LM-estimator with a gain in efficiency of 155.44%. The estimator $\widehat{Y}_{w_{0.75}}$ has the lowest gain in efficiency of 69.28% followed by the M-estimator with a gain in efficiency of 71.17%. The small gain obtained from the M-estimator is due to its substantial bias of -90.44. The estimator $\widehat{Y}_{w_{0.75}}$ is very sensitive to large units. The LM-estimator is very good at 137.99%.

As the sample size increases, the M-estimator loses its efficiency very fast. For a sample size of 60, the M-estimator yields a dramatic loss in efficiency of -48.84% which is quite deceptive. In fact, the MSE of the M-estimator is dominated by the squared bias. This confirms that the M-estimator will only work well for populations with some symmetric structure. As mentioned earlier, this problem has motivated our study on consistent M-estimators. For $n = 60$, the proposed robust estimators offer a much better performance. The gain in efficiency of the CM-estimator is 57.48% and that of the LM-estimator is 45.20%. These two estimators outperform all their competitors except $\widehat{Y}_{w_2}$ which led to a gain of 56.43%. For larger sample sizes as 100 or 120 the CM-estimator is the best. When the sample size is 100, the CM-estimator and the LM-estimator yield a gain in efficiency over the GREG esti-

mator of 22.57% and 6.37% respectively while the gain obtained from the estimator $\widehat{Y}_{w_2}$ is 10.03%.

Chambers' estimator seems to be an improvement over the traditional M-estimator. However, for skewed populations such as MU284, the loss in efficiency for large samples might be substantial and its performance in terms of relative bias is poor. For $n = 120$ for example, the loss in efficiency of Chambers' estimator is -13.74% with a relative bias of -14.98%. These results indicate that it is desirable for a good robust M-estimator to have a bias that depends on the sample size. This explains the good performance of the CM-estimator and that of the LM-estimator.

Among the estimators proposed by Fuller and Rivest, only $\widehat{Y}_{w_2}$ seems to be competitive. Unfortunately its performance in terms of relative bias is also poor. While the main advantage of W2 is simplicity, its main disadvantage seems to be the difficulty of extending its use to more complex survey problems.

2.5.5. Accuracy of the approximation to the MSE

In this section the accuracy of the approximations to the $\text{MSE}(\widehat{Y}_{\text{CM}})$ and $\text{MSE}(\widehat{Y}_{lm})$ obtained in section 2.4 are evaluated. This is done by calculating the relative error $\mathcal{R}_e()$ of the approximations. For an estimator $\widehat{Y}$ of the population mean $\bar{Y}$,

$$\mathcal{R}_e[\text{MSE}_a(\widehat{Y})] = \frac{\text{MSE}_a(\widehat{Y}_{\text{CM}}) - \text{MSE}(\widehat{Y}_{\text{CM}})}{\text{MSE}(\widehat{Y}_{\text{CM}})},$$

where $\text{MSE}_a(\widehat{Y}_{\text{CM}})$ is the approximate MSE of the CM estimator. In table 2.4. we have reported the relative error of the MSE approximations calculated on the

MU284 population for sample size values of $n = 20, 40, 60, 80, 100, 120$. These results indicate that for almost all sample sizes, the approximations are satisfactory. Even for a sample size as small as 20, the relative errors of the GREG M CM and LM estimators range from -17.43% to 2.24%. The asymptotic formulas derived in section 2.4 for the robust estimators are quite valid even for small populations and small samples. When the sample size is as small as 20, only the CM-estimator shows a high relative error of -17.43%. This percentage is difficult to explain since the relative error for this estimator decreases substantially as the sample size increases.

2.5.6. Bias of the MSE's estimators

In section 2.4 we have proposed estimators of the MSE of the robust estimators. The percent relative biases of these MSE estimators are reported in table 2.5. If $\widehat{Y}$ is any robust estimator, then the percent relative bias of its MSE estimator $\text{mse}(\widehat{Y})$ is given by,

$$RB[\text{mse}(\widehat{Y})] = \frac{E[\text{mse}(\widehat{Y}_r)]}{MSE(\widehat{Y}_r)} - 1.$$

The population under study here is again the MU284 population. It appears clearly from table 2.5 that the proposed MSE estimators perform very well. The MSE of the CM- and LM-estimators is usually underestimated. The MSE of the CM-estimator is underestimated by about 11% for small samples. This percentage decreases substantially at $n = 100$. The underestimation for $n = 20$ might also be due to the fact that the population under study is fairly small. The LM estimator is

the only robust estimator that seems to perform consistently well. Other techniques such as the jackknife variance estimator must be investigated.

Note that $B(\hat{Y})$ is the Monte Carlo bias of the estimator, $V(\hat{Y})$ is the Monte Carlo variance, $MSE(\hat{Y})$ is the Monte Carlo MSE and $RB(\hat{Y})$ is the relative bias given by the ratio of the bias and the population mean.

Table 2.2. Performance of the Estimators on the Symmetric Populations
(10,000 replicates & Sample size (n)= 20)

Population	$\hat{Y}$	$B(\hat{Y})$	$V(\hat{Y})$	$\Delta\xi(\%)$
POP I	GREG	-0.51	1198.24	-
	M	-0.62	1250.36	-4.20
	CM	-0.19	1220.03	-1.79
	LM	-0.15	1237.03	-3.14
	CHAM	0.12	1204.69	-0.54
POP II	GREG	-0.22	1032.21	-
	M	1.42	698.62	47.33
	CM	0.90	774.81	33.08
	LM	0.55	861.91	19.72
	CHAM	0.48	854.62	20.75
POP III	GREG	-0.01	77.45	-
	M	-0.96	36.60	106.46
	CM	-0.85	37.81	101.04
	LM	-0.97	36.76	105.43
	CHAM	-0.65	43.12	77.89
POP IV	GREG	-0.03	84.71	-
	M	-1.00	37.31	121.07
	CM	-0.90	39.09	112.25
	LM	-1.02	38.51	114.21
	CHAM	-0.71	45.43	84.37
POP V	GREG	-0.03	126.92	-
	M	-2.23	41.65	172.37
	CM	-1.95	46.84	150.61
	LM	-1.95	52.20	126.67
	CHAM	-1.50	59.43	105.73

Table 2.3. Performance of the Estimators on the MU284 Population

n	$\hat{Y}$	$B(\hat{Y})$	$V(\hat{Y})$	$MSE(\hat{Y})$	$RB(\hat{Y})$	$\Delta\xi(\%)$
20	GREG	-0.75	16639.70	16639.70	-0.31	-
	M	-90.44	1542.14	9720.91	-36.91	71.17
	CM	-44.54	5007.99	6991.85	-18.18	137.99
	LM	-50.61	3952.34	5514.16	-20.66	155.44
	W1	-50.83	5431.21	8015.29	-20.75	107.60
	W2	-69.06	2400.32	7168.95	-28.19	132.11
	W075	-25.79	9164.47	9829.76	-10.53	69.28
	CHAM	-57.14	3211.72	6476.20	-23.32	156.94
60	GREG	-0.52	4640.93	4640.93	-0.21	-
	M	-93.14	394.75	9070.62	-38.02	-48.84
	CM	-31.84	1933.26	2946.99	-13.00	57.48
	LM	-36.20	1885.51	3196.24	-14.78	45.20
	W1	-28.16	3119.24	3912.06	-11.49	18.63
	W2	-41.80	1219.57	2966.80	-17.06	56.43
	W075	-14.34	3658.31	3863.91	-5.85	20.11
	CHAM	-50.04	1082.48	3586.22	-20.43	29.41
100	GREG	0.24	2291.49	2291.49	0.09	-
	M	-93.35	189.63	8903.24	-38.10	-74.26
	CM	-21.88	1265.13	1743.90	-8.93	31.40
	LM	-25.06	1358.22	1986.06	-10.23	15.38
	W1	-16.40	2167.53	2436.41	-6.69	-5.95
	W2	-29.72	998.30	1881.41	-12.13	21.80
	W075	-8.08	2180.62	2245.93	-3.30	2.03
	CHAM	-40.87	718.47	2388.55	-16.68	-4.06
120	GREG	-0.14	1692.79	1692.79	-0.06	-
	M	-93.60	140.09	8900.68	-38.21	-80.98
	CM	-18.38	1043.40	1381.05	-7.50	22.57
	LM	-21.06	1140.59	1583.97	-8.59	6.37
	W1	-12.64	1751.06	1910.88	-5.16	-11.41
	W2	-25.42	892.39	1538.52	-10.38	10.03
	W075	-6.39	1697.34	1738.16	-2.61	-2.61
	CHAM	-36.71	615.14	1962.45	-14.98	-13.74

Table 2.4. Percent Relative Error of the MSE approximations

n	Estimators			
	GREG	M	CM	LM
20	1.81	2.24	-17.43	-6.32
40	0.91	1.66	-9.80	-11.58
60	0.80	1.19	-7.86	-16.64
80	2.90	0.47	-7.24	-14.45
100	-0.57	-0.37	-5.81	2.09
120	-0.93	0.65	-5.25	-0.21

Table 2.5. Percent Relative Bias of the MSE Estimators

n	Estimators			
	GREG	M	CM	LM
20	-1.01	-8.95	-11.11	-3.96
40	0.37	-5.27	-12.23	-2.18
60	-0.48	-3.83	-11.62	-1.83
80	1.02	-4.85	-12.14	-2.76
100	0.69	-0.56	-6.84	0.68
120	-0.36	-0.18	-6.29	-0.07

Table 2.6. Choice of $K_n(\alpha)$ for the CM-estimator

	$n = 20$			$n = 60$			$n = 100$		
	$\alpha = 2$	$\alpha = 3$	$\alpha = 4$	$\alpha = 2$	$\alpha = 3$	$\alpha = 4$	$\alpha = 2$	$\alpha = 3$	$\alpha = 4$
$B(\widehat{Y})^{(*)}$	-18.18	-23.15	-24.73	-13.00	-17.88	-19.48	-8.93	-13.23	-14.63
$V(\widehat{Y})$	5007.99	3485.92	3140.32	1933.26	1383.57	1266.29	1265.13	953.99	888.75
$MSE(\widehat{Y})$	6991.85	6703.30	6812.00	2946.99	3302.41	3544.73	1743.90	2004.29	2173.77
$\Delta\xi(\%)$	137.99	148.23	144.27	57.48	40.53	30.92	31.40	14.33	5.42

(*): Bias expressed in percentage of the population mean

CHAPTER 3

Efficiency-Robust Alternatives to the GREG Estimator

3.1. Introduction

In this chapter, the problem to be addressed is still the estimation of the population mean. However, we work this time with a more complex superpopulation model. In fact we assume that p auxiliary variables are related to the variable of interest through a linear relation. This is a natural extension of the theory developed in chapter 2 for the common mean superpopulation model and the resulting model is usually called the GREG model, where GREG stands for **generalized regression** (see Särndal et al., 1992). Some of the results we have proved for the simple case will be adapted to this general framework. However, not all of the proofs will be fully developed again. Straightforward generalizations of the proofs already developed will simply be outlined.

The finite population mean can be estimated under GREG models by using the GREG estimator that is designed to make efficient use of the p auxiliary variables. Our objective is to develop a different type of regression estimator that will keep most of the desirable properties of the GREG estimator but with the advantage of having a smaller mean square error (MSE) for skewed and symmetric long tail populations.

It is important to stress that most of the robust estimators that have been proposed in the literature for skewed populations do not make use of auxiliary information. The development of a general theory of CM- and LM- estimators that is our objective here, is intended to fill this gap. It is however appropriate to mention some attempts towards such generalizations that exist in the literature. Chambers (1982) proposed within the model-based framework an extension of the Winsorization approach to the case where a vector of auxiliary variables is available. The Winsorization parameter depends upon the auxiliary variables. Gwet & Rivest (1992) proposed a robust alternative to the ratio estimator that makes use of the auxiliary variable under the simple random sampling design. No extension to more general models and sampling designs is available. Hulliger (1995) discussed a “robustified” version of the Horvitz-Thompson estimator within the context of the PPS sampling design. The Hulliger estimator can only be used for that specific design and its lack of consistency is a problem in sample surveys.

In sections 3.2 and 3.3 we present the general theory of the robust regression esti-

mators. In section 3.4 we study an important special case of the GREG model, the ratio model, under various sampling designs. Section 3.5 is an extension of section 3.4 to the simple regression model with a nonzero intercept and where only one auxiliary variable is used.

3.2. Review of Existing Procedures

In this section we confine ourselves to the case where the auxiliary information is available at the unit level although this is not needed to implement the proposed robust procedures. Population totals are sufficient to many estimators, including the CM-estimator and the LM-estimator. The availability of the auxiliary information at the unit level is only needed for implementing some complex sampling designs such as the Probability-Proportional-to-Size or the stratified sampling designs.

Let $\mathcal{U}$ be the target population which comprises N units labeled $k = 1, \dots, N$. It is also assumed that a sample s of fixed size n is to be selected according to an arbitrary design p . Let $\mathcal{S}_n$ be the set of all possible samples of size n that can be drawn out of $\mathcal{U}$ according to the design $p(\cdot)$ and π_k the probability of selecting unit k in the sample. The variable and the parameter of interest are respectively given by y and $\bar{Y}$, where

$$\bar{Y} = \sum_{i=1}^N y_i / N.$$

Furthermore, it is assumed that a vector $\mathbf{x}_k = (x_{1k}, \dots, x_{pk})^\top$ exists for each population unit k and observable for $k \in s$. This assumption is needed for sampling purposes only. Two-phase sampling is not needed to collect more information. In

cases where this assumption is not satisfied, a minimum requirement will be that the vector $\bar{\mathbf{X}}$ is known where

$$\bar{\mathbf{X}} = \sum_{i=1}^N \mathbf{x}_i / N.$$

3.2.1. The GREG Estimator

The general form of the GREG model is given by

$$(\xi_0) \begin{cases} y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \text{ for } i \in \mathcal{U} \\ E_\xi(\varepsilon_i) = 0, V_\xi(\varepsilon_i) = \sigma_i^2, \text{ and } E_\xi(\varepsilon_i \varepsilon_j) = 0 \text{ for all } i \neq j. \end{cases}$$

Note that no specific structure is given to the variances σ_i^2 's. However there are two important structures for the variances that are worth noting. These are the cases where $\sigma_i^2 = \boldsymbol{\lambda}^\top \mathbf{x}_i$ or $\sigma_i^2 = \sigma^2$. In fact, with these variance structures, the GREG estimator takes a very simple form (see Särndal et al., 1992).

This model is assumed to provide a reasonable description of the finite population under study. It is then seen as a superpopulation model. The GREG estimator of $\bar{Y}$ suggested by this model is given by

$$\hat{\bar{Y}}_G = \bar{\mathbf{X}}^\top \hat{\mathbf{B}}_G + \sum_{i \in s} \frac{e_{Gi}}{\hat{N} \pi_i},$$

where

$$\begin{cases} \hat{N} = \sum_{i \in s} 1/\pi_i, \\ e_{Gi} = y_i - \mathbf{x}_i^\top \hat{\mathbf{B}}_G, \end{cases}$$

and the vector $\hat{\mathbf{B}}_G$ is solution to the following normal equations:

$$\sum_{i \in s} \frac{1}{\pi_i \sigma_i^2} \mathbf{x}_i (y_i - \mathbf{x}_i^\top \hat{\mathbf{B}}_G) = 0.$$

Under mild regularity conditions, Robinson & Särndal (1983) proved that $\widehat{Y}_G$ is a consistent estimator of $\bar{Y}$. Consistency is the first important property of the GREG estimator. The GREG estimator discussed extensively by Särndal et al. (1992): Cassel, Särndal and Wretman (1976); Wright (1983) Isaki and Fuller (1982).

Moreover, the GREG estimator can be rewritten as

$$\widehat{Y}_G = \sum_{i \in s} \frac{g_i}{\widehat{N}\pi_i} y_i,$$

where

$$g_i = 1 + (\bar{\mathbf{X}} - \widehat{\mathbf{X}})^\top \left(\sum_{i \in s} \frac{\mathbf{x}_i \mathbf{x}_i^\top}{\widehat{N}\pi_i \sigma_i^2} \right)^{-1} \frac{\mathbf{x}_i}{\sigma_i^2},$$

and $\widehat{\mathbf{X}}$ is given by

$$\widehat{\mathbf{X}} = \sum_{i \in s} \frac{1}{\widehat{N}\pi_i} \mathbf{x}_i.$$

This shows that the GREG estimator is of the form of a weighted mean, where the total weight of y_i is the product of the sampling weight $1/\widehat{N}\pi_i$ and the regression weight g_i .

A second important property of the GREG estimator is that the estimated population means of the auxiliary variables equal the corresponding population means.

Formally, we have

$$\sum_{i \in s} \frac{g_i}{\widehat{N}\pi_i} \mathbf{x}_i = \bar{\mathbf{X}}. \quad (3.1)$$

It can be shown that the set of weights $\{w_i = (\widehat{N}\pi_i)^{-1}g_i, i \dots, N\}$ attached to the y_i 's is the nearest possible, according to some distance measure, to the original

set of sampling weights that satisfies the constraint (3.1) (see Deville & Särndal, 1992). For that reason, the w_i 's are called the calibrated weights and the GREG estimator, a calibration estimator. In developing the robust estimators we like to preserve these two important properties of the GREG estimator. In fact, it will be shown that the robust estimators proposed here, which are extensions of the M-, CM- and LM- estimators developed previously, are also calibration estimators. They are consistent and the main difference between the GREG estimator and the robust estimators is that the bias of the latter will converge to 0 more slowly than that of the former.

3.2.2. The Gwet-Rivest M-Estimator

The Gwet-Rivest M-estimator was originally developed under the assumption of a simple random sampling design and a ratio superpopulation model given by

$$(\xi_R) \begin{cases} y_i = \beta x_i + \epsilon_i, \\ E_{\xi_R}(\epsilon_i) = 0 \text{ and } V_{\xi_R}(\epsilon_i) = \sigma^2 x_i. \end{cases}$$

Note that, $\bar{Y} = \bar{X}(\bar{Y}/\bar{X})$, where the ratio $\bar{Y}/\bar{X}$ to be estimated is the weighted least-squares estimator of β under the ratio model if the finite population is considered as a simple random sample from the superpopulation. Therefore, if a simple random sample of size n is selected from the finite population then $\bar{Y}$ can be estimated by the ratio estimator $\hat{\bar{Y}}_R = \bar{X}(\bar{y}/\bar{x})$.

In order to motivate the M-estimator, we assume in this section that the ϵ_i 's in the ratio model are symmetrically distributed. It follows from the ratio model that

$\bar{Y} \doteq \bar{X}\beta$ as $\sum_{i=1}^N \epsilon_i/N \doteq 0$ under the law of large numbers. Moreover, if the finite population is seen as a simple random sample of size N from the superpopulation, then a consistent estimator of β under the ratio model is given by B_M implicitly defined by the following equation,

$$\sum_{i=1}^N \sqrt{x_i} \psi_H \left(\frac{y_i - B_M x_i}{\sigma \sqrt{x_i}} \right) = 0,$$

where ψ_H is defined for all real numbers r by,

$$\psi_H(r) = \text{sgn}(r) \min(K, |r|),$$

K being the robustness constant. Therefore we have that $\bar{Y} \doteq \bar{X}B_M$. From a simple random sample of size n from the finite population, $\bar{Y}$ can be estimated by the Gwet-Rivest M-estimator given by $\hat{\bar{Y}}_M = \bar{X} \hat{B}_M$ where $\hat{B}_M$ is defined by the equation

$$\sum_{i=1}^n \sqrt{x_i} \psi_H \left(\frac{y_i - \hat{B}_M x_i}{\sigma \sqrt{x_i}} \right) = 0.$$

In practice however, σ is unknown and has to be estimated. In this case, it is appropriate to start by defining the population parameters (S_M, B_M) as solution to the following system of equations:

$$\begin{cases} \sum_{i=1}^N \sqrt{x_i} \psi_H \left(\frac{y_i - B_M x_i}{S_M \sqrt{x_i}} \right) = 0, \\ S_M = C_0 \text{ med}_{1 \leq i \leq N} |(y_i - B_M x_i)/\sqrt{x_i}|, \end{cases}$$

where C_0 is a constant that is usually set so that S_M is consistent for estimating the scale parameter σ . Under normality, $C_0 = 1/\Phi^{-1}(3/4)$, Φ being the cumulative

distribution function of the normal distribution. The choice of C_0 is not that critical and might simply result in some loss in efficiency.

Figure 3.1: Contaminated Normal Distributions for Various Contamination Rates ϵ .

The M-estimator of $\bar{Y}$ is still defined as previously, except that $\hat{B}_M$ must be defined simultaneously with an estimator s_M of S_M . These two estimators are implicitly defined as solution to the following system of equations:

$$\begin{cases} \sum_{i=1}^n \sqrt{x_i} \psi_H \left(\frac{y_i - \hat{B}_M x_i}{s_M \sqrt{x_i}} \right) = 0, \\ s_M = C_0 \text{ med}_{1 \leq i \leq n} |(y_i - \hat{B}_M x_i) / \sqrt{x_i}| \end{cases}$$

It is clear that the M-estimator will work very well if the model (ξ_R) is valid or approximately valid and the errors ϵ_i 's follow a symmetric distribution with a tail heavier than that of the normal distribution. Such a distribution might be double exponential or a distribution in the class of the contaminated normal distributions.

Figure 3.1 gives an example of contaminated normal distributions.

The extension of the M-estimator to multiple regression models is possible under certain conditions. The simplest case where the extension is straightforward is when $\sigma_i = \sigma$, $\forall i = 1, \dots, N$. Let us assume the following GREG model:

$$(\xi_0^*) \begin{cases} y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \\ E_{\xi_0^*}(\epsilon_i) = 0, \quad E_{\xi_0^*}(\epsilon_i \epsilon_j) = 0 \text{ for } i \neq j, \quad V_{\xi_0^*}(\epsilon_i) = \sigma^2, \end{cases}$$

Suppose that σ is known. Then under the assumption of symmetry and under the model, we have that $E_{\xi_0^*} \psi_H(\epsilon_i) = 0$ for all $i = 1, \dots, N$ and

$$\bar{Y} \doteq \bar{\mathbf{X}}^\top \boldsymbol{\beta} \doteq \bar{\mathbf{X}}^\top \mathbf{B}_M,$$

where $\mathbf{B}_M$ is defined by the following estimating equation,

$$\sum_{i=1}^N d_i \psi_H \left((y_i - \mathbf{x}_i^\top \mathbf{B}_M) / (d_i \sigma) \right) \mathbf{x}_i = 0, \quad (3.2)$$

$d_i = d(\mathbf{x}_i)$ being a function of leverage to be specified. This follows from the law of large numbers and the ξ_0^* -consistency of $\mathbf{B}_M$ as an estimator of $\boldsymbol{\beta}$.

An M-estimator of $\bar{Y}$ is given by $\widehat{\bar{Y}}_M = \mathbf{X}^\top \widehat{\mathbf{B}}_M$ where $\widehat{\mathbf{B}}_M$ is given by,

$$\sum_{i=1}^n d_i \psi_H \left((y_i - \mathbf{x}_i^\top \widehat{\mathbf{B}}_M) / (d_i \sigma) \right) \mathbf{x}_i = 0.$$

It should be stressed that this estimator can work only if $\mathbf{B}_M$ is ξ_0^* -consistent for estimating $\boldsymbol{\beta}$. Note also that the extension to the important GREG models where $\sigma_i^2 = \boldsymbol{\lambda}^\top \mathbf{x}_i$ is not known. The problem stems from the fact that a general standard procedure for estimating the parameter $\boldsymbol{\lambda}$ is not available.

3.3. M-Estimation Under the Assumption of Symmetry: Extension to Complex Designs

The Gwet-Rivest M-estimator was presented in the previous section under the simple random sampling design. The standard ratio model and the general regression model with a constant error variance were studied. It was seen that the M-estimator is essentially a model-based estimator that can work well only under certain assumptions on the superpopulation model. In some practical situations, these assumptions are approximately satisfied, justifying the use of the M-estimator.

In this section, the objective is to study the asymptotic properties of M-estimators for complex sampling designs and under the assumption of symmetry, as well as their calibration properties. Throughout this section, we consider the superpopulation model (ξ_0^*) with the assumption that $E_{\xi_0^*} \psi_H(\epsilon_i) = 0$. This amounts to assuming symmetry. As a finite population is seldom symmetric, it is more realistic to put the symmetry hypothesis on the superpopulation which is infinite.

3.3.1. The M-Estimator

If the model (ξ_0^*) is valid, then

$$\bar{Y} = \bar{\mathbf{X}}^\top \boldsymbol{\beta} + \mathcal{O}_p(1/\sqrt{N}), \quad (3.3)$$

where $\mathcal{O}_p(1/\sqrt{N})$ indicates that the stochastic order is taken relative to the model.

Moreover, it follows from standard results in robust estimation that

$$\mathbf{B}_M = \boldsymbol{\beta} + \mathcal{O}_p(1/\sqrt{N}).$$

where $\mathbf{B}_M$ is given by (3.2). Therefore, (3.3) implies that

$$\bar{Y} = \bar{\mathbf{X}}\mathbf{B}_M + \mathcal{O}_p(1/\sqrt{N}). \quad (3.4)$$

Equation (3.4) indicates that if the population is reasonably large, one can assume that $\bar{Y} \doteq \bar{\mathbf{X}}^\top \mathbf{B}_M$.

Suppose that a sample of population units is selected according to an arbitrary design and that the probability for selecting population unit i (the inclusion probability) is π_i . It is assumed that $\pi_i > 0 \forall i = 1, \dots, N$. Then, the M-estimator of $\bar{Y}$ is given by:

$$\hat{\bar{Y}}_M = \bar{\mathbf{X}}^\top \hat{\mathbf{B}}_M \text{ where } \sum_{i=1}^n \frac{d_i \mathbf{x}_i}{\pi_i} \psi_H \left(\frac{y_i - \mathbf{x}_i^\top \hat{\mathbf{B}}_M}{d_i \sigma} \right) = 0. \quad (3.5)$$

This is in fact a Generalized M(GM)-estimator of the Scweppe's form designed to dampen the influence of high leverage points via the $d_i = d(\mathbf{x}_i)$ function. We propose the following function:

$$d_i = \min \left(1, \frac{1 + \chi_{p-\lambda, 0.95}^2}{NH_i} \right) \text{ where } \begin{cases} H_i = \mathbf{x}_i^\top (\mathbf{X}_P^\top \mathbf{X}_P)^{-1} \mathbf{x}_i, \\ \mathbf{X}_P = \underset{1 \leq i \leq N}{\text{col}} \mathbf{x}_i^\top, \end{cases}$$

$\lambda = 1$ if the regression model has an intercept and $\lambda = 0$ otherwise. Note that for large values of N , $NH_i - 1$ is nothing but the squared Mahalanobis distance of the i -th population unit given by

$$MD_i^2 = (\mathbf{x}_i - \bar{\mathbf{X}})^\top \mathbf{C}^{-1} (\mathbf{x}_i - \bar{\mathbf{X}}),$$

Figure 3.2: Influence Function of the M-estimator.

where,

$$\mathbf{C} = \frac{1}{N-1} \sum_{k=1}^N (\mathbf{x}_k - \bar{\mathbf{X}})(\mathbf{x}_k - \bar{\mathbf{X}})^T.$$

That distance is usually compared with quantiles of the chi-squared distribution with $p - \lambda$ degrees of freedom. These d-factors were unnecessary for the common mean model discussed previously but are essential when auxiliary variables are used. For standard M-estimators, $d_i = 1, \forall i = 1, \dots, N$. To see the importance of the d-factors, note that the asymptotic variance V_0 of the M, the GM or the GREG estimator is of the form,

$$V_0 = V \left(\sum_{i=1}^n \frac{IF(\mathbf{x}_i, y_i)}{N\pi_i} \right).$$

The form of the influence function $IF()$ of the M- and the GM-estimator will be

specified in the next section. In figure 3.2 we made a plot of the influence function $IF(\cdot)$ of the M-estimator ($d_i = 1$) under the ratio model (ξ_R) mentioned earlier. Figure 3.2a indicates that for small values of x influence is fairly limited. It can also be seen from figure 3.2b that for any fixed value of x , the influence of (x, y) is bounded. However, when y is fixed, both graphs show that the influence of (x, y) is unbounded as x increases. That influence can even grow dramatically fast. As indicated in the legends, shaded areas indicate areas of high influence. The darker the area the higher the influence of the points in that area will be.

Figure 3.3: Influence Function of the GM-estimator.

The GM-estimator defined by (3.5) was designed to solve that problem. The figures 3.3a and 3.3b show that, in fact, the overall influence of (x, y) is bounded in the

case of GM-estimators.

For very large values of x , the influence of (x, y) has a tendency to decrease towards 0 beyond a certain upper bound. This shows clearly that the GM-estimator is to be used whenever high leverage points are expected.

The M-estimator $\widehat{Y}_M$ is a calibration estimator. To see this, suppose that the variable of interest is now x_1 . It follows from (3.5) that vector $\widehat{\mathbf{B}}_M = (1, 0, \dots, 0)^\top$ is a solution to the estimating equation. Since the function ψ_H is increasing, that solution is unique, which implies that $\widehat{Y}_M = \overline{X}_1$ where $\overline{X}_1$ is given by

$$\overline{X}_1 = \frac{1}{N} \sum_{k=1}^N x_{1k}.$$

Another problem of practical interest that will be addressed is that of a multipurpose survey dealing with a variety of characteristics. Producing estimates for all characteristics through M-estimation might be a time-consuming process. Moreover, a variety of other practical issues are of concern as well. Chambers and Kokic (1993) provide an account of many such problems. Among other issues these authors mentioned the consistency problem that arises when estimates of the same characteristic produced at a low level of aggregation do not add up to the estimate produced at a higher level.

For multipurpose surveys, we assume that there exists a key variable of interest called z . Robust M-weights will be produced for that variable and will be applied to the remaining variables for which estimates are required.

Let $u_{Mi} = z_i - \mathbf{x}_i^\top \hat{\mathbf{B}}_M$, $\bar{u}_{Mi} = u_{Mi}/d_i\sigma$ and α_i given by:

$$\alpha_i = \begin{cases} 1, & \text{if } u_{Mi} = 0, \\ \psi_H(\bar{u}_{Mi})/\bar{u}_{Mi}, & \text{otherwise.} \end{cases}$$

It follows from (3.5) that the M-estimator of the key variable of interest can be written as:

$$\hat{\bar{Z}}_M = \sum_{i=1}^n w_{Mi} z_i,$$

where,

$$w_{Mi} = \frac{g_{Mi}}{\hat{N}\pi_i} \text{ and } g_{Mi} = \bar{\mathbf{X}}^\top \left(\sum_{k=1}^n \frac{\alpha_k}{\hat{N}\pi_k} \mathbf{x}_k \mathbf{x}_k^\top \right)^{-1} \mathbf{x}_i \alpha_i.$$

The robust weight w_{Mi} is made up of three components: (i) the original sampling weight $1/\pi_i$, (ii) the robust regression weight g_{Mi} and (iii) the estimated population size $\hat{N}$. In practice, these robust weights will be calculated from a single key variable and will be applied as such to the other variables.

The M-estimator of an arbitrary variable y is by definition given by

$$\hat{\bar{Y}}_M = \sum_{i=1}^n w_{Mi} y_i.$$

This estimator has the advantage of solving the consistency problem mentioned earlier. However, it will only be efficient on characteristics that are correlated with the key variable of interest.

The M-estimator defined above depends on the scale parameter σ which is usually unknown in practice. It must be estimated from the data. This can be done along the lines of the technique developed in the univariate case by estimating

simultaneously the scale parameter and the parameter $\mathbf{B}_M$. Let the parameters $(\mathbf{B}_M, S_M)$ be defined by

$$\sum_{k=1}^N d_k \mathbf{x}_k \psi_H \left(\frac{y_k - \mathbf{x}_k^\top \mathbf{B}_M}{d_k S_M} \right) = 0, \quad S_M = C_0 \text{med}_{1 \leq i \leq N} |y_i - \mathbf{x}_i^\top \mathbf{B}_M|.$$

The M-estimator of $\bar{Y}$ is then given by $\hat{\bar{Y}}_M = \bar{\mathbf{X}}^\top \hat{\mathbf{B}}_M$ where $\hat{\mathbf{B}}_M$ is given by

$$\sum_{i=1}^n (1/\pi_i) \mathbf{x}_i d_i \psi_H \left(\frac{y_i - \mathbf{x}_i^\top \hat{\mathbf{B}}_M}{d_i \hat{S}_M} \right) = 0, \quad \hat{S}_M = C_0 \widehat{\text{med}}_{1 \leq i \leq N} |y_i - \mathbf{x}_i^\top \hat{\mathbf{B}}_M|.$$

In chapter 2, a method for estimating medians under complex designs has been discussed and can be used here.

However Rivest (1993) pointed out that the above equations may have more than one solution. This is somewhat troublesome. To solve this problem, S_M and $\hat{S}_M$ can be replaced with S_{L1} and $\hat{S}_{L1}$ respectively which are defined as follows:

$$S_{L1} = C_0 \text{med}_{1 \leq i \leq N} |y_i - \mathbf{x}_i^\top \mathbf{B}_{L1}| \quad \text{and} \quad \hat{S}_{L1} = C_0 \widehat{\text{med}}_{1 \leq i \leq N} |y_i - \mathbf{x}_i^\top \hat{\mathbf{B}}_{L1}|$$

where $\mathbf{B}_{L1}$ and $\hat{\mathbf{B}}_{L1}$ are implicitly defined by

$$\mathbf{B}_{L1} = \arg \min_{\mathbf{B}} \sum_{i=1}^N |y_i - \mathbf{x}_i^\top \mathbf{B}| \quad \text{and} \quad \hat{\mathbf{B}}_{L1} = \arg \min_{\mathbf{B}} \sum_{i=1}^n |y_i - \mathbf{x}_i^\top \mathbf{B}| / \pi_i.$$

3.3.2. Large-Sample Properties of the M-Estimator

In this section, the conditions under which the M-estimator is consistent for estimating $\bar{Y}$ are given. The asymptotic variance is derived through a linear representation of $\hat{\mathbf{B}}_M$ as in chapter 2. An estimator of variance is also obtained.

The M-estimator is typically a model-based estimator that is used within the design-based framework. Its consistency properties must therefore be studied with respect to the design and the model simultaneously. The bias of an estimator $\widehat{Y}$ will be given by

$$B(\widehat{Y}) = E_{\xi_0} E_d(\widehat{Y} - \overline{Y}),$$

and the variance of an estimator,

$$V(\widehat{Y}) = E_{\xi_0} E_d \left[\widehat{Y} - E_{\xi_0} E_d(\widehat{Y}) \right]^2.$$

In practice, however, only the design-based variance will be used.

Again, we assume a nested sequence of finite populations $\{\mathcal{U}\}_{t=1}^{+\infty}$. For a given $t \geq 1$, the population $\mathcal{U}_t$ is of size N_t and the sample drawn from that population is of size n_t . Moreover, we also have that

$$\lim_{t \rightarrow \infty} N_t = \lim_{t \rightarrow \infty} n_t = \infty.$$

If the scale parameter σ must be estimated by $\widehat{S}_M$, we will also assume that

$$\widehat{S}_M = S_M + o_p^{(d)}(1) \text{ and } S_M = \sigma + o_p^{(\xi_0^*)}(1),$$

where $o_p^{(d)}(\cdot)$ refers to the design and $o_p^{(\xi_0^*)}(\cdot)$ refers to the model. Under this assumption, $\widehat{\mathbf{B}}_M(\widehat{S}_M)$ and $\widehat{\mathbf{B}}(\sigma)$ have the same asymptotic distribution by Slutsky's theorem. It is then possible to assume without loss of generality that the scale parameter σ is known.

Since the population is allowed to grow indefinitely, it is more convenient to assume that the $\mathbf{x}_i$'s are a sequence of independent and identically distributed random vectors. Let $K(\mathbf{x})$ be the distribution function of the $\mathbf{x}_i$'s and $G(r)$ the distribution function of ϵ_i/σ 's. The joint density function of the $(\mathbf{x}_i, y_i)$'s is given by

$$\begin{aligned} f_{\xi_0^*}(\mathbf{x}, y) &= (1/\sigma)g[(y - \mathbf{x}^\top \boldsymbol{\beta})/\sigma]k(\mathbf{x}), \\ &= (1/\sigma)g(r)k(\mathbf{x}) \end{aligned}$$

where $r = (y - \mathbf{x}^\top \boldsymbol{\beta})/\sigma$ and $F_{\xi_0^*}$ the corresponding distribution function.

Let us introduce the following functions:

$$g_0(\mathbf{b}) = \int \psi_H\left(\frac{y - \mathbf{x}^\top \mathbf{b}}{d(\mathbf{x})\sigma}\right) \mathbf{x} d(\mathbf{x}) dF_{\xi_0^*}(\mathbf{x}, y),$$

where

$$d(\mathbf{x}) = \min\left(1, \frac{1 + \chi_{p-\lambda, 0.95}}{h_0(\mathbf{x})}\right) \text{ and } h_0(\mathbf{x}) = \mathbf{x}^\top \left(\int \mathbf{x}\mathbf{x}^\top dK(\mathbf{x})\right)^{-1} \mathbf{x}$$

and

$$g_t(\mathbf{b}) = \sum_{i=1}^{N_t} \psi_H\left(\frac{y_i - \mathbf{x}_i^\top \mathbf{b}}{d_i\sigma}\right) d_i \mathbf{x}_i / N \text{ and } \hat{g}_t(\mathbf{b}) = \sum_{i=1}^{n_t} \psi_H\left(\frac{y_i - \mathbf{x}_i^\top \mathbf{b}}{d_i\sigma}\right) \frac{d_i \mathbf{x}_i}{N\pi_i}.$$

We will need the following regularity conditions already used in chapter 2:

$$\bullet N_t \min_{1 \leq i \leq N_t} \pi_i \rightarrow \infty \text{ as } t \rightarrow \infty, \quad (3.6)$$

$$\bullet \max_{1 \leq i \neq j \leq N_t} \left| \frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right| \rightarrow 0 \text{ as } t \rightarrow \infty. \quad (3.7)$$

Definition 3.1. An estimator $\hat{\bar{Y}}_t$ of $\bar{Y}_t$ is consistent if

$$\lim_{t \rightarrow \infty} P\left(|\hat{\bar{Y}}_t - \bar{Y}_t| > \epsilon\right) = 0, \quad \forall \epsilon > 0.$$

Note that the probability is taken here with respect to the design as well as the model. That is,

$$P\left(|\widehat{\bar{Y}}_t - \bar{Y}_t| > \epsilon\right) = \int P_p\left(|\widehat{\bar{Y}}_t - \bar{Y}_t| > \epsilon \left| \begin{matrix} \mathbf{x}_1, \dots, \mathbf{x}_{N_t}, \\ y_1, \dots, y_{N_t} \end{matrix} \right.\right) \prod_{i=1}^{N_t} dF_{\xi_0^*}(\mathbf{x}_i, y_i), \quad (3.8)$$

where $P_p(\cdot)$ is the probability with respect to the design.

The next lemma gives the conditions under which $\widehat{\mathbf{B}}_{Mt}$ is a consistent estimator of $\boldsymbol{\beta}$ according to definition 3.1 .

Lemma 3.1. *Suppose that the conditions (C_π) are satisfied as well as the following two conditions,*

- (i) $(1/N_t) \sum_{i=1}^{N_t} \|\mathbf{x}_i\|^4 = O(1).$
- (ii) $0 < \sup_{\|\mathbf{t}\|=1} \int (\mathbf{x}^\top \mathbf{t})^2 dK(\mathbf{x}) < +\infty.$

Then, $\widehat{\mathbf{B}}_{Mt} - \boldsymbol{\beta} = O_p(n_t^{-1/2}).$

Note that $O_p(\cdot)$ refers to the stochastic order with respect to the model and the design, while $O_p^{(d)}(\cdot)$ indicates that only the design is involved and $\mathcal{O}_p(\cdot)$ refers to the model only.

The proof of lemma 3.1 can be obtained by a straightforward adaptation of the proof of theorem 3.3. This is done by using the functions $\mathbf{g}_0(\cdot)$, $\mathbf{g}_t(\cdot)$ and $\widehat{\mathbf{g}}_t(\cdot)$ and a version of the fixed point theorem (see the proof of theorem 3.3 in appendix for more details). This lemma is now used to prove the consistency of the M-estimator $\widehat{\bar{Y}}_{Mt}$ as an estimator of $\bar{Y}_t$.

Theorem 3.1. *Suppose that the conditions of the lemma 3.1 are all satisfied.*

Then $\widehat{Y}_{\text{Mt}} - \bar{Y}_t = O_p(n_t^{-1/2})$.

Proof.

We have that,

$$\bar{Y}_t = \bar{\mathbf{X}}_t^\top \boldsymbol{\beta} + \bar{\epsilon}_t, \text{ and } \widehat{Y}_{\text{Mt}} = \bar{\mathbf{X}}_t^\top \widehat{\mathbf{B}}_{\text{Mt}}.$$

Therefore, for all $\eta > 0$,

$$\begin{aligned} P\left(|\widehat{Y}_{\text{Mt}} - \bar{Y}_t| > \eta\right) &= P\left(|\bar{\mathbf{X}}_t^\top (\widehat{\mathbf{B}}_{\text{Mt}} - \boldsymbol{\beta}) - \bar{\epsilon}_t| > \eta\right), \\ &\leq P\left(|\bar{\mathbf{X}}_t^\top (\widehat{\mathbf{B}}_{\text{Mt}} - \boldsymbol{\beta})| + |\bar{\epsilon}_t| > \eta\right), \\ &\leq P\left(\|\bar{\mathbf{X}}_t^\top\| \|\widehat{\mathbf{B}}_{\text{Mt}} - \boldsymbol{\beta}\| > \eta/2\right) + P\left(|\bar{\epsilon}_t| > \eta/2\right). \end{aligned}$$

It follows from assumption (i) of lemma 3.1 that $\bar{\mathbf{X}}_t = O(1)$. This notation simply means that there exist t_0 and a constant $C_0 > 0$ such that if $t > t_0$ then $\mathcal{P}(\bar{\mathbf{X}}_t \leq C_0) = 1$, where $\mathcal{P}$ stands for the probability with respect to the model. It is possible to extend the theorem to the case where $\bar{\mathbf{X}}_t = \mathcal{O}_p(1)$. Now we have that

$$P\left(|\widehat{Y}_{\text{Mt}} - \bar{Y}_t| > \eta\right) \leq P\left(\|\widehat{\mathbf{B}}_{\text{Mt}} - \boldsymbol{\beta}\| > \eta/(2C_0)\right) + P\left(|\bar{\epsilon}_t| > \eta/2\right), \text{ if } t > t_0.$$

The result follows from lemma 3.1 and the fact that

$$P\left(|\bar{\epsilon}_t| > \eta/2\right) \leq 4 \frac{E(\bar{\epsilon}_t^2)}{\eta^2} = 4 \frac{\sigma^2}{\eta^2 N_t}.$$

■

The next theorem provides a design-based linear expansion of the M-estimator $\widehat{Y}_{Mt}$ and the conditions that ensure its validity. That representation will then be used to derive the asymptotic variance of the M-estimator.

Theorem 3.2. *Suppose that all the conditions of lemma 3.1 are satisfied as well as the following two conditions:*

- (i) *There exists t_0 such that if $t > t_0$ then the second- order derivative $\widehat{g}_t''$ of g_t exists in the neighborhood of $\mathbf{B}_{Mt}$ with probability 1.*
- (ii) *$|\psi'_H| \leq 1$ and $|\psi''_H| \leq 1$. Then the M-estimator has the following linear representation,*

$$\widehat{Y}_{Mt} = \overline{\mathbf{X}}^\top \mathbf{B}_{Mt} + \sum_{k=1}^{n_t} \frac{d_k \mathbf{x}_k^\top \mathbf{Q}^{-1} \overline{\mathbf{X}}}{N_t \pi_k} [S_{Mt} \psi_H(\overline{E}_{Mi})] + O_p^{(d)}(n_t^{-1}),$$

where $\mathbf{Q} = \frac{1}{N_t} \sum_{i=1}^{N_t} \psi'_H(\overline{E}_{Mi}) \mathbf{x}_i \mathbf{x}_i^\top$ and $\overline{E}_{Mi} = (y_i - \mathbf{x}_i^\top \mathbf{B}_{Mt}) / S_{Mt}$.

In this theorem, it is implicitly assumed that the matrix $\mathbf{Q}_t$ is non-singular for all t . This will be true very often in practice. The proof of this theorem is a straightforward adaptation of the proof of theorem 3.5 and is therefore omitted. Theorem 3.5 provides a similar result for a more general M-estimator and its proof features techniques that can be used to prove theorem 3.2 .

Theorem 3.2 is now used to derive the design-based asymptotic variance of the M-estimator. It indicates that the M-estimator has the same asymptotic distribution

as the quantity $\bar{Y}_M^{(0)}$ given by

$$\bar{Y}_M^{(0)} = \bar{\mathbf{X}}^\top \mathbf{B}_M + \sum_{k=1}^n \frac{d_k \mathbf{x}_k^\top \mathbf{Q}^{-1} \bar{\mathbf{X}}}{N \pi_k} E_{MMi}$$

where

$$E_{MMi} = S_M \psi_H(\bar{E}_{Mi}).$$

Therefore, the asymptotic variance of the M-estimator is given by, $AV(\hat{\bar{Y}}_M) = V(\bar{Y}_M^{(0)})$. Thus,

$$AV(\hat{\bar{Y}}_M) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) (g_i E_{MMi})(g_j E_{MMj}),$$

where $g_k = d_k \mathbf{x}_k^\top \mathbf{Q}^{-1} \bar{\mathbf{X}}$.

A consistent estimator of this asymptotic variance, is given by

$$\widehat{AV}(\hat{\bar{Y}}_M) = \frac{1}{\hat{N}^2} \sum_{i=1}^n \sum_{j=1}^n \frac{1}{\pi_{ij}} \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) (\hat{g}_i e_{MMi})(\hat{g}_j e_{MMj}),$$

where $e_{MMi} = s_M \psi_M(\bar{e}_{Mi})$, $\bar{e}_{Mi} = (y_i - \mathbf{x}_i^\top \hat{\mathbf{B}}_M)/s_M$ and $\hat{g}_i = d_i \mathbf{x}_i^\top \hat{\mathbf{Q}}^{-1} \bar{\mathbf{X}}$.

The gain in efficiency of the M-estimator over the GREG estimator will in general stem from the fact that the variability of the M-residuals E_{MMi} 's is often smaller than that of the GREG-residuals.

The M-estimator can work well for estimating $\bar{Y}$ only if the model is valid and the assumption of symmetry approximately satisfied. This is a major drawback that weakens the robustness of the M-estimator substantially. In the next section, we propose an alternative to the M-estimator that yields consistency regardless of the

validity of the model. It is called the CM-estimator. The price to be paid will necessarily be some loss of robustness under the superpopulation model.

3.4. Robust Estimation for Skewed Populations

In this section we introduce the CM-estimator and the LM-estimator within the generalized regression model framework. The two estimators will simply be introduced here, their properties being relegated to section 3.5.

3.4.1. The CM-Estimator

The M-estimator was introduced in the last section as a way to obtain stable estimates when the superpopulation is symmetric. It was also seen that for multipurpose surveys with a variety of characteristics of interest, robust weights w_{Mi} 's can be derived from the key variable of interest and be applied to the others. For any sample unit i the robust weight is given by,

$$w_{Mi} = g_{Mi}/(\hat{N}\pi_i) \text{ where } g_{Mi} = \bar{\mathbf{X}}^\top \left(\sum_{k=1}^n \frac{\alpha_k}{\hat{N}\pi_k} \mathbf{x}_k \mathbf{x}_k^\top \right)^{-1} \mathbf{x}_i \alpha_i.$$

As mentioned previously, if the superpopulation does not bear any symmetric structure, the M-estimator will lose its consistency property. This is a major drawback as the symmetry assumption is very restrictive.

It is important to mention at this stage that, if all the α_i 's were equal to 1, the M-estimator would be identical to the GREG estimator. Because these α -factors are not all equal to 1, the M-estimator is inconsistent although robust. To achieve

consistency, one option is to modify these factors so that the new factors $\alpha_{\bullet i}$'s be sample-size dependent and satisfy the condition

$$\text{plim}_{n \rightarrow \infty} \max_{1 \leq i \leq n} \alpha_{\bullet i} = 1.$$

Let $\alpha_{\bullet i}$ be given by:

$$\alpha_{\bullet i} = \begin{cases} 1, & \text{if } e_{\text{CM}i} = 0, \\ \psi_{\bullet H}(e_{\text{CM}i}^*)/e_{\text{CM}i}^*, & \text{otherwise,} \end{cases}$$

where $\psi_{\bullet H}$ is the influence function defined for the CM-estimator under the common mean superpopulation model, $e_{\text{CM}i} = y_i - \mathbf{x}_i^\top \hat{\mathbf{B}}_{\text{CM}}$, the statistic $\hat{\mathbf{B}}_{\text{CM}}$ being solution to the following equation

$$\sum_{i=1}^n \frac{d_i}{\pi_i \sigma_i} \mathbf{x}_i \psi_{\bullet H} \left(\frac{y_i - \mathbf{x}_i^\top \hat{\mathbf{B}}_{\text{CM}}}{d_i \sigma_i} \right) = 0. \quad (3.9)$$

We have $e_{\text{CM}i}^* = e_{\text{CM}i}/(d_i \sigma_i)$ if $\sigma_i^2 = \sigma^2 x_i$ and σ is known, or $e_{\text{CM}i}^* = e_{\text{CM}i}/(d_i \hat{S}_{\text{CM}i})$ otherwise. Note that $\alpha_{\bullet i} = \alpha_{\bullet i}(\mathbf{y})$ is a function of the survey variable y observed from the n sample units.

Definition 3.2. Let y be a variable of interest and assume that z is the key variable of the survey. The CM-estimator under the GREG model (ξ_0) of $\bar{Y}$ is given by

$$\begin{aligned} \hat{\bar{Y}}_{\text{CM}} &= \sum_{i=1}^n w_{\text{CM}i} y_i, \text{ where } w_{\text{CM}i} = \frac{g_{\text{CM}i}}{\hat{N} \pi_i}, \text{ and} \\ g_{\text{CM}i} &= \bar{\mathbf{X}}^\top \left(\sum_{k=1}^n \frac{\alpha_{\bullet k}(z)}{\hat{N} \pi_k \sigma_k^2} \mathbf{x}_k \mathbf{x}_k^\top \right)^{-1} \frac{\mathbf{x}_i}{\sigma_i^2} \alpha_{\bullet i}(z). \end{aligned}$$

It will be shown later that this estimator is consistent for estimating the population $\bar{Y}$. It is straightforward to see that the CM-estimator is still a calibration estimator. This follows from the definition of the CM-estimator.

3.4.2. The LM-Estimator

The CM-estimator was developed in section 3.4.1 under a suitable assumption on the superpopulation variance structure. That is, $V_{\xi_0^*}(\epsilon_i) = \boldsymbol{\lambda}^\top \mathbf{x}_i$ for some constant vector $\boldsymbol{\lambda}$. This variance structure leads to a simple form for the GREG estimator. If the hypothesized model possesses a variance structure that is different from the one mentioned above, the CM-estimator might no longer be consistent as it is not asymptotically equivalent to the GREG estimator. The purpose of this section is to propose another robust estimator called LM-estimator that is computationally more demanding than the CM-estimator, but that is always asymptotically equivalent to the GREG estimator, hence consistent.

The LM-estimator is developed here in the context of a multipurpose survey involving several variables of interest. For simplicity, it is assumed that the survey in question has a single key variable of interest and an arbitrary number of other variables with some interest. Let us assume that the key variable of interest is z . The LM-estimator is developed here along the lines of what was previously done for the common mean superpopulation model. The parameter of interest $\bar{Y}$ can be

written as

$$\begin{aligned}\bar{Z} &= \bar{\mathbf{X}}^\top \mathbf{B}_M + (\bar{Z} - \bar{\mathbf{X}}^\top \mathbf{B}_M), \\ &= \bar{\mathbf{X}}^\top \mathbf{B}_M + \sum_{i=1}^N \frac{E_{Mi}}{N},\end{aligned}$$

where

$$E_{Mi} = z_i - \bar{\mathbf{X}}^\top \mathbf{B}_M$$

and $\mathbf{B}_M$ is a population parameter that is implicitly defined by the following equation:

$$\sum_{i=1}^N \frac{d_i}{\sigma_i} \mathbf{x}_i \psi_H \left(\frac{z_i - \mathbf{x}_i^\top \mathbf{B}_M}{d_i \sigma_i} \right) = 0.$$

Note that, unless the errors ε_i 's in model (ξ_0) are symmetrically distributed, $\mathbf{B}_M$ will not be a ξ_0 -consistent estimator of $\boldsymbol{\beta}$ as the population size N goes to infinity. It will indicate however the location of the bulk of the data and the above approach still holds.

For the key variable z , the LM-estimator $\widehat{\bar{Z}}_{LM}$ of $\bar{Z}$ is obtained by replacing $\mathbf{B}_M$ with $\widehat{\mathbf{B}}_M$ and $\bar{E}_M$ with an L-statistic as done under the common mean model. Therefore,

$$\widehat{\bar{Z}}_{LM} = \bar{\mathbf{X}}^\top \widehat{\mathbf{B}}_M + \sum_{i \in s} \frac{e_{CMi}}{\widehat{N}\pi_i}, \text{ where } e_{CMi} = [w_n + (1 - w_n)\hat{J}_{ni}]e_{Mi}, \quad (3.10)$$

the weighting factor w_n as well as the trimming factors $\hat{J}_{ni}$'s are similar to those defined under the common mean superpopulation model. For all $i = 1, \dots, N$.

$$\begin{aligned}\hat{J}_{ni} &= J_n \left[\widehat{F}(e_{Mi}) \right], \\ \text{where } \begin{cases} \forall r \in \mathbb{R}, J_n(r) = \frac{1}{[1 - (\lambda + 1)\alpha_n]} \mathbf{I}(\lambda\alpha_n < r \leq 1 - \alpha_n), \\ \widehat{F}(r) = \sum_{i=1}^n (\widehat{N}\pi_i)^{-1} \mathbf{I}(e_{Mi} \leq r). \end{cases}\end{aligned}$$

Again, $\alpha_n = 1/\hat{n}$ with $\hat{n} = \hat{N} \min_{1 \leq i \leq N} \pi_i$.

In order to define the LM-estimator for any other variable of interest apart from y , we need to express $\hat{\bar{Z}}_{LM}$ as a weighted sum of the z_i 's. It follows from equation (3.10) that the LM-estimator of the key variable Y can be rewritten as,

$$\hat{\bar{Y}}_{LM} = \sum_{i=1}^n w_{LMi} y_i,$$

where

$$w_{LMi} = \frac{g_{LMi}}{\hat{N} \pi_i},$$

and

$$g_{LMi} = \hat{J}_{\bullet ni} + \left(\bar{\mathbf{X}}^\top - \sum_{i=1}^n \frac{\hat{J}_{\bullet ni}}{\hat{N} \pi_i} \mathbf{x}_i^\top \right) \left(\sum_{i=1}^n \frac{\alpha_i}{\hat{N} \pi_i \sigma_i^2} \mathbf{x}_i \mathbf{x}_i^\top \right)^{-1} \frac{\alpha_i}{\sigma_i^2} \mathbf{x}_i,$$

$\hat{J}_{\bullet ni}$ being given by $\hat{J}_{\bullet ni} = w_n + (1 - w_n) \hat{J}_{ni}$.

Definition 3.3. *The LM-estimator under the GREG model (ξ_0) of the population mean of any variable of interest y is given by*

$$\hat{\bar{Y}}_{LM} = \sum_{i=1}^n w_{LMi} y_i,$$

where $w_{LMi} = w_{LMi}(z)$ is obtained from the key variable z .

The question to be asked at this stage is whether this estimator preserves the two important properties of the GREG estimator mentioned earlier. The answer is yes. This LM-estimator is consistent and its consistency will be proved in the next section devoted to the asymptotics. Moreover the LM-estimator is a calibration estimator. This is in fact a direct consequence of definition 3.3 .

There is another way to derive the LM-estimator by adopting an approach similar to that of Deville & Särndal (1992). We start with the Horvitz-Thompson estimator that uses the original sampling weights $(\hat{N}\pi_i)^{-1}$ and we like to derive a new estimator with new weights w_i 's that will satisfy the condition

$$\sum_{i=1}^n w_i \mathbf{x}_i = \bar{\mathbf{X}}. \quad (3.11)$$

Given our goal to have an estimator with good robustness properties, it seems that by staying fairly close to the original sampling weights, there is little hope to have a robust estimator. For, the resulting estimator is likely to be close to the Horvitz-Thompson estimator which is known to be very sensitive to outliers (see Hulliger, 1995). Our approach is to replace the original sampling weights $d_k = 1/\hat{N}\pi_k$ by the modified weights $d_k^* = \hat{J}_{\bullet\pi_i} d_k$. Then the LM weights w_{LMi} 's are obtained by minimizing the distance

$$\sum_{k \in s} \frac{(w_{LMk} - d_k^*)^2}{d_k q_k} \text{ where } q_k = 1/\phi_k,$$

while satisfying the calibration equations (3.11).

It is also appropriate to make a comment about the robustness of this estimator. We anticipate that all variables of interest that are correlated with the key variable use to derive the LM weights will lead to estimators with good robustness properties. Those with a poor correlation with the key variable are likely to yield estimators more sensitive to outlying units. This problem is similar to the efficiency problem encountered with the PPS sampling design.

In the next section we state and prove some asymptotic results of the two robust estimators. Again in the course of our theoretical investigation, we assume for sake of simplicity that the σ_i^2 is proportional to a know parameter σ^2 .

3.5. Asymptotic Properties of the Estimators

The strategy for proving consistency remains fairly close to that used for the common mean superpopulation model. It amounts to showing that the robust estimator is asymptotically equivalent to the GREG estimator which is known to be consistent. Furthermore, the framework for the asymptotics is still that of a sequence of nested finite populations $\{\mathcal{U}\}_{t \geq 1}$. It is assumed that as t goes to ∞ , the corresponding populations grow continuously. The CM-estimator and the LM-estimator are fairly different in nature and cannot be studied simultaneously. The CM-estimator is designed to “robustify” the cosmetic form of the GREG estimator while the LM-estimator will “robustify” the GREG estimator in its general form. We start with the asymptotic properties of the CM-estimator.

3.5.1. Asymptotic Properties of the CM-estimator

Let us introduce some notations. Let $\hat{\mathbf{f}}$ and $\mathbf{f}$ be two functions from $\mathbb{R}^p$ to $\mathbb{R}^p$ defined as follows:

$$\begin{aligned}\hat{\mathbf{f}}(\mathbf{B}) &= \sum_{i=1}^n \frac{d_i}{N\pi_i\sigma_i} \mathbf{x}_i \psi_{\bullet_H} \left(\frac{y_i - \mathbf{x}_i^\top \mathbf{B}}{d_i\sigma_i} \right), \\ \mathbf{f}(\mathbf{B}) &= \sum_{i=1}^N \frac{d_i}{N\sigma_i} \mathbf{x}_i \psi_{\bullet_H} \left(\frac{y_i - \mathbf{x}_i^\top \mathbf{B}}{d_i\sigma_i} \right).\end{aligned}$$

The CM-residuals are denoted by $e_{\text{CM}i} = y_i - \mathbf{x}_i^\top \hat{\mathbf{B}}_{\text{CM}}$ and the theoretical CM-

residuals are given by $E_{\text{CM}i} = y_i - \mathbf{x}_i^\top \mathbf{B}_{\text{CM}}$. These residuals will be sometimes standardized by the factor $d_i \sigma_i$ and the corresponding standardized residuals will be denoted by $e_{\text{CM}i}^{(s)}$ for $e_{\text{CM}i}$ and $E_{\text{CM}i}^{(s)}$ for $E_{\text{CM}i}$. The $\alpha_{\bullet i}$'s have been defined in the beginning of section 3.4. Sometimes it will be more convenient to use the notation $\alpha_{\bullet i}(\mathbf{z})$ that specifies the survey variable used to calculate these factors. Note that $\alpha_{\bullet i}$ estimates the following quantity:

$$A_{\bullet i} = \begin{cases} 1, & \text{if } E_{\text{CM}i} = 0, \\ \psi_{\bullet H}(E_{\text{CM}i}^{(s)})/E_{\text{CM}i}^{(s)}, & \text{otherwise} \end{cases}$$

The following matrices will also be used very often:

$$\begin{aligned} \hat{\mathbf{T}}_{\text{G}} &= \sum_{k=1}^n \frac{\mathbf{x}_k \mathbf{x}_k'}{\hat{N} \pi_k \sigma_k^2}, & \mathbf{T}_{\text{G}} &= \sum_{k=1}^N \frac{\mathbf{x}_k \mathbf{x}_k'}{N \sigma_k^2}, & \hat{\mathbf{T}}_{\bullet}(\mathbf{y}) &= \sum_{k=1}^n \frac{\alpha_{\bullet k}(\mathbf{z})}{\hat{N} \pi_k \sigma_k^2} \mathbf{x}_k \mathbf{x}_k', \\ \tilde{\mathbf{T}}_{\bullet}(\mathbf{y}) &= \sum_{k=1}^n \frac{A_{\bullet k}(\mathbf{z})}{N \pi_k \sigma_k^2} \mathbf{x}_k \mathbf{x}_k', & \mathbf{T}_{\bullet}(\mathbf{y}) &= \sum_{k=1}^N \frac{A_{\bullet k}(\mathbf{z})}{N \sigma_k^2} \mathbf{x}_k \mathbf{x}_k'. \end{aligned}$$

Our first result is about the asymptotic equivalence of the statistics $\hat{\mathbf{B}}_{\text{CM}}$ defined by (3.9) and $\hat{\mathbf{B}}_{\text{G}}$ defined earlier. First, we make the following general assumptions that will be used regularly in the theorems to follow:

$$\text{A1. } \max_{1 \leq k \neq l \leq N_t} \left| \frac{\pi_{kl}}{\pi_k \pi_l} - 1 \right| = O(n_t^{-1}).$$

$$\text{A1'}. \lim_{t \rightarrow \infty} \max_{1 \leq k \neq l \leq N_t} \left| \frac{\pi_{kl}}{\pi_k \pi_l} - 1 \right| = 0.$$

$$\text{A2. } N_t \min_{1 \leq k \leq N_t} \pi_k = O(n_t).$$

$$\text{A2'}. \lim_{t \rightarrow \infty} N_t \min_{1 \leq k \leq N_t} \pi_k = \infty.$$

$$\text{A3. } \lim_{t \rightarrow \infty} \min_{1 \leq k \leq N_t} \pi_k > 0.$$

$$\text{A4. } \exists \lambda_0 > 0, \delta_0 > 0 \text{ and } t_0 > 0 \text{ such that}$$

$$\left. \begin{array}{l} t > t_0 \\ \|\mathbf{B} - \mathbf{B}_{Gt}\| < \delta_0 \end{array} \right\} \Rightarrow \inf_{\|\mathbf{t}\|=1} \frac{1}{N_t} \sum_{i=1}^{N_t} \left(\frac{\mathbf{x}_i^\top \mathbf{t}}{\sigma_i} \right)^2 \psi'_{\bullet_H}(E_{\theta i}^*) \geq \inf_{\|\mathbf{t}\|=1} \frac{\lambda_0}{N_t} \sum_{i=1}^{N_t} (\mathbf{x}_i^\top \mathbf{t} / \sigma_i)^2.$$

where $E_{\theta i}^* = (y_i - \mathbf{x}_i^\top \mathbf{B}_\theta) / (d_i \sigma_i)$, $\mathbf{B}_\theta = \theta \mathbf{B} + (1 - \theta) \mathbf{B}_{Gt}$, $(0 < \theta < 1)$.

$$\text{A5. } 0 < \lim_{t \rightarrow \infty} \inf_{\|\mathbf{t}\|=1} \frac{1}{N_t} \sum_{i=1}^{N_t} (\mathbf{x}_i^\top \mathbf{t} / \sigma_i)^2 < \infty.$$

$$\text{A6. } \lim_{t \rightarrow \infty} \sup_{\|\mathbf{t}\|=1} \frac{1}{N_t} \sum_{i=1}^{N_t} (\mathbf{x}_i^\top \mathbf{t} / \sigma_i)^4 < \infty.$$

$$\text{A7. } \frac{1}{N_t} \sum_{i=1}^{N_t} (\|\mathbf{x}_i\| / \sigma_i)^4 = O(1).$$

A8. There exists a unique solution to the equation $\hat{\mathbf{f}}_t(\mathbf{B}) = 0$ and that solution is

$$\hat{\mathbf{B}}_{\text{CMT}}.$$

$$\text{A9. } \frac{1}{N_t} \sum_{i=1}^{N_t} E_{Gi}^{*4} = O(1), \text{ where } E_{Gi}^* = \frac{y_i - \mathbf{x}_i^\top \mathbf{B}_{Gt}}{d_i \sigma_i} \text{ and } \mathbf{B}_{Gt} = \mathbf{T}_G^{-1} \sum_{i=1}^{N_t} \frac{\mathbf{x}_i y_i}{N_t \sigma_i^2}.$$

$$\text{A10. } \lim_{t \rightarrow \infty} w_{n_t} = w_0 \text{ where } 0 \leq w_0 \leq 1.$$

$$\text{A11. } \frac{1}{N_t} \sum_{i=1}^{N_t} \left[A_{\bullet_H}(E_{Gi}^*) - E_{Gi}^* \right]^2 = O(n_t^{-1}).$$

$$\text{A12. } \lim_{t \rightarrow \infty} \frac{1}{N_t} \sum_{i=1}^{N_t} (y_i / \sigma_i^2)^2 < \infty.$$

$$\text{A13. } \lim_{t \rightarrow \infty} \frac{1}{N_t} \sum_{i=1}^{N_t} (\bar{\mathbf{X}}^\top [\mathbf{T}_\bullet(z)]^{-1} \mathbf{x}_i)^4 < \infty.$$

$$\text{A14. } \max_{1 \leq i \leq N_t} N_t \mathbf{x}_i^\top (\mathbf{X}_p' \mathbf{X}_p)^{-1} \mathbf{x}_i = O(1).$$

Remarks

- The assumptions A1 and A2 are used by Robinson & Särndal (1983) to prove the consistency of the GREG estimator.
- Assumption A3 guarantees that each unit in the population maintains at least a small chance of being selected as the population grows.

- Assumption A4 is not restrictive and is satisfied for most influence functions in the literature. For the Huber-type influence function ψ_H , for example, by taking a small enough δ_0 so that the derivative $\psi'_H(\cdot)$ is equal to 1 for some θ the assumption is satisfied. If $\psi_{\bullet H}$ is replaced with the identity function then $\psi'_{\bullet F} \equiv 1$ and the condition is satisfied for $\lambda_0 = 1$ and for any t_0 and δ_0 .

Under the common mean superpopulation model of chapter 2, this condition must be satisfied to guarantee the validity of the order of magnitude of $\widehat{B}_M - B_M$ and $\widehat{B}_{CM} - B_{CM}$ used as assumptions in theorems 2.2 and 2.4. Under this set-up, condition A4 becomes

$\exists \lambda_0 > 0, \delta_0 > 0$ and $t_0 > 0$ such that

$$(t > t_0 \text{ and } B \in (\bar{Y} - \delta_0, \bar{Y} + \delta_0)) \implies \frac{1}{N_t} \sum_{i=1}^{N_t} \psi'_{\bullet H} \left(\frac{y_i - B_0}{\sigma} \right) \geq \lambda_0,$$

where B_0 lies between B and $\bar{Y}$. Note that under the common mean superpopulation model, $\forall i = 1, \dots, N$ $\sigma_i = \sigma$, $\mathbf{x}_i = 1$. It follows from the definition of $\psi_{\bullet H}$ that it is sufficient to find λ_0 such that

$$\lambda_0 \leq w_{n_t} + (1 - w_{n_t}) \frac{1}{N_t} \sum_{i=1}^{N_t} \psi'_H \left(\frac{y_i - B_0}{\sigma} \right).$$

Note that $\psi'_H(\cdot)$ takes two values: 0 and 1. Therefore, if $\lim_{t \rightarrow \infty} w_{n_t} = w_0 > 0$ then one can take $\lambda_0 = w_0$. Otherwise, since B_0 is in the neighborhood of $\bar{Y}$ there always exists a unit i such that $\psi'_H((y_i - B_0)/\sigma) = 1$ which implies that

$$\lim_{t \rightarrow \infty} \frac{1}{N_t} \sum_{i=1}^{N_t} \psi'_H \left(\frac{y_i - B_0}{\sigma} \right) > 0$$

and take λ_0 equals that limit.

- Assumptions A5, A6 and A7 are standard moment conditions.
- The assumptions such as A9 and A13 that control the skewness of the population are not needed if $w_n = 0$ or the ψ -function used is bounded.

Lemma 3.2. (Ortega & Rheinbolt, 1970, page 163). Let U be a bounded neighborhood of a p -dimensional real vector $\boldsymbol{\beta}$. Suppose that $\mathbf{f} : \bar{U} \rightarrow \mathbb{R}^p$ is a function defined on the closure $\bar{U}$ of U . If $\mathbf{f}$ is continuous and satisfies the condition $(\mathbf{x} - \boldsymbol{\beta})' \mathbf{f}(\mathbf{x}) < 0 \ \forall \mathbf{x} \in \dot{U}$ where $\dot{U}$ is the boundary of U , then the equation $\mathbf{f}(\mathbf{x}) = \mathbf{0}$ has a solution in $\bar{U}$.

Lemma 3.3. Under the assumptions A1', A2', A4, A5 and A6, we have that for any vector $\mathbf{B}$ satisfying $\|\mathbf{B} - \mathbf{B}_{Gt}\| \leq \delta_0$ (see A4),

$$\exists \eta_0 > 0 : \lim_{t \rightarrow \infty} P \left(\inf_{\|\mathbf{t}\|=1} \sum_{i=1}^{n_t} \frac{(\mathbf{x}_i^\top \mathbf{t})^2}{N \pi_i \sigma_i^2} \psi'_{\bullet_H} [\theta_i(\mathbf{B})] \geq \eta_0 \right) = 1.$$

The proof of this lemma is given in Appendix A.

Lemma 3.4. Let $\mathbf{b}_1, \mathbf{b}_2, \dots$ be a sequence of p -dimensional vectors, $\mathbf{b}$ a p -dimensional vector and $\hat{\mathbf{g}}$ a vector-valued sample-dependent function defined on $\mathbb{R}^p$. Suppose that $\hat{\mathbf{g}}$ is continuous and differentiable for all samples and that there exists a function $\mathbf{g}$ also continuous satisfying $\hat{\mathbf{g}}(\mathbf{b}) \xrightarrow{p} \mathbf{g}(\mathbf{b})$ for any random vector $\mathbf{b}$. If $\mathbf{b}_t \xrightarrow{p} \mathbf{b}$ then $\hat{\mathbf{g}}(\mathbf{b}_t) \xrightarrow{p} \mathbf{g}(\mathbf{b})$.

The proof of this lemma is omitted. But a similar result is proved within the framework of parametric statistical inference by Serfling (1980, page 24).

Theorem 3.3. *If the assumptions A1, A2, A4 to A10 are satisfied then*

$$\widehat{\mathbf{B}}_{\text{CMT}} - \mathbf{B}_{\text{CMT}} = O_p(n_t^{-1/2}) \text{ and } \widehat{\mathbf{B}}_{\text{CMT}} - \mathbf{B}_{\text{GT}} = O_p(n_t^{-1/2}).$$

The proof of this theorem involves some techniques that are not needed in the proof of the consistency of the CM-estimator. It can be found in appendix A. Note also that the conditions A9 and A10 imply A11 for the $\psi_{\bullet H}$ we are using. However, a ψ -function satisfying A11 does not have to satisfy neither A9 nor A10.

We need the following lemma for proving consistency of the CM-estimator. We recall that the $\mathbf{x}_i$'s are p dimensional vectors. Sometimes it might be more convenient to use the notation p_t instead of p where p_t refers to the dimension of vector $\mathbf{x}$ for the t -th finite population.

Lemma 3.5. *Suppose that conditions of theorem 3.3 and the following three conditions are satisfied:*

- i. $\max_{1 \leq k \leq N_t} \|\mathbf{x}_k\|^2 = O(p_t)$.
- ii. $p_t/n_t \rightarrow 0$ as $t \rightarrow \infty$.
- iii. $\forall \epsilon > 0, \exists \delta_\epsilon > 0 : |r| \leq \delta_\epsilon \implies |\mathcal{A}(r) - 1| \leq \epsilon/2$, where,

$$\mathcal{A}(r) = \begin{cases} 1 & \text{if } r = 0, \\ \psi_{\bullet H}(r)/r & \text{otherwise.} \end{cases}$$

Then $\max_{1 \leq k \leq N_t} |\alpha_{\bullet k} - A_{\bullet k}| = o_p(1)$.

Proof

Let $\epsilon_0 > 0$. We have to prove that,

$$\lim_{t \rightarrow \infty} P\left(\max_{1 \leq k \leq N_t} |\alpha_{\bullet k} - A_{\bullet k}| > \epsilon_0\right) = 0.$$

It follows from assumption (iii) that there exists $\delta_{\epsilon_0} > 0$ such that if $|r| \leq \delta_{\epsilon_0}$ then

$|\mathcal{A}(r) - 1| \leq \epsilon_0$. We have that for all k ,

$$|\alpha_{\bullet k} - A_{\bullet k}| = |\alpha_{\bullet k} - A_{\bullet k}|(I_{1k} + I_{2k} + I_{3k} + I_{4k}),$$

where

$$I_{1k} = I(|e_{CMk}^{(s)}| \leq \delta_{\epsilon_0} \ \& \ |E_{CMk}^{(s)}| \leq \delta_{\epsilon_0}), \ I_{2k} = I(|e_{CMk}^{(s)}| \leq \delta_{\epsilon_0} \ \& \ |E_{CMk}^{(s)}| > \delta_{\epsilon_0}),$$

$$I_{3k} = I(|e_{CMk}^{(s)}| > \delta_{\epsilon_0} \ \& \ |E_{CMk}^{(s)}| \leq \delta_{\epsilon_0}), \ I_{4k} = I(|e_{CMk}^{(s)}| > \delta_{\epsilon_0} \ \& \ |E_{CMk}^{(s)}| > \delta_{\epsilon_0}).$$

We have that

$$\begin{aligned} P\left(\max_{1 \leq k \leq N_t} |\alpha_{\bullet k} - A_{\bullet k}| > \epsilon_0\right) &\leq P\left(\max_{1 \leq k \leq N_t} I_{1k} |\alpha_{\bullet k} - A_{\bullet k}| > \epsilon_0\right) \\ &\quad + P\left(\max_{1 \leq k \leq N_t} I_{2k} |\alpha_{\bullet k} - A_{\bullet k}| > \epsilon_0\right) \\ &\quad + P\left(\max_{1 \leq k \leq N_t} I_{3k} |\alpha_{\bullet k} - A_{\bullet k}| > \epsilon_0\right) \\ &\quad + P\left(\max_{1 \leq k \leq N_t} I_{4k} |\alpha_{\bullet k} - A_{\bullet k}| > \epsilon_0\right) \end{aligned}$$

To complete the proof, it is sufficient to prove that each of the four terms on the right-hand-side of the above equality goes to 0 as t goes to infinity.

a) Note that,

$$\begin{aligned} \max_{1 \leq k \leq N_t} I_{1k} |\alpha_{\bullet k} - A_{\bullet k}| &\leq \max_{1 \leq k \leq N_t} I_{1k} |\alpha_{\bullet k} - 1| + \max_{1 \leq k \leq N_t} I_{1k} |A_{\bullet k} - 1|, \\ &\leq \epsilon_0 \text{ (from (iii)).} \end{aligned}$$

Thus,

$$P\left(\max_{1 \leq k \leq N_t} I_{1k} |\alpha_{\bullet k} - A_{\bullet k}| > \epsilon_0\right) = 0.$$

b) We want to prove that

$$P\left(\max_{1 \leq k \leq N_t} I_{4k} |\alpha_{\bullet k} - A_{\bullet k}| > \epsilon_0\right) = 0.$$

We have that

$$\begin{aligned} \max_{1 \leq k \leq N_t} I_{4k} |\alpha_{\bullet k} - A_{\bullet k}| &= \max_{1 \leq k \leq N_t} I_{4k} \left| \psi_{\bullet H}(E_{CMk}^{(s)}) / E_{CMk}^{(s)} - \psi_{\bullet H}(e_{CMk}^{(s)}) / e_{CMk}^{(s)} \right|, \\ &\leq 2\delta_0^{-1} \|\hat{\mathbf{B}}_{CM} - \mathbf{B}_{CM}\| \max_{1 \leq k \leq N_t} \|\mathbf{x}_k\|. \end{aligned}$$

The above inequality follows from the mean value theorem that ensures that

$$\psi_{\bullet H}(e_{CMk}^{(s)}) = \psi_{\bullet H}(E_{CMk}^{(s)}) + \mathbf{x}_k^\top (\mathbf{B}_{CM} - \hat{\mathbf{B}}_{CM}) \psi'_{\bullet H}(E_{CMk}^{(s)}(\theta)),$$

where $E_{CMk}^{(s)}(\theta) = y_k - \mathbf{x}_k^\top \mathbf{B}_\theta$ and $\mathbf{B}_\theta = \theta \hat{\mathbf{B}}_{CM} - \mathbf{B}_{CM}$.

c) We will only prove that the second probability involving the I_{2k} 's goes to 0 as

t goes to infinity. The case involving the I_{4k} 's can be carried out similarly.

Note that for all $k = 1, \dots, N_t$,

$$I_{2k} = I_{2k}^* + I_{2k}^{**},$$

where

$$I_{2k}^* = I\left(\delta_{\epsilon_0}/2 \leq |e_{\text{CM}k}^{(s)}| \leq \delta_{\epsilon_0} \ \& \ |E_{\text{CM}k}^{(s)}| > \delta_{\epsilon_0}\right)$$

and

$$I_{2k}^{**} = I\left(|e_{\text{CM}k}^{(s)}| < \delta_{\epsilon_0}/2 \ \& \ |E_{\text{CM}k}^{(s)}| > \delta_{\epsilon_0}\right).$$

We have that

$$\begin{aligned} P\left(\max_{1 \leq k \leq N_t} I_{2k} |\alpha_{\bullet k} - A_{\bullet k}| > \epsilon_0\right) &\leq P\left(\max_{1 \leq k \leq N_t} I_{2k}^* |\alpha_{\bullet k} - A_{\bullet k}| > \epsilon_0\right) \\ &\quad + P\left(\max_{1 \leq k \leq N_t} I_{2k}^{**} = 1\right). \end{aligned}$$

Note that for all k ,

$$\begin{aligned} |E_{\text{CM}k}^{(s)}| &\leq |e_{\text{CM}k}^{(s)}| + |\mathbf{x}_k^\top (\hat{\mathbf{B}}_{\text{CM}} - \mathbf{B}_{\text{CM}})|, \\ &\leq \|\hat{\mathbf{B}}_{\text{CM}} - \mathbf{B}_{\text{CM}}\| \max_{1 \leq k \leq N_t} \|\mathbf{x}_k\| + |e_{\text{CM}k}^{(s)}|. \end{aligned}$$

It follows that if $\max_{1 \leq k \leq N_t} I_{2k}^{**} = 1$ then there exists k_0 such that $|E_{\text{CM}k_0}^{(s)}| > \delta_{\epsilon_0}$ and $|e_{\text{CM}k_0}^{(s)}| < \delta_{\epsilon_0}/2$ which implies that

$$\|\hat{\mathbf{B}}_{\text{CM}} - \mathbf{B}_{\text{CM}}\| \max_{1 \leq k \leq N_t} \|\mathbf{x}_k\| > \delta_{\epsilon_0}.$$

Thus,

$$\begin{aligned} P\left(\max_{1 \leq k \leq N_t} I_{2k}^{**} = 1\right) &\leq P\left(\|\hat{\mathbf{B}}_{\text{CM}} - \mathbf{B}_{\text{CM}}\| \max_{1 \leq k \leq N_t} \|\mathbf{x}_k\| > \delta_{\epsilon_0}\right), \\ &= o(1) \text{ (from theorem 3.3 , (i) and (ii).)} \end{aligned}$$

Now, the fact that

$$P\left(\max_{1 \leq k \leq N_t} I_{2k} |\alpha_{\bullet k} - A_{\bullet k}| > \epsilon_0\right) \leq P\left(\max_{1 \leq k \leq N_t} I_{2k}^* |\alpha_{\bullet k} - A_{\bullet k}| > \epsilon_0\right) = o(1)$$

can be proved using the same techniques as in case b).

■

Theorem 3.4. *Suppose that the conditions of theorem 3.3 and those of lemma 3.5 are all satisfied as well as assumptions A12, A13 and A14. Then $\widehat{Y}_{\text{CM}}$ is a consistent estimator of the population mean $\bar{Y}$ for any variable of interest y .*

Proof of Theorem 3.4

Let y be a variable of interest and z the key survey variable the robust regression weights are built upon. Note that these two variables might well be the same variable. We have to prove that for all $\epsilon > 0$, $\lim_{t \rightarrow \infty} P(|\widehat{Y}_{\text{CM}t} - \bar{Y}_t| > \epsilon) = 0$. Let $\widehat{Y}_{\text{CM}0t}$ be defined by

$$\widehat{Y}_{\text{CM}0t} = \sum_{i=1}^{n_t} \frac{G_{\text{CM}i}}{\widehat{N}_t \pi_i} y_i$$

where

$$\begin{cases} G_{\text{CM}i} = \bar{\mathbf{X}}^\top [\mathbf{T}_\bullet(z)]^{-1} \mathbf{x}_i A_{\bullet i}(z) / \sigma_i^2, \\ \mathbf{T}_\bullet(z) = \sum_{k=1}^{N_t} \frac{A_{\bullet k}(z)}{N_t \sigma_k^2} \mathbf{x}_k \mathbf{x}_k'. \end{cases}$$

This is a population entity that is not a statistic but is used here simply as a tool for proving consistency of the CM-estimator. Let $\epsilon_0 > 0$. We have that

$$\begin{aligned} P(|\widehat{Y}_{\text{CM}t} - \bar{Y}_t| > \epsilon_0) &\leq \\ &\leq P(|\widehat{Y}_{\text{CM}t} - \widehat{Y}_{\text{CM}0t}| > \epsilon_0/3) + P(|\widehat{Y}_{\text{CM}0t} - \widehat{Y}_{\text{G}t}| > \epsilon_0/3) + \\ &\quad + P(|\widehat{Y}_{\text{G}t} - \bar{Y}_t| > \epsilon_0/3). \end{aligned} \tag{3.12}$$

Thus we have to prove that each of the three terms on the right hand side of the above inequality goes to 0 as t goes to ∞ . We have that

$$\begin{aligned}
\widehat{Y}_{\text{CM}t} - \widehat{Y}_{\text{CM}0t} &= \sum_{i=1}^{n_t} \frac{1}{\widehat{N}_t \pi_i} (g_{\text{CM}i} - G_{\text{CM}i}) y_i, \\
&= \overline{\mathbf{X}}^\top [\mathbf{T}_\bullet(z)]^{-1} \sum_{i=1}^{n_t} \frac{1}{\widehat{N}_t \sigma_i^2 \pi_i} \left(\mathbf{T}_\bullet(z) \alpha_{\bullet i}(z) - \widehat{\mathbf{T}}_\bullet(z) A_{\bullet i}(z) \right) [\mathbf{T}_\bullet(z)]^{-1} \mathbf{x}_i y_i, \\
&= \overline{\mathbf{X}}^\top [\mathbf{T}_\bullet(z)]^{-1} \sum_{i=1}^{n_t} \frac{1}{\widehat{N}_t \pi_i} \left\{ \mathbf{T}_\bullet(z) \left(\alpha_{\bullet i}(z) - A_{\bullet i}(z) \right) + \left(\mathbf{T}_\bullet(z) - \widehat{\mathbf{T}}_\bullet(z) \right) \times \right. \\
&\quad \left. \times A_{\bullet i}(z) \right\} [\mathbf{T}_\bullet(z)]^{-1} \mathbf{x}_i y_i / \sigma_i^2, \\
&= \sum_{i=1}^{n_t} \left(\alpha_{\bullet i}(z) - A_{\bullet i}(z) \right) \frac{\overline{\mathbf{X}}^\top [\widehat{\mathbf{T}}_\bullet(z)]^{-1} \mathbf{x}_i y_i}{\widehat{N}_t \sigma_i^2 \pi_i} + \\
&\quad + \sum_{i=1}^{n_t} \frac{y_i A_{\bullet i}(z)}{\widehat{N}_t \sigma_i^2 \pi_i} \overline{\mathbf{X}}^\top [\widehat{\mathbf{T}}_\bullet(z)]^{-1} \widehat{\Delta}_{\text{CM}} [\mathbf{T}_\bullet(z)]^{-1} \mathbf{x}_i = a_t + b_t,
\end{aligned}$$

where $\widehat{\Delta}_{\text{CM}} = \mathbf{T}_\bullet(z) - \widehat{\mathbf{T}}_\bullet(z)$. We have to prove that $a_t = o_p(1)$ and $b_t = o_p(1)$.

Conditions A1 and A2 imply that $\widehat{N}_t/N_t$ converges in probability to 1. Moreover,

Assumption A14 and lemma 3.5 imply that $\widehat{\mathbf{T}}_\bullet(z)$ converges in probability to TM .

It follows from lemma 3.4 that in order to prove that $b_t = o_p(1)$, it is sufficient to

prove that $b_{0t} = o_p(1)$ where

$$b_{0t} = \overline{\mathbf{X}}^\top [\mathbf{T}_\bullet(z)]^{-1} \widehat{\Delta}_{\text{CM}}^{(0)} [\mathbf{T}_\bullet(z)]^{-1} \widehat{U}_0$$

and

$$\widehat{\Delta}_{\text{CM}}^{(0)} = \frac{1}{N_t} \sum_{i=1}^{N_t} \sigma_i^{-2} \mathbf{x}_i \mathbf{x}_i^\top (A_{\bullet i}(z) - \alpha_{\bullet i} I_i / \pi_i), \quad \widehat{U}_0 = \frac{1}{N_t} \sum_{i=1}^{n_t} A_{\bullet k}(z) y_i \mathbf{x}_i / \sigma_i^2 \pi_i.$$

Note that

$$\widehat{\Delta}_{\text{CM}}^{(0)} = \frac{1}{N_t} \sum_{i=1}^{N_t} \sigma_i^{-2} A_{\bullet i}(z) \left(1 - I_i / \pi_i \right) \left(\overline{\mathbf{X}}^\top [\mathbf{T}_\bullet(z)]^{-1} \mathbf{x}_i \right) \left(\mathbf{x}_i^\top [\mathbf{T}_\bullet(z)]^{-1} \widehat{U}_0 \right)$$

$$\begin{aligned}
 & + \frac{1}{N_t} \sum_{i=1}^{N_t} \sigma_k^{-2} \left(A_{\bullet i}(z) - \alpha_{\bullet i}(z) \right) \frac{I_i}{\pi_i} \left(\bar{\mathbf{X}}^\top [\mathbf{T}_\bullet(z)]^{-1} \mathbf{x}_i \right) \left(\mathbf{x}_i^\top [\mathbf{T}_\bullet(z)]^{-1} \widehat{\mathbf{U}}_0 \right). \\
 & = b_{0t}^* + b_{0t}^{**}.
 \end{aligned}$$

We have that

$$\begin{aligned}
 |b_{0t}^{**}| & \leq \max_{1 \leq k \leq N_t} |A_{\bullet k} - \alpha_{\bullet k}| \times \frac{1}{N_t} \sum_{i=1}^{N_t} \sigma_i^{-2} \frac{I_i}{\pi_i} \left(\bar{\mathbf{X}}^\top [\mathbf{T}_\bullet(z)]^{-1} \mathbf{x}_i \right) \left(\mathbf{x}_i^\top [\mathbf{T}_\bullet(z)]^{-1} \widehat{\mathbf{U}}_0 \right). \\
 & = o_p(1) O_p(1) = o_p(1).
 \end{aligned}$$

The fact that $b_{0t}^* = o_p(1)$ is trivial. Moreover, $|a_t| = o_p(1)$ is a direct consequence of lemma 3.5 .

■

Theorem 3.5. *Assume all the conditions of Theorem 3.3 and A14. In addition to that, we assume that there exists t_0 such that if $t > t_0$ then the second-order derivative $\widehat{\mathbf{f}}_t''$ of $\widehat{\mathbf{f}}_t$ exists in the neighborhood of $\mathbf{B}_{\text{CM}t}$ in which case $|\psi'_{\bullet H}| \leq 1$ and $|\psi''_{\bullet H}| \leq 1$. Then the CM-estimator takes the following linear form,*

$$\begin{aligned}
 \widehat{\mathbf{Y}}_{\text{CM}t} & = \bar{\mathbf{X}}^\top \mathbf{B}_{\text{CM}t} + \\
 & + \bar{\mathbf{X}}^\top \left(\frac{1}{N_t} \sum_{i=1}^{N_t} \sigma_i^{-2} \psi'_{\bullet H}(E_{\text{CM}i}^{(s)}) \mathbf{x}_i \mathbf{x}_i^\top \right)^{-1} \sum_{k=1}^{n_t} \frac{d_k}{N_t \pi_k \sigma_k} \mathbf{x}_k \psi_{\bullet H}(E_{\text{CM}k}^{(s)}) + O_p(n_t^{-1}).
 \end{aligned}$$

where $E_{\text{CM}i}^{(s)} = (y_i - \mathbf{x}_i^\top \mathbf{B}_{\text{CM}t}) / (d_i \sigma_i)$.

Proof of theorem 3.5 .

It follows from the mean value theorem that

$$\begin{aligned}
 \widehat{\mathbf{f}}_t(\widehat{\mathbf{B}}_{\text{CM}t}) & = \widehat{\mathbf{f}}_t(\mathbf{B}_{\text{CM}t}) + [\widehat{\mathbf{f}}_t'(\mathbf{B}_{\text{CM}t})](\widehat{\mathbf{B}}_{\text{CM}t} - \mathbf{B}_{\text{CM}t}) + \\
 & + (1/2)(\widehat{\mathbf{B}}_{\text{CM}t} - \mathbf{B}_{\text{CM}t})' \widehat{\mathbf{f}}_t''(\mathbf{B}_{\theta \text{CM}t})(\widehat{\mathbf{B}}_{\text{CM}t} - \mathbf{B}_{\text{CM}t}),
 \end{aligned}$$

where $\mathbf{B}_{\theta Mt} = \theta \hat{\mathbf{B}}_{Cmt} + (1 - \theta) \mathbf{B}_{Cmt}$ for some $\theta \in (0, 1)$. We have

$$\hat{\mathbf{f}}'_t(\mathbf{B}_{Cmt}) = \sum_{i=1}^{n_t} \frac{\psi'_{\bullet H}(E_{Cmi}^{(s)})}{N\pi_i\sigma_i^2} \mathbf{x}_i \mathbf{x}_i^\top \text{ and } \hat{\mathbf{f}}''_t(\mathbf{B}_{\theta Mt}) = \sum_{i=1}^{n_t} \frac{\psi''_{\bullet H}(E_{\theta Mi}^{(s)})}{N\pi_i d_i \sigma_i^3} \mathbf{x}_i (\mathbf{x}_i^\top \mathbf{x}_i).$$

Using this, we can see that

$$\begin{aligned} (1/2)(\hat{\mathbf{B}}_{Cmt} - \mathbf{B}_{Cmt})' \hat{\mathbf{f}}''_t(\mathbf{B}_{\theta Mt}) (\hat{\mathbf{B}}_{Cmt} - \mathbf{B}_{Cmt}) &= \\ &= O_p \left(\|\hat{\mathbf{B}}_{Cmt} - \mathbf{B}_{Cmt}\|^2 \frac{1}{N_t} \sum_{i=1}^{N_t} d_i^{-1} \psi''_{\bullet H}(E_{Cmi}^{(s)}) (\|\mathbf{x}_i\|/\sigma_i)^3 \right). \end{aligned}$$

It follows from theorem 3.3, and from the assumptions A7 and A14 that the order of magnitude of the above quantity is $O_p(n_t^{-1})$. Therefore,

$$\begin{aligned} \hat{\mathbf{f}}_t(\hat{\mathbf{B}}_{Cmt}) &= \hat{\mathbf{f}}_t(\mathbf{B}_{Cmt}) + [\hat{\mathbf{f}}'_t(\mathbf{B}_{Cmt})](\hat{\mathbf{B}}_{Cmt} - \mathbf{B}_{Cmt}) + \\ &+ \left([\hat{\mathbf{f}}'_t(\mathbf{B}_{Cmt})] - [\mathbf{f}'_t(\mathbf{B}_{Cmt})] \right) (\hat{\mathbf{B}}_{Cmt} - \mathbf{B}_{Cmt}) + O_p(n_t^{-1}). \end{aligned}$$

It remains to show that the third term on the right hand side of the above equation is $O_p(n_t^{-1})$. For any $p \times p$ -matrix $\mathbf{A} = (a_{ij})_{i,j=1,\dots,p}$ let $\|\mathbf{A}\|_m$ be defined by

$$\|\mathbf{A}\|_m^2 = \sum_{i=1}^p \sum_{j=1}^p a_{ij}^2.$$

It is straightforward to check that for any p -dimensional vector $\mathbf{x}$ we have the inequality $\|\mathbf{Ax}\|_2 \leq \|\mathbf{A}\|_m \|\mathbf{x}\|_2$ where $\|\cdot\|_2$ stands for the euclidean norm. Using the definition of $\|\cdot\|_m$ we obtain

$$\begin{aligned} \left\| \left([\hat{\mathbf{f}}'_t(\mathbf{B}_{Cmt})] - [\mathbf{f}'_t(\mathbf{B}_{Cmt})] \right) (\hat{\mathbf{B}}_{Cmt} - \mathbf{B}_{Cmt}) \right\| &\leq \\ \left\| [\hat{\mathbf{f}}'_t(\mathbf{B}_{Cmt})] - [\mathbf{f}'_t(\mathbf{B}_{Cmt})] \right\|_m \|\hat{\mathbf{B}}_{Cmt} - \mathbf{B}_{Cmt}\| &= a_t \|\hat{\mathbf{B}}_{Cmt} - \mathbf{B}_{Cmt}\|. \end{aligned}$$

It is enough to show that $a_t = O(n_t^{-1/2})$. A sufficient condition for that is that $E|a_t| = O(n_t^{-1/2})$. Note that

$$a_t^2 = \sum_{i=1}^p \sum_{j=1}^p \left\{ \frac{1}{N_t} \sum_{k=1}^{N_t} \psi'_{\bullet H}(E_{Cmk}^*) x_{ki} x_{kj} / \sigma_k^2 \left(\frac{I_{k,t}}{\pi_k} - 1 \right) \right\}^2.$$

Therefore,

$$\begin{aligned} Ea_t^2 &\leq \sum_{i=1}^p \sum_{j=1}^p \left(\frac{1}{\min_{1 \leq l \leq N_t} \pi_l} + \max_{1 \leq l \neq l' \leq N_t} \left| \frac{\pi_{ll'}}{\pi_l \pi_{l'}} - 1 \right| \right) \frac{1}{N_t} \sum_{k=1}^{N_t} x_{ki}^2 x_{kj}^2 / \sigma_k^4, \\ &\leq O(n_t^{-1}) O\left(\frac{1}{N_t} \sum_{i=1}^{N_t} (\|\mathbf{x}_k\| / \sigma_k)^4 \right) = O(n_t^{-1}) \end{aligned}$$

■

Theorem 3.5 can be used to calculate the asymptotic variance of the CM-estimator.

It follows from that theorem that the CM-estimator is asymptotically equivalent to

$\tilde{\bar{Y}}_{\text{CM}}$ given by

$$\tilde{\bar{Y}}_{\text{CM}} = \bar{\mathbf{X}}^\top \mathbf{B}_{\text{CM}} + \sum_{k=1}^n \frac{(\bar{\mathbf{X}}^\top \mathbf{T}_\bullet^{-1} \mathbf{x}_k / \sigma_k^2)}{N \pi_k} \sigma_k d_k \psi_{\bullet H}(E_{1mk}^{(s)}),$$

where

$$\mathbf{T}_\bullet = \frac{1}{N} \sum_{i=1}^N \sigma_i^{-2} \psi'_{\bullet H}(E_{\text{CM}i}^{(s)}) \mathbf{x}_i \mathbf{x}_i^\top.$$

To simplify the notations, the robustness function will be denoted by ψ instead of $\psi_{\bullet H}$ when there is no confusion. The asymptotic variance of the CM-estimator is then given by

$$AV(\hat{\bar{Y}}_{\text{CM}}) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) (g_i R_i) (g_j R_j), \quad (3.13)$$

where $R_i = \sigma_i \psi(E_{\text{CM}i}^{(s)})$ and $g_i = (d_i \mathbf{x}_i^\top \mathbf{T}_\bullet^{-1} \bar{\mathbf{X}} / \sigma_i^2)$.

A natural asymptotically unbiased estimator we propose for estimating this variance is given by

$$\widehat{AV}(\hat{\bar{Y}}_{\text{CM}}) = -\frac{1}{2\widehat{N}^2} \sum_{i=1}^n \sum_{j=1}^n \left(1 - \pi_i \pi_j / \pi_{ij} \right) \left(\frac{\hat{g}_i r_i}{\pi_i} - \frac{\hat{g}_j r_j}{\pi_j} \right)^2, \quad (3.14)$$

where $r_i = \hat{\sigma}_i \psi(e_{\text{CM}i}^{(s)})$ and $\hat{g}_i = (d_i \mathbf{x}_i^\top \hat{\mathbf{T}}_\bullet^{-1} \bar{\mathbf{X}} / \hat{\sigma}_i^2)$.

Note that if the ψ -function is replaced with the traditional identity function, then this variance estimator reduces to the estimator proposed by Särndal et al. (1989) if we assume the cosmetic form of the GREG estimator. Our general approach provides another justification for that estimator.

The CM-estimator although biased, might have a non negligible bias especially for small samples. It follows from theorem 3.5 that this bias can be approximated by

$$B_a(\hat{Y}_{\text{CM}}) = \bar{\mathbf{X}}^\top (\mathbf{B}_{\text{CM}} - \mathbf{B}_G).$$

It follows from the definition of $\mathbf{B}_{\text{CM}}$ that the order of magnitude for most populations of this bias is at most $O(n^{-1/\alpha})$. Note that this order of magnitude can be further reduced by decreasing the value of α . Given that the variance will in general be of order $O(n^{-1})$, imposing $\alpha < 2$ guarantees that in any case, the bias ratio goes to 0 as the sample size increases.

A natural estimator of this bias is given by $\hat{B}_a(\hat{Y}_{\text{CM}})$. However, in estimating the MSE of the CM-estimator $\hat{B}_a^2(\hat{Y}_{\text{CM}})$ might lead to a severe underestimation of the squared bias. It is more efficient to estimate the squared bias by

$$\hat{B}_{2a}(\hat{Y}_{\text{CM}}) = \max\left\{0, \hat{B}_a^2(\hat{Y}_{\text{CM}}) - \hat{V}\left[\hat{B}_a(\hat{Y}_{\text{CM}})\right]\right\}.$$

A consistent estimator of the MSE of the CM-estimator is given by

$$\widehat{MSE}(\hat{Y}_{\text{CM}}) = \hat{V}(\hat{Y}_{\text{CM}}) + \hat{B}_{2a}(\hat{Y}_{\text{CM}}).$$

3.5.2. Asymptotic Properties of the LM-estimator

We start this section by proving the consistency of the LM-estimator. Then a linear approximation to the LM-estimator is developed. This linearization is important in deriving the asymptotic variance and the approximate bias for small samples.

Note that the LM-estimator is built upon the M-estimator $\hat{\mathbf{B}}_{\mathbf{M}}$ defined in equation 3.5 . It is possible to prove a theorem similar to theorem 3.3 for $\hat{\mathbf{B}}_{\mathbf{M}}$ where $\mathbf{B}_{\mathbf{CM}}$ is replaced with $\mathbf{B}_{\mathbf{M}}$. That theorem would simply be a particular case of theorem 3.3 and would require less regularity conditions. For that reason we will often appeal to theorem 3.3 when the equation $\|\hat{\mathbf{B}}_{\mathbf{M}} - \mathbf{B}_{\mathbf{M}}\| = O_p(n^{-1/2})$ is needed.

Theorem 3.6. *Assume the conditions of theorem 3.3 are satisfied as well as the following ones:*

$$C1 \quad \lim_{t \rightarrow \infty} \mathbf{B}_{\mathbf{M}t} = \boldsymbol{\beta}_{\mathbf{M}}.$$

$$C2 \quad \frac{1}{N_t} \sum_{i=1}^{N_t} \|\mathbf{x}_i\|^2 = O(1).$$

$$C3 \quad \frac{1}{N_t} \sum_{i=1}^{N_t} y_i^2 = O(1).$$

Then the LM-estimator is a consistent estimator of the population mean $\bar{Y}$.

Proof of theorem 3.6 .

We have to prove that for all $\epsilon > 0$

$$\lim_{t \rightarrow \infty} P(|\hat{Y}_{\mathbf{LM}t} - \bar{Y}_t| > \epsilon) = 0.$$

Recall that the LM-estimator is given by

$$\widehat{Y}_{\text{LM}t} = \overline{\mathbf{X}}^\top \widehat{\mathbf{B}}_{\text{M}} + \sum_{i=1}^{n_t} \frac{e_{\text{Mi}}^*}{\widehat{N}_t \pi_i} \text{ where } e_{\text{Mi}}^* = [\mathbf{w}_n + (1 - \mathbf{w}_n) \hat{j}_{ni}] e_{\text{Mi}}.$$

After some straightforward algebra, one gets

$$\begin{aligned} \widehat{Y}_{\text{LM}t} - \overline{Y}_t &= \left(\overline{\mathbf{X}}^\top - \sum_{i=1}^{n_t} \frac{\mathbf{x}'_i}{\widehat{N}_t \pi_i} \right) \mathbf{B}_{\text{M}t} + \left(\overline{\mathbf{X}}^\top - \sum_{i=1}^{n_t} \frac{\mathbf{x}'_i}{\widehat{N}_t \pi_i} \right) (\widehat{\mathbf{B}}_{\text{M}t} - \mathbf{B}_{\text{M}t}) - \\ &\quad - (1 - \mathbf{w}_n) \left(\sum_{i=1}^{n_t} (1 - \hat{j}_{ni}) \frac{\mathbf{x}'_i}{\widehat{N}_t \pi_i} \right) \mathbf{B}_{\text{M}t} \\ &\quad + (1 - \mathbf{w}_n) \left(\sum_{i=1}^{n_t} (1 - \hat{j}_{ni}) \frac{\mathbf{x}'_i}{\widehat{N}_t \pi_i} \right) (\widehat{\mathbf{B}}_{\text{M}t} - \mathbf{B}_{\text{M}t}) - \\ &\quad - (1 - \mathbf{w}_n) \sum_{i=1}^{n_t} (1 - \hat{j}_{ni}) \frac{y_i}{\widehat{N}_t \pi_i} + \left(\sum_{i=1}^{n_t} \frac{y_i}{\widehat{N}_t \pi_i} - \overline{Y}_t \right), \\ &= \Delta_{1t} + \Delta_{2t} + \Delta_{3t} + \Delta_{4t} + \Delta_{5t} + \Delta_{6t}. \end{aligned}$$

Let $\epsilon_0 > 0$. We have

$$P(|\widehat{Y}_{\text{LM}t} - \overline{Y}_t| > \epsilon_0) \leq \sum_{i=1}^6 P(|\Delta_{i,t}| > \epsilon_0/6).$$

Therefore, it is enough to prove that $\forall i \quad \lim_{t \rightarrow \infty} P(|\Delta_{i,t}| > \epsilon_0/6) = 0$. Proving this is pretty straightforward for some of these Δ_i 's. In fact, it follows from the assumptions A1 and A2 that $\Delta_{1,t} = O_p(n_t^{-1/2})$. Theorem 3.3 and the assumptions A1 and A2 imply that $\Delta_{2t} = O_p(n_t^{-1/2}) O_p(n_t^{-1/2}) = O_p(n_t^{-1})$. Theorem 3.3 and the condition C2 imply that $\Delta_{4t} = O_p(n_t^{-1/2})$. The proof that $\Delta_{6t} = O_p(n_t^{-1/2})$ is a standard one (see Robinson & Särndal, 1983). It remains to prove that $\Delta_{3t} = o_p(1)$ and that $\Delta_{5t} = o_p(1)$. But these two proofs are similar and therefore only the fact that $\Delta_{3t} = o_p(1)$ will be proved.

Using Markov's theorem we can see that it is enough to prove that $\lim_{t \rightarrow \infty} |\Delta_{3t}| = 0$ or that

$$\lim_{t \rightarrow \infty} E \left| \sum_{i=1}^{N_t} (1 - \hat{J}_{ni}) \frac{\mathbf{x}_i' I_{it}}{N_t \pi_i} \right| = 0.$$

But

$$\begin{aligned} E \left| \sum_{i=1}^{N_t} (1 - \hat{J}_{ni}) \frac{\mathbf{x}_i' I_{it}}{N_t \pi_i} \right| &\leq \sum_{i=1}^{N_t} \frac{1}{N_t \pi_i} \|\mathbf{x}_i\| E(1 - \hat{J}_{ni}) I_{it}, \\ &\leq \sum_{i=1}^{N_t} \frac{\|\mathbf{x}_i\|}{N_t \pi_i} E^{1/2} (1 - \hat{J}_{ni})^2 E^{1/2} I_{it}^2, \\ &\leq \left(\sum_{i=1}^{N_t} E(1 - \hat{J}_{ni})^2 \right)^{1/2} \left(\sum_{i=1}^{N_t} \frac{\|\mathbf{x}_i\|^2 E(I_{it}^2)}{N_t \pi_i^2} \right)^{1/2}, \\ &\leq \left\{ \left(\min_{1 \leq i \leq N_t} \pi_i \right)^{-1} \frac{1}{N_t} \sum_{i=1}^{N_t} \|\mathbf{x}_i\|^2 \right\}^{1/2} \left(\frac{1}{N_t} \sum_{i=1}^{N_t} E(1 - \hat{J}_{ni})^2 \right)^{1/2}. \end{aligned}$$

It follows from the definition of $\hat{J}_{ni}$ that

$$\begin{aligned} E(1 - \hat{J}_{ni})^2 &= \left\{ 1 - [1 - (\lambda + 1)\alpha_n]^{-1} \right\}^2 P(\lambda\alpha_n < \hat{\hat{F}}(e_{Mi}) \leq 1 - \alpha_n) + \\ &\quad + 1 - P(\lambda\alpha_n < \hat{\hat{F}}(e_{Mi}) \leq 1 - \alpha_n), \\ &= 1 + \left\{ \left[1 - (1 - (\lambda + 1)\alpha_n)^{-1} \right]^2 - 1 \right\} P(\lambda\alpha_n < \hat{\hat{F}}(e_{Mi}) \leq 1 - \alpha_n). \end{aligned}$$

Thus,

$$\begin{aligned} \frac{1}{N_t} \sum_{i=1}^{N_t} E(1 - \hat{J}_{ni})^2 &= 1 + \left\{ \left[1 - (1 - (\lambda + 1)\alpha_n)^{-1} \right]^2 - 1 \right\} \times \\ &\quad \times \frac{1}{N_t} \sum_{i=1}^{N_t} P(\lambda\alpha_{n_i} < \hat{\hat{F}}(e_{Mi}) \leq 1 - \alpha_{n_i}). \end{aligned}$$

As $\lim_{t \rightarrow \infty} \alpha_{n_t} = 0$, we have that

$$\lim_{t \rightarrow \infty} \frac{1}{N_t} \sum_{i=1}^{N_t} P(\lambda\alpha_{n_i} < \hat{\hat{F}}(e_{Mi}) \leq 1 - \alpha_{n_i}) = 1.$$

and the proof follows.

■

Remark.

In proving theorem 3.6 , we have used the fact that $\|\hat{\mathbf{B}}_{\mathbf{M}} - \mathbf{B}_{\mathbf{M}}\| = O_p(n^{-1/2})$.

This has been possible by appealing to theorem 3.3 . However, it is important to mention that the above relation is true even if the assumption A7 is replaced with a much weaker one given by

$$\frac{1}{N_t} \sum_{i=1}^{N_t} (d_i \|\mathbf{x}_i\| / \sigma_i)^4 = O(1).$$

Furthermore, the assumptions A9 and A10 can simply be dropped for the M-estimator.

Now we like to derive a linear representation for the LM-estimator to make the calculation of the asymptotic variance possible. Let $\hat{\hat{F}}$ be the distribution of the empirical residuals. That is for any real number r ,

$$\hat{\hat{F}}(r) = \sum_{i=1}^n \frac{\Delta(e_{\mathbf{M}i} \leq r)}{\hat{N}\pi_i}, \quad e_{\mathbf{M}i} = y_i - \mathbf{x}'_i \hat{\mathbf{B}}_{\mathbf{M}}.$$

Let $\tilde{F}$ (resp. F) be the empirical (resp. theoretical) distribution of the theoretical residuals $E_{\mathbf{M}i}$. For any real number r , we have

$$\hat{\hat{F}}(r) = \sum_{i=1}^n \frac{\Delta(E_{\mathbf{M}i} \leq r)}{\hat{N}\pi_i}, \quad F(r) = \frac{1}{N} \sum_{i=1}^N \Delta(E_{\mathbf{M}i} \leq r), \quad \text{where } E_{\mathbf{M}i} = y_i - \mathbf{x}'_i \mathbf{B}_{\mathbf{M}}.$$

The LM-estimator can be written as follows,

$$\hat{\hat{Y}}_{\text{LM}t} = \bar{\mathbf{X}}^\top \hat{\mathbf{B}}_{\mathbf{M}} + w_n T(\hat{\hat{F}}) + (1 - w_n) \Lambda(\hat{\hat{F}}),$$

where T and Λ are functionals respectively given for any distribution F by

$$T(F) = \int r dF(r) \text{ and } \Lambda(F) = \int r J_n[F(r)] dF(r).$$

$T(\hat{\hat{F}})$ is already linear in r . Therefore we only have to linearize $\bar{\mathbf{X}}^\top \hat{\mathbf{B}}_{\mathbf{M}}$ and $\Lambda(\hat{\hat{F}})$.

But under the conditions of theorem 3.5 we have that

$$\bar{\mathbf{X}}^\top \hat{\mathbf{B}}_{\mathbf{M}} = \bar{\mathbf{X}}^\top \mathbf{B}_{\mathbf{M}} + \sum_{i=1}^n \frac{d_i \bar{\mathbf{X}}^\top \mathbf{M}^{-1} \mathbf{x}_i}{N \sigma_i \pi_i} \psi_H(E_{\mathbf{M}i}^*) + O_p(1/n). \quad (3.15)$$

where,

$$\mathbf{M} = \frac{1}{N} \sum_{i=1}^N \sigma_i^{-2} \psi'_H(E_{\mathbf{M}i}^*) \mathbf{x}_i \mathbf{x}_i' \text{ and } E_{\mathbf{M}i}^* = E_{\mathbf{M}i} / (d_i \sigma_i).$$

Our main problem will be the linearization of $\Lambda(\hat{\hat{F}})$. We start by proving some lemmas.

Lemma 3.6. *Suppose that*

$$C1 \max_{1 \leq k \leq N_t} \|\mathbf{x}_k\| = O(1).$$

Then

$$\begin{aligned} \hat{\hat{F}}_t(E_{\mathbf{M}i} - 2 \max_{1 \leq k \leq N_t} \|\mathbf{x}_k\| \|\hat{\mathbf{B}}_{\mathbf{M}} - \mathbf{B}_{\mathbf{M}}\|) &\leq \hat{\hat{F}}(e_{\mathbf{M}i}) \leq \\ &\leq \hat{\hat{F}}_t(E_{\mathbf{M}i} + 2 \max_{1 \leq k \leq N_t} \|\mathbf{x}_k\| \|\hat{\mathbf{B}}_{\mathbf{M}} - \mathbf{B}_{\mathbf{M}}\|). \end{aligned}$$

Proof of lemma 3.6 .

For any robust residual $e_{\mathbf{M}i}$ we have

$$\begin{aligned} \hat{\hat{F}}(e_{\mathbf{M}i}) &= \sum_{k=1}^n \frac{\Delta(e_{\mathbf{M}k} \leq e_{\mathbf{M}i})}{\hat{N} \pi_k}, \\ &= \sum_{i=1}^n \frac{\Delta[E_{\mathbf{M}k} \leq E_{\mathbf{M}i} + (\mathbf{x}_k - \mathbf{x}_i)'(\hat{\mathbf{B}}_{\mathbf{M}} - \mathbf{B}_{\mathbf{M}})]}{\hat{N} \pi_k}. \end{aligned}$$

But it is clear that

$$-\left(2 \max_{1 \leq k \leq N} \|\mathbf{x}_k\|\right) \|\hat{\mathbf{B}}_{\mathbf{M}} - \mathbf{B}_{\mathbf{M}}\| \leq (\mathbf{x}_k - \mathbf{x}_i)' (\hat{\mathbf{B}}_{\mathbf{M}} - \mathbf{B}_{\mathbf{M}}) \leq \left(2 \max_{1 \leq k \leq N} \|\mathbf{x}_k\|\right) \|\hat{\mathbf{B}}_{\mathbf{M}} - \mathbf{B}_{\mathbf{M}}\|$$

■

Lemma 3.7. Suppose that A1 and A2 are satisfied as well as the following two conditions.

C1 The sequence of distribution functions $\{F_t\}_{t \geq 1}$ is tight.

C2 There exist two constants c_1 and c_2 such that for all $t \geq 1$ we have $F_t(c_1) < 1/N_t$ and $F_t(c_2) > 1 - 1/N_t$.

Then the following expansion holds.

$$\Lambda(\hat{F}_t) = \Lambda(F_t) + \hat{Z}_t + o_p(n_t^{-1/2}),$$

where $\hat{Z}_t = \sum_{i=1}^{n_t} \frac{Z_i}{\hat{N}_t \pi_i}$ and $Z_i = - \int [\Delta(E_{\mathbf{M}i} \leq r) - F_t(r)] J_n[F_t(r)] dr$.

The proof of this lemma is quite similar to that of the common mean superpopulation model and is omitted here.

Theorem 3.7. Assume all the conditions of lemmas 3.6 and 3.7 as well as the following:

C1) There exists $\epsilon_0 > 0$ such that

$$\max_{1 \leq k \leq N_t} P\left(\hat{F}_t(E_{\mathbf{M}i} + \epsilon_0) - \hat{F}_t(E_{\mathbf{M}i} - \epsilon_0) > 0\right) = O(1/n_t).$$

Then the following expansion holds:

$$\widehat{Y}_{\text{LM}t} = \left[\overline{\mathbf{X}}^\top \mathbf{B}_M + (1 - w_n) \Lambda(F_t) \right] + \sum_{k=1}^{n_t} \frac{w_n E_{\text{M}k} + (1 - w_n) Z_k}{\widehat{N} \pi_k} + o_p(n_t^{-1/2}).$$

Proof of theorem 3.7 .

By definition we have

$$\widehat{Y}_{\text{LM}t} = \overline{\mathbf{X}}^\top \widehat{\mathbf{B}}_M + w_n T(\widehat{F}) + (1 - w_n) \Lambda(\widehat{F}). \quad (3.16)$$

The first term on the right-hand side admits the expansion given by the equation (3.15). Concerning the second term, we have

$$\begin{aligned} T(\widehat{F}) &= \sum_{i=1}^{n_t} \frac{e_{\text{M}i}}{\widehat{N}_t \pi_i}, \\ &= \sum_{i=1}^{n_t} \frac{E_{\text{M}i}}{\widehat{N}_t \pi_i} - \overline{\mathbf{X}}^\top (\widehat{\mathbf{B}}_M - \mathbf{B}_M) + \left(\overline{\mathbf{X}} - \sum_{i=1}^{n_t} \frac{\mathbf{x}_i}{\widehat{N}_t \pi_i} \right)^\top (\widehat{\mathbf{B}}_M - \mathbf{B}_M). \end{aligned}$$

It follows from the assumptions A1, A2 and from theorem 3.3 that the order of magnitude of the third term is $O_p(n_t^{-1})$. Therefore

$$T(\widehat{F}) = \sum_{i=1}^{n_t} \frac{E_{\text{M}i} - Q_{\text{M}i}}{\widehat{N}_t \pi_i} + O_p(1/n_t), \text{ where } Q_{\text{M}i} = d_i(\overline{\mathbf{X}}^\top \mathbf{M}^{-1} \mathbf{x}_i / \sigma_i) \psi_H(E_{\text{M}i}^*)$$

The third term on the right-hand side of equation 3.16 can be transformed as follows

$$\begin{aligned} \Lambda(\widehat{F}) &= \sum_{i=1}^{n_t} \frac{\hat{J}_{ni} e_{\text{M}i}}{\widehat{N}_t \pi_i} = \sum_{i=1}^{n_t} \frac{\hat{J}_{ni} [E_{\text{M}i} - \mathbf{x}_i' (\widehat{\mathbf{B}}_M - \mathbf{B}_M)]}{\widehat{N}_t \pi_i}, \\ &= \sum_{i=1}^{n_t} \frac{\hat{J}_{ni} E_{\text{M}i}}{\widehat{N}_t \pi_i} + \sum_{i=1}^{n_t} \frac{\hat{J}_{ni} - \hat{J}_{ni}}{\widehat{N}_t \pi_i} E_{\text{M}i} - \overline{\mathbf{X}}^\top (\widehat{\mathbf{B}}_M - \mathbf{B}_M) + \\ &\quad + \left(\sum_{i=1}^{N_t} \frac{\mathbf{x}_i}{N_t} - \sum_{i=1}^{n_t} \frac{\hat{J}_{ni} \mathbf{x}_i}{\widehat{N}_t \pi_i} \right)' (\widehat{\mathbf{B}}_M - \mathbf{B}_M). \end{aligned}$$

It is quite straightforward to prove that the stochastic order of the last term in the above inequality is $o_p(1)O_p(n_t^{-1/2}) = o_p(n_t^{-1/2})$. Concerning the second term, we have that

$$E \left| \frac{1}{N_t} \sum_{i=1}^{n_t} (\hat{J}_{ni} - \hat{J}_{ni}) \frac{E_{Mi}}{\pi_i} \right| \leq E \left| \frac{1}{N_t} \sum_{i=1}^{N_t} (\hat{J}_{ni} - \hat{J}_{ni}) \frac{I_{i,t} y_i}{\pi_i} \right| + \\ + \left| \frac{1}{N_t} \sum_{i=1}^{n_t} (\hat{J}_{ni} - \hat{J}_{ni}) \frac{I_{i,t} \mathbf{x}_i' \mathbf{B}_M}{\pi_i} \right|,$$

or

$$E \left| \frac{1}{N_t} \sum_{i=1}^{n_t} (\hat{J}_{ni} - \hat{J}_{ni}) \frac{E_{Mi}}{\pi_i} \right| \leq \left(\min_{1 \leq k \leq N_t} \pi_k \right)^{-1} \left\{ \frac{1}{N_t} \sum_{i=1}^{N_t} y_i E |\hat{J}_{ni} - \hat{J}_{ni}| + \right. \\ \left. + \frac{1}{N_t} \sum_{i=1}^{N_t} |\mathbf{x}_i' \mathbf{B}_M| E |\hat{J}_{ni} - \hat{J}_{ni}| \right\}.$$

Note that

$$E |\hat{J}_{ni} - \hat{J}_{ni}| \leq [1 - (\lambda + 1)\alpha_n]^{-1} P \left(|\hat{F}_t(e_{Mi}) - \hat{F}_t(E_{Mi})| > 0 \right), \\ \leq [1 - (\lambda + 1)\alpha_n]^{-1} P \left[\hat{F}_t \left(E_{Mi} + 2 \max_{1 \leq k \leq N_t} \|\mathbf{x}_k\| \|\hat{\mathbf{B}}_M - \mathbf{B}_M\| \right) - \right. \\ \left. - \hat{F}_t \left(E_{Mi} - 2 \max_{1 \leq k \leq N_t} \|\mathbf{x}_k\| \|\hat{\mathbf{B}}_M - \mathbf{B}_M\| \right) > 0 \right], \\ \leq P \left(\hat{F}_t(E_{Mi} + \epsilon_0) - \hat{F}_t(E_{Mi} - \epsilon_0) > 0 \right) + P(\|\hat{\mathbf{B}}_M - \mathbf{B}_M\| > \epsilon_0), \\ \leq O(1/n_t) + \epsilon_0^{-2} E \|\hat{\mathbf{B}}_M - \mathbf{B}_M\|^2 = O(1/n_t).$$

Therefore $E |\hat{J}_{ni} - \hat{J}_{ni}| = O(1/n_t)$ uniformly in i .

■

Theorem 3.7 can now be used to derive the asymptotic variance of the LM-estimator. Let $E_{Mk}^* = w_n E_{Mk} + (1 - w_n) Z_k$ for any population unit k . The asymptotic variance $V_a(\hat{Y}_{LMt})$ of the LM-estimator is given by

$$V_a(\hat{Y}_{LMt}) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) E_{Mi}^* E_{Mj}^*.$$

The bias of the LM-estimator can be approximated by

$$B_a(\widehat{Y}_{LMt}) = -(1 - w_n)(\overline{E}_M - \overline{E}_{\alpha_n}),$$

where $\overline{E}_{\alpha_n} = \Lambda(\widehat{F})$. An asymptotically unbiased estimator of $V_a(\widehat{Y}_{LMt})$ is given by

$$\widehat{V}_a(\widehat{Y}_{LMt}) = -\frac{1}{2N^2} \sum_{i=1}^n \sum_{j=1}^n \left(1 - \pi_i \pi_j / \pi_{ij}\right) \left(\frac{e_{Mi}^*}{\pi_i} - \frac{e_{Mj}^*}{\pi_j}\right)^2,$$

where $e_{Mi}^* = w_n e_{Mi} + (1 - w_n) z_i$, and $e_{Mi} = y_i - \mathbf{x}_i' \widehat{\mathbf{B}}_M$. z_i is obtained from Z_i by replacing the E_{Mi} 's with the e_{Mi} 's and the population distribution F with $\widehat{F}$.

For a consistent estimator of $MSE(\widehat{Y}_{LMt})$, we can use the proposal suggested for the common mean superpopulation model.

In section 3.6, the general robust estimators are studied for some well-known models and sampling designs. The two superpopulation models that will be studied are the ratio model and the simple regression model. For each model the study will involve two sampling designs: Simple random sampling (SRS) and stratified sampling (STR) where the sample units are selected within strata following the SRS design. These sections are justified by the fact that in practice the inclusion probabilities are not calculated very often.

3.6. CM- and LM-estimators under the Ratio Model

An important superpopulation model that is popular in practice is the ratio model given by

$$(\xi_r) \begin{cases} y_i = \beta x_i + \epsilon_i, \\ V_{\xi_r}(\epsilon_i) = \sigma^2 x_i. \end{cases}$$

It is a well-known fact in the model-based framework that the ratio estimator is the best linear unbiased estimator of the population mean under this model and is also identical to the regression estimator generated by (ξ_r) . This estimator has some limited robustness properties. In this section we will focus on ratio models with a more general variance structure given by

$$(\xi_v) \begin{cases} y_i = \beta x_i + \epsilon_i, \\ V_{\xi_v}(\epsilon_i) = \sigma^2 v_i. \end{cases}$$

3.6.1. SRS Design.

It is assumed throughout this section that the sample units are selected without replacement according to the simple random sampling design. Each population unit has a positive probability of being selected.

3.6.1.1. CM-Estimator

The CM-estimator is derived under the standard ratio model where $v_i = x_i$ for all $i = 1, \dots, N$. Therefore, it is assumed throughout this sub-section that the standard ratio model provides an acceptable description of the population under study. The GREG estimator of the population mean takes its cosmetic form under the name ratio estimator and is given by

$$\bar{y}_R = \bar{X} \left(\frac{\bar{y}}{\bar{x}} \right),$$

where

$$\bar{y} = \sum_{k=1}^n y_i/n, \quad \bar{x} = \sum_{k=1}^n x_i/n.$$

The CM-estimator of $\bar{Y}$ generated by this model is given by

$$\bar{y}_{\text{CM}} = \bar{X} \hat{B}_{\text{CM}} \text{ where } \begin{cases} \sum_{i=1}^n d_i \sqrt{x_i} \psi_{\bullet H} \left(\frac{y_i - \hat{B}_{\text{CM}} x_i}{\sigma d_i \sqrt{x_i}} \right) = 0, \\ d_i = \min \left(1, c_0 \frac{\sqrt{\bar{X}_2}}{x_i} \right), \end{cases}$$

$c_0^2 = 1 + \chi_{1,0.95}^2$ and $\bar{X}_2$ is the population mean of the x_i^2 . This estimator is close at least in spirit to the Gwet-Rivest M-estimator. The d -factor proposed by Gwet & Rivest (1992) is given by $d_i = (\bar{X}/x_i)^{1/2}$. This choice is good so long as the smallest x_i is not too small. In fact, a very small value for x_i is likely to increase d_i dramatically which is not the case with our proposal.

The second important difference with the Gwet-Rivest estimator is the ψ -function used. We have

$$\psi_{\bullet H}(r) = w_n r + (1 - w_n) \psi_H(r, K_n),$$

It follows from the equation 3.13 that the asymptotic variance of this estimator is given by

$$AV(\bar{y}_{\text{CM}}) = \frac{1-f}{n(N-1)} \left(\frac{\bar{X}}{\bar{X}_e} \right)^2 \sum_{i=1}^N \left[\sigma \sqrt{x_i} d_i \psi_{\bullet H}(E_{\text{CM},i}^*) \right]^2$$

where

$$\bar{X}_e = \frac{1}{N} \sum_{k=1}^N \psi'_{\bullet H}(E_{\text{CM},k}^*) x_k.$$

The form of this asymptotic variance is very close to that of the Gwet-Rivest estimator except the d -factors and the ψ -functions that are different.

The variance estimator in this case is given by

$$\text{av}(\bar{y}_{\text{CM}}) = \frac{1-f}{n(n-1)} \left(\frac{\bar{X}}{\bar{x}_e} \right)^2 \sum_{i=1}^N \left[\hat{S}_{\text{CM}} \sqrt{x_i} d_i \psi_{\bullet H}(e_{\text{CM},i}^*) \right]^2$$

where

$$\bar{x}_e = \frac{1}{N} \sum_{k=1}^N \psi'_{\bullet H}(e_{CM,k}^*) x_k \text{ and } e_{CM,k}^* = (\hat{S}_{CM} d_k \sqrt{x_k})^{-1} e_{CM,k}.$$

The bias of the CM-estimator is given by

$$B_a(\bar{y}_{CM}) = -\bar{X} \left(\frac{\bar{Y}}{\bar{X}} - B_{CM} \right).$$

For small samples, the difference $\bar{Y}/\bar{X} - B_{CM}$ might not be negligible. Therefore, it is important to use the estimator for the MSE we have proposed instead of using the variance only.

3.6.1.2. The LM-estimator

Now let us assume the more general ratio model (ξ_V) . The GREG estimator derived from this model is given by

$$\begin{aligned} \bar{y}_g &= \bar{y} + (\bar{X} - \bar{x}) \hat{B} \text{ where } \hat{B} = \frac{\frac{1}{n} \sum_{k=1}^n (x_k/v_k) y_k}{\frac{1}{n} \sum_{k=1}^n x_k^2/v_k}, \\ &= \frac{1}{n} \sum_{k=1}^n g_k y_k \text{ where } g_k = 1 + (\bar{X} - \bar{x}) \left(\frac{1}{n} \sum_{k=1}^n x_k^2/v_k \right)^{-1} x_k/v_k. \end{aligned}$$

The robust alternative to this estimator is the LM-estimator. This estimator is now defined. Let $\hat{B}_m$ be implicitly defined as solution to the following equation:

$$\begin{cases} \sum_{i=1}^n \frac{d_i x_i}{\sqrt{v_i}} \psi_H \left(\frac{y_i - \hat{B}_m x_i}{\hat{\sigma}_M d_i \sqrt{v_i}} \right) = 0, \\ \hat{\sigma}_M = (1/\Phi^{-1}(0.75)) \text{ med}_{1 \leq k \leq n} |y_k - \hat{B}_m x_k|. \end{cases}$$

the d_i 's being defined as previously. The LM-estimator of $\bar{Y}$ is then given by

$$\hat{\bar{Y}}_{\text{LM}} = \bar{X}\hat{B}_M + \frac{1}{n} \sum_{i=1}^n r_{Mi}$$

where r_{Mi} is defined as follows:

$$r_{Mi} = \begin{cases} w_n e_{Mi} + (1 - w_n) \frac{e_{Mi}}{1 - (\lambda + 1)\alpha_n} & \text{if } e_{m(\lambda)} < e_{Mi} \leq e_{m(n-1)}, \\ w_n e_{Mi} & \text{otherwise,} \end{cases}$$

and $e_{Mi} = y_i - \hat{B}_M x_i$. Under the simple random sampling design the choice $w_n = n/N$ is likely to be satisfactory.

The asymptotic variance of this estimator is given by

$$V_a(\hat{\bar{Y}}_{\text{LM}}) = \frac{1-f}{n} S_{\text{LM}}^2 \text{ where } \begin{cases} S_{\text{LM}}^2 = \frac{1}{N-1} \sum_{k=1}^N (E_{Mk}^* - \bar{E}_M^*)^2, \\ E_{Mk}^* = w_n E_{Mk} + (1 - w_n) Z_k, \quad E_{Mk} = y_k - B_M x_k. \end{cases}$$

Note that $\bar{E}_M^* = w_n \bar{E}_M$ and if for example $\lambda = 1$ in the definition of the LM-estimator, $Z_k = -\bar{E}_{M(1/n)} + (1 - 2/n)^{-1} [U_i - (E_{M(1)} + E_{M(n-1)})/n]$ and $U_i = E_{Mi} \Delta(E_{Mi} \leq E_{M(n-1)}) + E_{M(n-1)} I(E_{Mi} > E_{M(n-1)})$. For most populations where the number of outliers remains low, the variance of the LM-estimator and that of the CM-estimator will be fairly close.

The bias of the of the LM-estimator is given by

$$B_a(\hat{\bar{Y}}_{\text{LM}}) = -(1 - w_n)(\bar{E}_M - \bar{E}_{m(1/n)}) \text{ where } \bar{E}_M = \frac{1}{N} \sum_{k=1}^N E_{Mk},$$

and $\bar{E}_{m(1/n)}$ is the $1/n$ -trimmed mean of the E_{Mk} 's. If the distribution of the residuals is heavily skewed this bias might be substantial for small samples. However, if the model fits the population units reasonably well the bias will be small.

3.6.2. Stratified Random Sampling

It is assumed that the population is divided into L nonoverlapping strata. Each stratum h comprises N_h population units. A simple random sample of size n_h is drawn without replacement within each stratum h .

3.6.2.1. CM-estimator

The CM-estimator of the population mean under the stratified random sampling design is given by

$$\bar{y}_{\text{CM}} = \bar{X} \hat{B}_{\text{ST}}$$

where

$$\sum_{h=1}^L \frac{N_h}{n_h} \sum_{i=1}^{n_h} d_{hi} \sqrt{x_{hi}} \psi_{\bullet H} \left(\frac{y_{hi} - \hat{B}_{\text{ST}} x_{hi}}{\sigma d_{hi} \sqrt{x_{hi}}} \right) = 0.$$

It is clear from this equation that if one stratum is much larger than the others it will have more influence on the robust estimator. This type of influence is due to the design and has nothing to do with the population distribution. Our estimators are not designed to reduce this type of influence.

It is quite straightforward to see that the variance of this estimator is given by

$$AV(\bar{y}_{\text{CM}}) = \left(\frac{\bar{X}}{\bar{X}_e} \right)^2 \sum_{h=1}^L W_h^2 \frac{1 - f_h}{n_h} S_{eh}^2.$$

where

$$S_{eh}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (E_{CM,hi}^{**} - \bar{E}_{CM,h}^{**})^2, \quad E_{CM,hi}^{**} = \sigma \sqrt{x_{hi}} d_{hi} \psi_{\bullet H}(E_{CM,hi}^*),$$

$$E_{CM,hi}^* = E_{CM,hi} / (\sigma d_{hi} \sqrt{x_{hi}}), \quad E_{CM,hi} = y_{hi} - B_{CM} x_{hi},$$

$$\text{and } \bar{E}_{CM,h}^{**} = \frac{1}{N_h} \sum_{i=1}^{N_h} E_{CM,hi}^{**}.$$

If the $\psi_{\bullet H}$ function is replaced with the identity function, then we obtain the standard variance of the ratio estimator in stratified random sampling (see Cochran, 1977, page 166). Therefore, the use of the CM-estimator can result in a substantial reduction of the variance. The superiority of this estimator over the ratio estimator depends upon the size of its bias. This bias is given by

$$B_a(\bar{y}_{CM}) = \bar{X}(B_{CM} - \bar{Y}/\bar{X}).$$

The magnitude of this bias depends essentially upon the population structure. For most populations, this bias will be pretty small. A consistent estimator of the variance of $\bar{y}_{CM}$ is given by

$$\text{av}(\bar{y}_{CM}) = \left(\frac{\bar{X}}{\bar{x}_{1e}} \right)^2 \sum_{h=1}^L W_h^2 \frac{1-f_h}{n_h} \hat{S}_{eh}^2$$

where

$$\begin{cases} \hat{S}_{eh}^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (e_{CM,hi}^{**} - \bar{e}_{CM,h}^{**})^2, \\ \bar{x}_{1e} = \sum_{h=1}^L \sum_{k=1}^{n_h} \frac{W_h}{n_h} \psi'_{\bullet H}(e_{CM,hk}^*) x_{hk}. \end{cases}$$

A consistent estimator of the squared bias can be constructed as follows. First, obtain the natural estimator of the bias given by

$$b_a(\bar{y}_{CM}) = - \sum_{h=1}^L \sum_{i=1}^{n_h} \frac{W_h}{n_h} e_{CM,hi}.$$

Then calculate an estimate of the asymptotic variance of this bias using the following formula:

$$\text{av}[b_a(\bar{y}_{CM})] = \sum_{h=1}^L \sum_{i=1}^{n_h} W_h^2 \frac{1-f_h}{n_h} \hat{S}_{1eh}^2, \text{ where } \hat{S}_{1eh}^2 = \frac{1}{n_h-1} \sum_{i=1}^{n_h} (e_{CM,hi} - \bar{e}_{CM,h})^2.$$

A consistent estimator of the squared bias is then given by

$$b_{2a}(\bar{y}_{CM}) = \max\left\{0, b_a^2(\bar{y}_{CM}) - \text{av}[b_a(\bar{y}_{CM})]\right\}.$$

3.6.2.2. LM-Estimator

We start by introducing some notation. Let c_0 be defined by $c_0 = 1/\Phi^{-1}(0.75)$ and the statistic $\hat{B}_{Ms}$ implicitly defined as solution to the following equation:

$$\sum_{h=1}^L \frac{N_h}{n_h} \sum_{i=1}^{n_h} d_{hi} \sqrt{x_{hi}} \psi_H \left(\frac{y_{hi} - \hat{B}_{Ms} x_{hi}}{\hat{\sigma}_M d_{hi} \sqrt{x_{hi}}} \right) = 0, \quad (3.17)$$

where $\hat{\sigma}_M = c_0 \text{Inf}\{|e_{Mi}|; \hat{F}(|e_{Mi}|) \geq 1/2\}$ and $\hat{F}$ is the empirical distribution function of the $|e_{Mk}|$'s given by

$$\hat{F}(r) = \sum_{h=1}^L W_h \hat{F}_h(r) \text{ where } \hat{F}_h(r) = \frac{1}{n_h} \sum_{k=1}^{n_h} \Delta(|e_{Mk}| \leq r) \quad W_h = N_h/N.$$

The LM-estimator of the population mean is given by

$$\hat{Y}_{LM} = \bar{X} \hat{B}_{Ms} + \sum_{h=1}^L W_h \bar{r}_{Mh}$$

where

$$\bar{r}_{Mh} = \frac{1}{n_h} \sum_{i=1}^{n_h} r_{Mhi},$$

and r_{mhi} is given by $r_{mhi} = [w_n + (1 - w_n)\lambda_{hi}]e_{mhi}$, and

$$\lambda_{hi} = J_n[\hat{G}(e_{mhi})] \text{ where } \hat{G}(r) = \sum_{h=1}^L W_h \hat{G}_h(r) \text{ and } \hat{G}_h(r) = \frac{1}{n_h} \sum_{i=1}^{n_h} \Delta(e_{mhi} \leq r).$$

The variance of the LM-estimator under the stratified random sampling design is given by

$$V_a(\hat{Y}_{LM}) = \sum_{h=1}^L W_h^2 \frac{1 - f_h}{n_h} S_{LMh}^2 \text{ where } \begin{cases} S_{LMh}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} U_{hi}^2, \\ U_{hi} = w_n(E_{mhi} - \bar{E}_{mh}) + (1 - w_n)Z_{hi}. \end{cases}$$

Note that for the GREG estimator, the variance has a similar form except that the U_{hi} is given by

$$U_{gi}^{(g)} = E_{ghi} - \bar{E}_{gh}$$

where

$$\begin{cases} E_{ghi} = y_{hi} - B_G x_{hi}, \\ B_G = \frac{\sum_{h=1}^L W_h \left(\frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi} y_{hi} / v_i \right)}{\sum_{h=1}^L W_h \left(\frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi}^2 / v_i \right)} \end{cases}$$

It follows from this that the variance of the LM-estimator will be smaller than that of the GREG estimator for many populations, except perhaps if one decides to take $w_n = 1$.

The variance of the CM-estimator has also a similar form. But the corresponding U_{hi} is given by

$$U_{hi}^{(c)} = w_n E_{cmhi} + (1 - w_n) R_{hi} \text{ where } R_{hi} = \sigma \sqrt{x_{hi}} d_{hi} \psi_H(E_{cmhi} / \sigma \sqrt{x_{hi}} d_{hi}).$$

and $E_{CMhi} = y_{hi} - B_{CM}x_{hi}$. If the standard ratio model holds, the variances of the CM-estimator and that of the LM-estimator are likely to be very close. The LM-estimator is expected to be better if the model ξ_v is more accurate.

3.7. CM- and LM-estimators under the Simple Regression Model and the Simple Random Sampling Design.

The simple regression often used to describe the population is given by

$$(\xi_{rg}) \begin{cases} y_k = \alpha + \beta x_k + \epsilon_k, \\ V_{\xi_{rg}}(\epsilon_k) = \sigma^2 \eta_k. \end{cases}$$

Note that very often $\eta_k = 1$ for all $k = 1, \dots, N$. But it is quite possible for some populations that η_k is different from 1 at least for some units. In such a case, the CM-estimator will not be consistent and the LM-estimator should be used. We will confine ourselves to the simple random sampling design without replacement. The general theory developed previously can be used to derive the robust estimators for some other interesting designs.

3.7.1. CM-Estimator

Let us assume at the outset the common case where $\eta_k = 1$. The GREG estimator generated by that model is given by

$$\bar{y}_G = \widehat{B}\bar{X} + \bar{e}_G$$

where

$$\begin{cases} \bar{e}_G = \frac{1}{n} \sum_{i=1}^n \bar{e}_{G,i}, \\ \bar{e}_{G,i} = y_i - \hat{B}_G x_i, \end{cases}$$

and the statistic $\hat{B}_G$ is given by,

$$\hat{B}_G = \frac{\frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})(y_k - \bar{y})}{\frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2}.$$

The CM-estimator generated by this simple regression model where $\eta_k = 1$, is given by $\bar{y}_{CM} = a_{CM} + \hat{B}_{CM} \bar{X}$, where a_{CM} and $\hat{B}_{CM}$ are implicitly defined by the following system of equations:

$$\begin{cases} \sum_{k=1}^n d_k \psi_{\bullet H} \left(\frac{y_k - (a_{CM} + \hat{B}_{CM} x_k)}{d_k \hat{\sigma}_c} \right) = 0, \\ \sum_{k=1}^n d_k x_k \psi_{\bullet H} \left(\frac{y_k - (a_{CM} + \hat{B}_{CM} x_k)}{d_k \hat{\sigma}_c} \right) = 0, \end{cases}$$

where $\hat{\sigma}_c = c_0 \text{Inf}\{|e_{CM,i}|; \hat{F}(|e_{CM,i}|)/ge1/2\}$,

$$d_k = \min \left(1, \frac{c_0}{\sqrt{1 + (x_k - \bar{X})^2 / S_x^2}} \right) \text{ and } S_x^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{X})^2.$$

The asymptotic variance of this estimator is given by

$$AV(\bar{y}_{CM}) = \frac{1-f}{n(N-1)} \sum_{i=1}^N U_{CM,i}^2 \text{ where } U_{CM,i} = \left(1 - \frac{(x_i - \bar{X}_\psi)(\bar{X}_\psi - \bar{X})}{S_\psi^2} \right) \frac{E_{CM,i}^+}{\bar{\Psi}_{1H}}.$$

The residuals involved here are given by

$$E_{CM,i}^+ = \hat{\sigma}_c d_i \psi_{\bullet H}(E_{CM,i}^{(s)}), \quad E_{CM,i}^{(s)} = [y_i - (A_{CM} + B_{CM} x_i)] / (d_i \sigma).$$

and we have $S_\psi^2 = \bar{X}_{2\psi} - \bar{X}_\psi^2$ where

$$\bar{X}_\psi = \frac{1}{N\bar{\Psi}_{1H}} \sum_{i=1}^N \bar{\psi}'_f(E_{\text{CM}i}^{(s)})x_i,$$

$$\bar{X}_{2\psi} = \frac{1}{N\bar{\Psi}_{1H}} \sum_{i=1}^N \bar{\psi}'_f(E_{\text{CM}i}^{(s)})x_i^2, \text{ and } \bar{\Psi}_{1H} = \frac{1}{N} \sum_{i=1}^N \bar{\psi}'_f(E_{\text{CM}i}^{(s)})x_i.$$

A consistent estimator of this asymptotic variance is given by

$$\widehat{AV}(\bar{y}_{\text{CM}}) = \frac{1-f}{n(n-1)} \sum_{i=1}^n u_{\text{CM},i}^2 \text{ where } u_{\text{CM},i} = \left(1 - \frac{(x_i - \bar{x}_\psi)(\bar{x}_\psi - \bar{X})}{\widehat{S}_\psi^2}\right) \frac{e_{\text{CM},i}^+}{\widehat{\Psi}_{1H}}.$$

We have

$$e_{\text{CM},i}^{(s)} = [y_i - (a_{\text{CM}} + \widehat{B}_{\text{CM}}x_i)] / (d_i \widehat{\sigma}_c), \quad e_{\text{CM},i}^+ = \widehat{\sigma}_c d_i \psi_{\bullet H}(e_{\text{CM},i}^{(s)}),$$

$$\bar{x}_\psi = \frac{1}{n\widehat{\Psi}_{1H}} \sum_{i=1}^n \bar{\psi}'_f(e_{\text{CM},i}^{(s)})x_i, \quad \widehat{S}_\psi^2 = \frac{1}{n\widehat{\Psi}_{1H}} \sum_{i=1}^n \bar{\psi}'_f(e_{\text{CM},i}^{(s)})(x_i - \bar{x}_\psi)^2.$$

It is important to stress at this stage that if the ψ -function is replaced with the identity function throughout, the variance estimator proposed above becomes

$$v_0 = \frac{1-f}{n(n-1)} \sum_{i=1}^n u_{Gi}^2$$

where

$$u_{Gi} = \left(1 - \frac{(x_i - \bar{x})(\bar{x} - \bar{X})}{S_x^2}\right) \bar{e}_{G,i},$$

which corresponds to the well-known estimator proposed by Särndal et al. (1989).

It is obtained by replacing the regression residuals with the residuals multiplied by the g -weight.

3.7.2. LM-Estimator

In this section we assume that the η_k 's in model ξ_{rg} are arbitrary but known. This model leads naturally to the LM-estimator. Let (a_M, b_M) be implicitly defined by

$$\begin{cases} \sum_{k=1}^n d_k \eta_k \psi_H \left(\frac{y_k - (a_M + b_M x_k)}{d_k \hat{\sigma}_M / \eta_k} \right) = 0, \\ \sum_{k=1}^n d_k \eta_k x_k \psi_H \left(\frac{y_k - (a_M + b_M x_k)}{d_k \hat{\sigma}_M / \eta_k} \right) = 0, \end{cases}$$

where $\hat{\sigma}_M = c_0 \inf\{|e_{Mi}|; \hat{F}(|e_{Mi}|) \geq 1/2\}$. The LM-estimator of $\bar{Y}$ (assuming for simplicity that $\lambda = 1$) is then given by

$$\hat{\bar{Y}}_{LM} = (a_M + b_M \bar{X}) + \frac{1}{n} \sum_{i=1}^n e_{LM,i},$$

where $e_{LM,i}$ is given by

$$e_{LM,i} = \begin{cases} w_n e_{Mi} + (1 - w_n) \left(\frac{e_{Mi}}{1 - 2\alpha_n} \right) & \text{if } e_{m(1)} < e_{Mi} \leq e_{M(n-1)}, \\ w_n e_{Mi} & \text{otherwise.} \end{cases}$$

We have $e_{Mi} = y_i - (a_M + b_M x_i)$. The asymptotic variance of the LM-estimator is given by

$$V_a(\hat{\bar{Y}}_{LM}) = \frac{1-f}{n(N-1)} \sum_{i=1}^N U_{LM,i}^2, \text{ where } U_{LM,i} = w_n (E_{Mi} - \bar{E}_M) + (1 - w_n) Z_i,$$

and Z_i is given by

$$\begin{cases} Z_i = -\bar{E}_{m(1/n)} + (1 - 2/n)^{-1} [H_i - (E_{m(1)} + E_{M(n-1)})/n], \\ H_i = E_{Mi} \Delta(E_{Mi} \leq E_{M(n-1)}) + E_{M(n-1)} \Delta(E_{Mi} > E_{M(n-1)}). \end{cases}$$

Note that for large samples, the asymptotic variance of the CM-estimator is approximately equal to

$$V_a(\bar{y}_{CM}) = \frac{1-f}{n(N-1)} \sum_{i=1}^N U_{CM,i}^2$$

where

$$U_{\text{CM},i} = w_n E_{\text{CM},i} + (1 - w_n) \hat{\sigma}_c d_{\mathbf{i}} \psi_H(E_{\text{CM},i}^{(s)}).$$

Because of the term $E_{\text{M},i} - \bar{E}_{\text{M}}$ in $U_{\text{LM},i}$, the variance of the LM-estimator is likely to be higher than that of the CM-estimator for many populations. But that larger variance will very often be offset by a smaller bias. However if the η_k 's are fairly variable, then the variance of the LM-estimator might be smaller.

CHAPTER 4

Jackknife Variance Estimation for the CM and LM Estimators

4.1. Introduction

The CM- and LM-estimators have been introduced and studied in chapters 2 and 3. The estimation of the mean squared error was carried out using a Taylor series linear representation of the corresponding estimators. In general, such an approach leads to a formula for the variance estimator that is valid for large sample sizes.

The simulation results, reported in chapter 2, indicate that the performance of the methods can be satisfactory for estimating the variance of the CM- and the LM-estimators. However, these methods exploit the linear representations of the CM- and LM-estimators. Consequently modifying these estimators even slightly will require the derivation of new linear representations of the estimators and new

variance formulas. Moreover, the implementation of the variance formulas can be time-consuming in terms of programming.

Some additional problems incidental to the Taylor-series-based approach are worth mentioning. The linear representation of the CM-estimator of $\bar{Y}$ is obtained under the assumption of a known scale parameter σ . The impact of such an assumption on the accuracy of the estimator is not well-studied. Concerning the LM-estimator, we had to assume that α_n - the proportion of LM-residuals to be trimmed - is constant. This may not always be true for complex designs. Another difficulty encountered in the estimation of the LM-estimator variance is the calculation of the Z_i -values. If the population mean of the predicted errors is estimated by means of an L-estimator other than the trimmed mean, then the calculation of the Z_i -values will become very difficult, making the variance estimation almost impossible.

The problem of estimating the variance of complex estimators is classical in statistics. Resampling methods have gained in popularity in recent years as a means to avoid tedious derivations. The basic idea underlying these methods is to draw many subsamples from the original sample and then use the cheap computing time nowadays to apply the estimation procedure repeatedly on each subsample capturing thereby the variability of the estimator. Among the most popular resampling methods are the jackknife with all its variants, the bootstrap with all its variants and the balanced repeated replication (BRR). The Jackknife method was introduced within the parametric statistical framework by Quenouille (1949,1956) as a

means of reducing the bias of some estimators. The extension of this technique to the problem of variance estimation was proposed by Tukey (1958). The BRR method on the other hand was proposed by McCarthy (1966), while Efron(1979) introduced the bootstrap method as a means to overcome the poor performance of the jackknife method for nonsmooth statistics.

Our goal in this chapter is to show the potential of resampling methods for estimating the variance of the proposed robust estimators. Some limited theoretical results exist on the use of the jackknife for estimating the variance of M-estimators. Reeds (1978) proved under some restrictive conditions on the ψ -function and on the distribution of errors that the delete-one jackknife variance estimator is consistent for estimating the asymptotic variance of the M-estimator. But Reeds's investigation was conducted within the infinite population framework and under the assumption that the observations are independent and identically distributed.

We confine ourselves here to one version of the jackknife, the delete-one jackknife since it is the most popular jackknife method and most jackknife methods developed to date having a theoretical basis, involve the removal of one sample unit at a time. Krewski and Rao (1981) studied theoretically some jackknife variance estimators under the stratified multistage sampling design when the parameter of interest is a function (linear or not) of population means. Shao (1994) also investigated the asymptotic properties of the delete-one jackknife for estimating the variance of some L-estimators under the stratified multistage sampling design.

In section 4.2 we will present a short overview of the jackknife methodology. In section 4.3 the CM- and the LM-estimators are introduced under the Simple Random Sampling Without Replacement (SRSWOR) design as well as the corresponding jackknife variance estimators. The asymptotic properties of these jackknife variance estimators are studied in section 4.4 and some simulation results are presented in section 4.5 to assess the performance of the jackknife.

4.2. An Overview of the Jackknife Methodology

Let us assume at the outset that the population is infinite, that θ is the parameter of interest and that a sample of n independent and identically distributed random variables $y_1, \dots, y_n$ is available. An estimator $\hat{\theta}$ of θ using the full sample is also available. Quenouille's estimator of θ based upon $\hat{\theta}$ is given by

$$\hat{\theta}_Q = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_{P,i}$$

where

$$\hat{\theta}_{P,i} = n\hat{\theta} - (n-1)\hat{\theta}_{(-i)},$$

$\hat{\theta}_{(-i)}$ is an estimator of θ calculated the same way as $\hat{\theta}$ but based on the subsample where the i -th unit is removed. The $\hat{\theta}_{P,i}$'s are usually called the pseudo-values. Quenouille's estimator has the property of reducing the bias of many estimators from order $O(1/n)$ to order $O(1/n^2)$.

Tukey (1958) suggested that the pseudovalues can reasonably be considered as being independent and identically distributed. Under the latter assumption, the jackknife

variance estimator of $\hat{\theta}_Q$ is given by

$$\begin{aligned}\hat{V}_{JK}(\hat{\theta}_Q) &= \frac{1}{n(n-1)} \sum_{i=1}^n (\hat{\theta}_{P,i} - \hat{\theta}_Q)^2, \\ &= \frac{n-1}{n} \sum_{i=1}^n (\hat{\theta}_{(-i)} - \hat{\theta}_{(\bullet)})^2,\end{aligned}\tag{4.1}$$

where

$$\hat{\theta}_{(\bullet)} = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_{(-i)}.$$

This estimator is also used in practice as an estimator of the variance of $\hat{\theta}$ and $\hat{\theta}_{(\bullet)}$ is sometimes replaced with $\hat{\theta}$, the resulting loss in efficiency being very small.

A substantial number of theoretical results are now available that support Tukey's conjecture. Miller (1964) proves that the jackknife variance estimators works for estimating the variance of a function of means. This result generalizes to U -statistics and functions thereof (see Arvesen, 1969). More details about U -statistics can be found in Serfling (1980, chapter 5).

Another important class of statistics where some interesting results about the jackknife variance estimators are available is the functionals. A functional $\hat{\theta}_n$ is an estimator that is defined as a function of the empirical distribution function F_n . Therefore, a functional is of the form $\hat{\theta}_n = T(F_n)$. It will be seen in the context of finite populations that the CM- and LM-estimators introduced in chapters 2 and 3 are either functionals or functions of functionals. Shao and Tu (1995) proved strong consistency of the jackknife variance estimators of a class of functionals within the model-based framework. Shao's proofs involve the concepts of Gâteaux and ρ -Fréchet differentiability (see Shao and Tu, 1995, page 32).

Now suppose that we are dealing with a finite population of size N . The variable of interest is y which can only be observed on sample units. The standard parameter of interest is $\bar{Y} = (1/N)(y_1 + \dots + y_N)$, but can also be a function of the mean. In this context, the number of results available about the jackknife variance estimator is more limited. This is particularly true for nonlinear statistics.

Suppose that the parameter of interest is a function of the mean $g(\bar{Y})$ that is estimated by $g(\bar{y})$ from a simple random sample selected without replacement. The jackknife variance estimator of that estimator is given by

$$\hat{V}_{JK}[g(\bar{y})] = \frac{1-f}{n}(n-1) \sum_{i=1}^n \left[g(\bar{y}_{(-i)}) - \frac{1}{n} \sum_{k=1}^n g(\bar{y}_{(-k)}) \right]^2.$$

It is a well-known fact that if the first-order derivative g' of g exists, is continuous and nonzero in the neighborhood of $\bar{Y}$ then this jackknife variance estimator is consistent. Krewski and Rao (1981) as well as Shao and Tu (1995) proved a similar result assuming a stratified two-stage sampling design which will be discussed later. Another result worth noting is the consistency of the jackknife variance estimator of smooth L-statistics in complex survey problems proved by Shao (1994). Some of his findings will be used here for proving the consistency of the jackknife estimator of the LM-estimator asymptotic variance.

There are no large sample results available for finite populations that can be used to justify jackknifing the CM- or the LM- estimators as introduced in chapter 2. Therefore, we start in section 4.3 with a simple case where the sample of size n is selected without replacement according to the simple random sampling (SRS-

WOR) design, from a finite population of size N . In section 4.4, some large sample results are discussed. A simulation study is presented in section 4.5.

We assume throughout this chapter that the parameter of interest is the population mean $\bar{Y}$ given by

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N y_i.$$

4.3. Jackknife Variance Estimation Under the SRS-WOR Design

The CM and the LM estimators will be treated separately because different techniques are needed for proving the asymptotic results. In each case we recall the definition of the estimator followed by the proposed jackknife estimator for estimating the asymptotic variance. An estimator of the mean squared error (MSE) is also proposed and the consistency of these estimators proved.

4.3.1. Jackknifing the CM-Estimator

We recall briefly the definition of the CM-estimator (see chapter 2 for more details).

Let $\psi_{\bullet H}$ be the function defined for all real number t by $\psi_{\bullet H}(t) = w_n t + (1 - w_n)\psi_H(t, K_n)$. The CM-estimator is given by $\hat{\bar{Y}}_{\text{CM}} = \hat{B}_{\text{CM}}$ where $\hat{B}_{\text{CM}}$ is implicitly

defined by

$$\sum_{i=1}^n \psi_{\bullet_H} \left(\frac{y_i - \hat{B}_{CM}}{\hat{\sigma}_{CM}} \right) = 0,$$

where for some $\lambda_0 > 0$ and $e_{CM,i} = y_i - \hat{B}_{CM}$,

$$\hat{\sigma}_{CM} = \lambda_0 \inf\{|e_{CM,i}| : \hat{F}(|e_{CM,i}|) \geq 1/2\}$$

and

$$\hat{F}(r) = \frac{1}{n} \sum_{i=1}^n I(|e_{CM,i}| \leq r).$$

Now, we like to explain the choice of both $\hat{Y}_{CM}$ and $\hat{B}_{CM}$ to denote the CM-estimator.

For a fixed sample size n , let us define the population parameter B_{CM} by the following equation

$$\sum_{i=1}^N \psi_{\bullet_H} \left(\frac{y_i - B_{CM}}{\sigma_{CM}} \right) = 0,$$

where for some $\lambda_0 > 0$ and $E_{CM,i} = y_i - B_{CM}$,

$$\sigma_{CM} = \lambda_0 \inf\{|E_{CM,i}| : F(|E_{CM,i}|) \geq 1/2\}$$

and

$$F(r) = \frac{1}{N} \sum_{i=1}^N I(|E_{CM,i}| \leq r). \quad (4.2)$$

When $n \ll N$ and n is small to moderate, the two parameters B_{CM} and $\bar{Y}$ are different and that difference vanishes as n increases. The CM-estimator we have proposed can be primarily seen as estimating B_{CM} and is denoted as $\hat{B}_{CM}$. But it is also used as an estimator of $\bar{Y}$. For that reason it is sometimes denoted as $\hat{Y}_{CM}$.

The jackknife estimator of the variance of $\widehat{Y}_{\text{CM}} = \widehat{B}_{\text{CM}}$ is given by

$$\begin{aligned}\widehat{V}_{\text{JK}}(\widehat{B}_{\text{CM}}) &= \frac{1-f}{n}(n-1) \sum_{i=1}^n \left(\widehat{B}_{\text{CM}(-i)} - \widehat{B}_{\text{CM}(\bullet)} \right)^2, \\ &= \widehat{V}_{\text{JK}}(\widehat{Y}_{\text{CM}})\end{aligned}\tag{4.3}$$

where $\widehat{B}_{\text{CM}(-i)}$ is implicitly defined by the following equation

$$\sum_{k \neq i}^n \psi_{\bullet H} \left(\frac{y_k - \widehat{B}_{\text{CM}(-i)}}{\widehat{\sigma}_{\text{CM}(-i)}} \right) = 0,\tag{4.4}$$

and

$$\widehat{B}_{\text{CM}(\bullet)} = \frac{1}{n} \sum_{i=1}^n \widehat{B}_{\text{CM}(-i)}.$$

Note that the statistic $\widehat{\sigma}_{\text{CM}(-i)}$ is defined by

$$\widehat{\sigma}_{\text{CM}(-i)} = \lambda_0 \inf \left\{ |e_{\text{CM},k}^{(-i)}| : \widehat{F}_{(-i)}(|e_{\text{CM},k}^{(-i)}|) \geq 1/2 \right\},\tag{4.5}$$

where

$$e_{\text{CM},k}^{(-i)} = y_k - \widehat{B}_{\text{CM}(-i)}$$

and

$$\widehat{F}_{(-i)}(r) = \frac{1}{n-1} \sum_{k \neq i}^n I(|e_{\text{CM},k}^{(-i)}| \leq r)\tag{4.6}$$

is the empirical distribution of the absolute CM-residuals $|e_{\text{CM},k}^{(-i)}|$'s.

It is important to recalculate the scale estimate $\widehat{\sigma}_{\text{CM}}$ each time a sample unit is removed. Otherwise the jackknife estimator will not capture the extra variability entailed by the estimation of σ_{CM} . The replacement of $\widehat{\sigma}_{\text{CM}}$ with $\widehat{\sigma}_{\text{CM}(-i)}$ is probably what will make the jackknife perform better than the linearization method for finite samples.

The proposed jackknife method for estimating the variance of the CM-estimator is computationally intensive. In fact each time a unit i is deleted from the sample it takes an iterative process to obtain $\hat{B}_{CM(-i)}$ and $\hat{\sigma}_{CM(-i)}$. But this is possible as shown by the Monte Carlo study presented in section 4.5. However, a faster and more affordable method is to obtain these estimates by performing a one-step CM-estimator as follows:

$$\hat{B}_{CM}^{(i)} = \frac{\sum_{k \neq i}^n y_k w\{(y_k - \hat{B}_{CM})/\hat{\sigma}_{CM}^{(i)}\}}{\sum_{k \neq i}^n w\{(y_k - \hat{B}_{CM})/\hat{\sigma}_{CM}^{(i)}\}} \quad (4.7)$$

where $w(\cdot)$ is defined for all real numbers r by

$$w(r) = \begin{cases} 1 & \text{if } r = 0, \\ \psi_H(r)/r & \text{otherwise.} \end{cases}$$

The statistic $\hat{\sigma}_{CM}^{(i)}$ is given by

$$\hat{\sigma}_{CM}^{(i)} = \lambda_0 \inf \left\{ |e_{CM,k}| : \hat{F}^{(i)}(|e_{CM,k}|) \geq 1/2 \right\} \quad (4.8)$$

where $\hat{F}^{(i)}(\cdot)$ is given by

$$\hat{F}^{(i)}(r) = \frac{1}{n-1} \sum_{k \neq i}^n I(|e_{CM,k}| \leq r). \quad (4.9)$$

Note the difference between the two estimators of scale $\hat{\sigma}_{CM}^{(i)}$ given by (4.8) and $\hat{\sigma}_{CM(-i)}$ given by (4.5). The first is calculated from $\hat{B}_{CM}$ while the second is obtained from $\hat{B}_{CM(-i)}$. The same difference is noticed between the two distribution functions $\hat{F}^{(i)}$ and $\hat{F}_{(-i)}$. But the one-step estimator $\hat{B}_{CM}^{(i)}$ defined by (4.7) and its fully iterated version defined by (4.4) will have approximately the same efficiency.

An alternative approach for estimating the variance of the CM-estimator is by the Taylor series method used in chapter 2. It follows from that general theory that the variance estimator based upon the Taylor series method is given by

$$\hat{V}_{TL}(\hat{Y}_{CM}) = \frac{1-f}{n} \frac{1}{n-1} \sum_{i=1}^n \left\{ \frac{\hat{\sigma}_{CM} \psi_{\bullet H}[(y_i - \hat{B}_{CM})/\hat{\sigma}_{CM}]}{\frac{1}{n} \sum_{k=1}^n \psi'_{\bullet H}[(y_k - \hat{B}_{CM})/\hat{\sigma}_{CM}]} \right\}^2.$$

This estimator follows from the formula of the asymptotic variance of the CM-estimator which is given by

$$V_a(\hat{Y}_{CM}) = \frac{1-f}{n} \frac{1}{N-1} \sum_{i=1}^N Q_{CM,i}^2, \quad (4.10)$$

where $Q_{CM,i}$ is given by

$$Q_{CM,i} = \frac{\sigma_{CM} \psi_{\bullet H}[(y_i - B_{CM})/\sigma_{CM}]}{\frac{1}{N} \sum_{k=1}^N \psi'_{\bullet H}[(y_k - B_{CM})/\sigma_{CM}]}. \quad (4.11)$$

Evaluating this formula is much easier once $\hat{\sigma}_{CM}$ and $\hat{B}_{CM}$ are available. But this formula must be derived first and this will sometimes be quite complicated or even impossible. Furthermore this formula was obtained using the fact that $\hat{\sigma}_{CM}$ converges in probability to σ_{CM} and the Slutsky theorem. This will necessarily result in a loss in efficiency for finite samples as shown by the simulation results. The jackknife estimator provides a more efficient way of capturing the variability entailed by $\hat{\sigma}_{CM}$.

For small to moderate sample sizes, the CM-estimator might have a nonnegligible bias. Therefore, it is important to be able to estimate not only the variance but the

mean squared error (MSE) as well. The variance can be estimated by the proposed jackknife estimator. We now propose an estimator of the squared bias.

The squared bias of the CM estimator is given by

$$\begin{aligned} B^2(\widehat{Y}_{\text{CM}}) &= \left[E(\widehat{Y}_{\text{CM}}) - \bar{Y} \right]^2, \\ &\doteq \left[E(\widehat{Y}_{\text{CM}} - \widehat{Y}_{\text{G}}) \right]^2, \\ &= E(\widehat{Y}_{\text{CM}} - \widehat{Y}_{\text{G}})^2 - V(\widehat{Y}_{\text{CM}} - \widehat{Y}_{\text{G}}). \end{aligned}$$

It follows from the linear representation of $\widehat{Y}_{\text{CM}}$ in chapter 2 that the asymptotic variance of $\widehat{Y}_{\text{CM}} - \widehat{Y}_{\text{G}}$ is given by,

$$V_a(\widehat{Y}_{\text{CM}} - \widehat{Y}_{\text{G}}) = \frac{1-f}{n} \frac{1}{N-1} \sum_{i=1}^N (Q_{\text{CM}i} - E_{\text{G}i})^2, \quad (4.12)$$

where $E_{\text{G}i} = y_i - \bar{Y}$. The asymptotic *MSE* of the CM-estimator is then given by

$$MSE_a(\widehat{Y}_{\text{CM}}) = V_a(\widehat{Y}_{\text{CM}}) + B_a(\widehat{Y}_{\text{CM}}) \quad (4.13)$$

where $V_a(\widehat{Y}_{\text{CM}})$ is given by (4.10) and $B_a(\widehat{Y}_{\text{CM}}) = E(\widehat{Y}_{\text{CM}} - \widehat{Y}_{\text{G}})^2 + V_a(\widehat{Y}_{\text{CM}} - \widehat{Y}_{\text{G}})$.

We now propose the following jackknife estimator of the squared bias of the CM-estimator:

$$b_{2\text{JK}}(\widehat{Y}_{\text{CM}}) = (\widehat{Y}_{\text{CM}} - \widehat{Y}_{\text{G}})^2 - \widehat{V}_{\text{JK}}(\widehat{Y}_{\text{CM}} - \widehat{Y}_{\text{G}}),$$

where $\widehat{V}_{\text{JK}}(\widehat{Y}_{\text{CM}} - \widehat{Y}_{\text{G}})$ is given by

$$\begin{aligned} \widehat{V}_{\text{JK}}(\widehat{Y}_{\text{CM}} - \widehat{Y}_{\text{G}}) &= \\ &= \frac{1-f}{n} (n-1) \sum_{i=1}^n \left[(\widehat{Y}_{\text{CM}(-i)} - \widehat{Y}_{\text{CM}(\bullet)}) - (\widehat{Y}_{\text{G}(-i)} - \widehat{Y}_{\text{G}(\bullet)}) \right]^2. \end{aligned}$$

Note that this statistic can take negative values for some samples. For that reason we might prefer the following proposal

$$b_{2JK}^*(\hat{\bar{Y}}_{CM}) = \max[0, b_{2JK}(\hat{\bar{Y}}_{CM})]$$

Consequently, the jackknife estimator of the CM-estimator MSE is given by

$$mse_{JK}(\hat{\bar{Y}}_{CM}) = \hat{V}_{JK}(\hat{\bar{Y}}_{CM}) + b_{2JK}^*(\hat{\bar{Y}}_{CM}). \quad (4.14)$$

It will be shown in the next section that this estimator is consistent for estimating the MSE of the CM-estimator.

It is also possible to derive another less accurate estimator of the variance using the linearization method. Let $e_{CM,i}$ and $\hat{Q}_{CM,i}$ be defined respectively by

$$e_{CM,i} = y_i - \hat{B}_{CM} \text{ and } \hat{Q}_{CM,i} = \frac{\hat{\sigma}_{CM} \psi_{\bullet H}(e_{CM,i}/\hat{\sigma}_{CM})}{\frac{1}{n} \sum_{k=1}^n \psi'_{\bullet H}(e_{CM,i}/\hat{\sigma}_{CM})}$$

The squared bias of the CM-estimator can be estimated by

$$b_{2TL}(\hat{\bar{Y}}_{CM}) = (\hat{\bar{Y}}_{CM} - \hat{\bar{Y}}_G)^2 - \hat{V}_{TL}(\hat{\bar{Y}}_{CM} - \hat{\bar{Y}}_G),$$

where $\hat{V}_{TL}(\hat{\bar{Y}}_{CM} - \hat{\bar{Y}}_G)$ is given by

$$\hat{V}_{TL}(\hat{\bar{Y}}_{CM} - \hat{\bar{Y}}_G) = \frac{1-f}{n} \frac{1}{n-1} \sum_{i=1}^n (u_i - \bar{u})^2,$$

where

$$\begin{cases} u_i = \hat{Q}_{CM,i} - e_{CM,i}, \\ \bar{u} = (1/n) \sum_{i=1}^n u_i. \end{cases}$$

Again, $b_{2\text{TL}}(\widehat{Y}_{\text{CM}})$ can take negative values. We might therefore prefer $b_{2\text{TL}}^*(\widehat{Y}_{\text{CM}})$ given by

$$b_{2\text{TL}}^*(\widehat{Y}_{\text{CM}}) = \max[0, b_{2\text{TL}}(\widehat{Y}_{\text{CM}})]$$

Therefore, the linearization method leads to the following estimator of the CM-estimator MSE:

$$\text{mse}_{\text{TL}}(\widehat{Y}_{\text{CM}}) = \widehat{V}_{\text{TL}}(\widehat{Y}_{\text{CM}}) + b_{2\text{TL}}^*(\widehat{Y}_{\text{CM}}).$$

4.3.2. Jackknifing the LM-Estimator

We start by recalling the definition of the LM-estimator. Let B_{M} be the parameter defined by

$$\sum_{i=1}^N \psi_H\left(\frac{y_i - B_{\text{M}}}{\sigma_{\text{M}}}\right) = 0,$$

where σ_{M} is of the form

$$\sigma_{\text{M}} = \lambda_0 \inf\{|E_{\text{M},i}| : F_a(|E_{\text{M},i}|) \geq 1/2\}, \quad F_a(r) = \frac{1}{N} \sum_{k=1}^N \mathbf{I}(|E_{\text{M},k}| \leq r),$$

and $E_{\text{M},i} = y_i - B_{\text{M}}$. These two parameters have been motivated by the fact if the population under study can be generated by the common mean superpopulation model where the errors are normally distributed with mean 0 and variance σ^2 then B_{M} and σ_{M} would estimate the common mean β and σ consistently.

The parameters B_{M} and σ_{M} can be consistently estimated from the sample by $\widehat{B}_{\text{M}}$

and $\hat{\sigma}_M$ implicitly defined by the following equation:

$$\sum_{i=1}^n \psi_H \left(\frac{y_i - \hat{B}_M}{\hat{\sigma}_M} \right) = 0,$$

where $\hat{\sigma}_M$ is of the form

$$\hat{\sigma}_M = \lambda_0 \inf\{|e_{M,i}| : \hat{F}_a(e_{M,i}) \geq 1/2\}. \quad \hat{F}_a(r) = \frac{1}{n} \sum_{k=1}^n I(|e_{M,k}| \leq r).$$

and the empirical M-residual $e_{M,i}$ is given by $e_{M,i} = y_i - \hat{B}_M$.

Let us introduce another empirical distribution function $\hat{F}$ given for all real number r by

$$\hat{F}(r) = \frac{1}{n} \sum_{i=1}^n I(e_{M,i} \leq r).$$

The LM-estimator of the population mean $\bar{Y}$ is given by

$$\hat{\bar{Y}}_{LM} = \hat{B}_M + \frac{1}{n} \sum_{i=1}^n r_i$$

where $r_i = w_n e_{M,i} + (1 - w_n) e_{M,i}^*$ and

$$e_{M,i}^* = \frac{e_{M,i}}{1 - 2\alpha_n} I(\alpha_n < \hat{F}(e_{M,i}) \leq 1 - \alpha_n).$$

In chapter 2 we suggested to use $\alpha_n = 1/n$ and w_n is a predetermined weight that can be taken as n/N . For more complex designs, one might modify it to take account of the design. More research is needed in this area.

The asymptotic variance of the LM-estimator has been derived in chapter 2. Under SRS design it is given by

$$V_a(\hat{\bar{Y}}_{LM}) = \frac{1-f}{n} \frac{1}{N-1} \sum_{i=1}^N U_i^2 \quad (4.15)$$

where $U_i = w_n(E_{M,i} - \bar{E}_M) + (1 - w_n)Z_i$, and Z_i is given by

$$Z_i = - \int [I(E_{M,i} \leq r) - F(r)] J(F(r)) dr.$$

In the formula above, $J(\cdot)$ is the function defined for all real number t by $J(t) = (1 - 2\alpha_n)^{-1} I(\alpha_n < t \leq 1 - \alpha_n)$ and F is the population distribution function of the theoretical residuals $E_{M,i}$'s. The corresponding variance estimator is given by

$$\hat{V}_{TL}(\hat{\bar{Y}}_{LM}) = \frac{1-f}{n} \frac{1}{n-1} \sum_{i=1}^n u_i^2 \quad (4.16)$$

where $u_i = w_n(e_{M,i} - \bar{e}_M) + (1 - w_n)z_i$ and z_i is given by

$$z_i = - \int [I(e_{M,i} \leq r) - \hat{F}(r)] J(\hat{F}(r)) dr.$$

A major drawback of this method is the need of an explicit form for z_i in order to estimate the asymptotic variance. For the J function mentioned above such a form is tractable as shown in chapter 2. However, for many J functions available in the literature (see Serfling, 1980) a simple form for z_i is intractable. To overcome this problem we propose a jackknife estimator of the asymptotic variance. It is given by

$$\hat{V}_{JK}(\hat{\bar{Y}}_{LM}) = \frac{1-f}{n} (n-1) \sum_{i=1}^n \left(\hat{\bar{Y}}_{LM(-i)} - \hat{\bar{Y}}_{LM(\bullet)} \right)^2, \quad (4.17)$$

where $\hat{\bar{Y}}_{LM(\bullet)}$ is given by

$$\hat{\bar{Y}}_{LM(\bullet)} = \frac{1}{n} \sum_{i=1}^n \hat{\bar{Y}}_{LM(-i)},$$

and $\hat{\bar{Y}}_{LM(-i)}$ is the LM-estimator based on the original sample after the i -th sample unit is removed. More formally, $\hat{\bar{Y}}_{LM(-i)}$ is given by

$$\hat{\bar{Y}}_{LM(-i)} = \hat{B}_{M(-i)} + \frac{1}{n-1} \sum_{k \neq i}^n r_k$$

where $r_k = w_n e_{M,k} + (1 - w_n) e_{M,k}^*$ and

$$e_{M,k}^* = \frac{e_{M,k}}{1 - 2\alpha_{n-1}} I(\alpha_{n-1} < \widehat{F}_{(-i)}(e_{M,k}) \leq 1 - \alpha_n).$$

The distribution function $\widehat{F}_{(-i)}$ is similar to $\widehat{F}$ except the calculations are based upon the original sample after removing the i -th sample unit. The statistics $\widehat{B}_{M(-i)}$ is obtained in the same way $\widehat{B}_{CM(-i)}$ was obtained except that $\psi_{\bullet H}$ will be replaced with ψ_H .

Since the bias of the LM-estimator will not be negligible for small to moderate sample sizes, again we have to estimate the MSE, not only the variance. It follows from the general results of chapter 2 that the Taylor series method leads to the following estimator of the squared bias:

$$b_{2TL}(\widehat{Y}_{LM}) = \max\left\{0, (1 - w_n)^2 \left[(\bar{e}_M - \bar{e}_{M(\alpha_n)})^2 - \widehat{V}_{TL}(\bar{e}_M - \bar{e}_{M(\alpha_n)}) \right] \right\}.$$

where $\widehat{V}_{TL}(\bar{e}_M - \bar{e}_{M(\alpha_n)})$ is given by

$$\widehat{V}_{TL}(\bar{e}_M - \bar{e}_{M(\alpha_n)}) = \frac{1-f}{n} \frac{1}{n-1} \sum_{i=1}^n t_i^2 \text{ where } t_i = (e_{M,i} - \bar{e}_M) - z_i.$$

The estimator of the squared bias based on the jackknife method is given by

$$b_{2JK}(\widehat{Y}_{LM}) = \max\left[0, (\widehat{Y}_{LM} - \widehat{Y}_G)^2 - \widehat{V}_{JK}(\widehat{Y}_{LM} - \widehat{Y}_G)\right].$$

where

$$\begin{aligned} \widehat{V}_{JK}(\widehat{Y}_{LM} - \widehat{Y}_G) &= \\ &= \frac{1-f}{n} (n-1) \sum_{i=1}^n \left[\left(\widehat{Y}_{LM(-i)} - \widehat{Y}_{LM(\bullet)} \right) - \left(\widehat{Y}_{G(-i)} - \widehat{Y}_{G(\bullet)} \right) \right]^2. \end{aligned}$$

The jackknife estimator of the MSE of the LM-estimator is given by

$$mse_{JK}(\widehat{Y}_{LM}) = \widehat{V}_{JK}(\widehat{Y}_{LM}) + b_{2JK}(\widehat{Y}_{LM}).$$

Likewise, the MSE estimator of $\widehat{Y}_{LM}$ based on the linearization method is given by

$$mse_{TL}(\widehat{Y}_{LM}) = \widehat{V}_{TL}(\widehat{Y}_{LM}) + b_{2TL}(\widehat{Y}_{LM}).$$

4.4. Large Sample Results

Let $\{\mathcal{U}_t, t = 1, 2, \dots\}$ be a sequence of finite populations such that for all t , $\mathcal{U}_t \subseteq \mathcal{U}_{t+1}$. The population size increases as t goes to ∞ . The population index t will sometimes be omitted for simplicity. In any case, all limits will be understood to be as $t \rightarrow \infty$.

The objective of this section is to prove the consistency of the jackknife variance estimator as well as that of the jackknife MSE estimator of $\widehat{Y}_{CM}$ and $\widehat{Y}_{LM}$. Note that the jackknife variance estimators are consistent estimators of the asymptotic variances. Furthermore, the leading term of the squared bias of the proposed robust estimators is of order $O(1/n)$ which is equal to the order of the asymptotic variance. It will be shown in this section that the MSE estimators proposed earlier are consistent estimators of the leading term of the MSE which is of order $O(1/n)$.

4.4.1. The CM-Estimator

We start by introducing some notations first. Let $\hat{F}_t$ and $\hat{F}_{\bullet t}$ be the empirical distribution functions of the CM-residuals $e_{\text{CM}k}$'s and $E_{\text{CM}k}$'s respectively. More formally, we have for all $t \in \mathbb{R}$,

$$\hat{F}_t(t) = \frac{1}{n_t} \sum_{k=1}^{n_t} I(e_{\text{CM},k} \leq r) \text{ and } \hat{F}_{\bullet t}(t) = \frac{1}{n_t} \sum_{k=1}^{n_t} I(E_{\text{CM},k} \leq r)$$

where $e_{\text{CM},k} = y_k - \hat{B}_{\text{CM}}$ and $E_{\text{CM},k} = y_k - B_{\text{CM}}$. When the i -th sample unit is omitted, we consider two versions of $\hat{F}_t$: $\hat{F}_{t(i)}$ and $\hat{F}_{t(-i)}$ defined for all real number r by

$$\hat{F}_{t(i)}(r) = \frac{1}{n_t - 1} \sum_{k \neq i} I(e_{\text{CM},k} \leq r)$$

and

$$\hat{F}_{t(-i)}(r) = \frac{1}{n_t - 1} \sum_{k \neq i} I(e_{\text{CM},k}^{(-i)} \leq r)$$

where $e_{\text{CM},k}^{(-i)} = y_k - \hat{B}_{\text{CM}(-i)}$. Two versions $\hat{F}_{\bullet t(i)}$ and $\hat{F}_{\bullet t(-i)}$ can be defined for $\hat{F}_{\bullet t}$ in a similar way. In this section, the distribution functions of the absolute values of the CM-residuals are often used. These distribution functions are denoted by $\hat{F}_{At}$. $\hat{F}_{\bullet At}$, $\hat{F}_{At(i)}$, $\hat{F}_{At(-i)}$, $\hat{F}_{\bullet At(i)}$ and $\hat{F}_{\bullet At(-i)}$.

The estimator $\hat{\sigma}_{\text{CM}t}$ of the scale parameter $\sigma_{\text{CM}t}$ is used in the definition of the CM-estimator. We recall that

$$\begin{aligned} \hat{\sigma}_{\text{CM}t} &= \lambda_0 \hat{F}_{At}^{-1}(1/2), \\ &= \lambda_0 \inf_{1 \leq k \leq n_t} \left\{ |e_{\text{CM}k}| : \hat{F}_{At}(|e_{\text{CM}k}|) \geq 1/2 \right\} \end{aligned}$$

and

$$\begin{aligned}\sigma_{\text{CM}t} &= \lambda_0 F_{At}^{-1}(1/2), \\ &= \lambda_0 \inf_{1 \leq k \leq N_t} \left\{ |E_{\text{CM}k}| : F_{At}(|E_{\text{CM}k}|) \geq 1/2 \right\}\end{aligned}$$

where F_{At} is defined by

$$F_{At}(r) = \frac{1}{N_t} \sum_{i=1}^{N_t} I(|E_{\text{CM}i}| \leq r).$$

The next theorem states the conditions under which $\hat{\sigma}_{\text{CM}t}^{(i)} = \lambda_0 \hat{F}_{At(i)}^{-1}$ converges in probability to $\sigma_{\text{CM}t}$ uniformly in i . We assume throughout this section that the following conditions are satisfied:

(C1) There is a sequence of functions $f_t(\cdot)$ such that

$$\text{plim}_{t \rightarrow \infty} \left[\frac{\hat{F}_{\bullet t}(r + \delta_t) - \hat{F}_{\bullet t}(r)}{\delta_t} - f_t(r) \right] = 0,$$

for any sequence of random variables $\{\delta_t\}$ of the order $O_p(n_t^{-1/2})$.

(C2) $\lim_{\epsilon \rightarrow 0} \max_{1 \leq i \leq n_t} |I(e_{\text{CM}i} \leq r) - I(e_{\text{CM}i} \leq r - \epsilon)| = o_p(1)$ uniformly in r .

(C3) $\lim_{\epsilon \rightarrow 0} \max_{1 \leq i \leq N_t} |I(E_{\text{CM}i} \leq r) - I(E_{\text{CM}i} \leq r - \epsilon)| = o(1)$ uniformly in r .

Condition (C1) is the stochastic version of condition A7 in Shao (1994). Assumptions (C2) and (C3) are smoothness conditions on the limiting distributions.

Theorem 4.1. *Suppose that the conditions C1, C2 and C3 are satisfied as well as the following two conditions:*

(i) $\hat{B}_{\text{CM}} - B_{\text{CM}} = o_p(n_t^{-1/2})$.

(ii) $\forall a > 0, \exists t_a \text{ \& } \eta_a > 0 : t > t_a \implies F_{\bullet At}(\sigma_{\text{CM}}/\lambda_0 + a) > 1/2 + \eta_a$.

Then

$$\max_{1 \leq i \leq n} (\hat{\sigma}_{\text{CM}}^{(i)} / \sigma_{\text{CM}}) \xrightarrow{p} 1.$$

Proof.

We have to prove that for all $\epsilon > 0$,

$$\lim_{t \rightarrow \infty} P\left(\max_{1 \leq i \leq n_t} |\hat{\sigma}_{\text{CM}}^{(i)} - \sigma_{\text{CM}}| > \epsilon\right) = 0.$$

Let us introduce the following notations:

$$i_0 = \underset{i}{\operatorname{argmin}} \left\{ |\hat{\sigma}_{\text{CM}} - \lambda_0 |e_{\text{CM}i}| : \hat{\sigma}_{\text{CM}} \geq \lambda_0 |e_{\text{CM}i}| \right\}, \quad i_m = \underset{1 \leq i \leq n_t}{\operatorname{argmin}} \{ |e_{\text{CM}i}| \}.$$

$$i_M = \underset{1 \leq i \leq n_t}{\operatorname{argmax}} \{ |e_{\text{CM}i}| \}.$$

Let $\epsilon_0 > 0$. Note that if $\hat{\sigma}_{\text{CM}} = \lambda_0 |e_{\text{CM}i_0}|$, then we have that

$$\hat{\sigma}_{\text{CM}}^{(i)} = \begin{cases} \hat{\sigma}_{\text{CM}}^{(i_0)} & \text{if } i = i_0, \\ \hat{\sigma}_{\text{CM}}^{(i_m)} & \text{if } i < i_0, \\ \hat{\sigma}_{\text{CM}}^{(i_M)} & \text{if } i > i_0 \end{cases}$$

where $\hat{\sigma}_{\text{CM}}^{(i_M)} \leq \hat{\sigma}_{\text{CM}}^{(i_0)} \leq \hat{\sigma}_{\text{CM}}^{(i_m)}$. Moreover if $\hat{\sigma}_{\text{CM}} \neq \lambda_0 |e_{\text{CM}i_0}|$ then we have that

$$\hat{\sigma}_{\text{CM}}^{(i)} = \begin{cases} \hat{\sigma}_{\text{CM}}^{(i_m)} & \text{if } i \leq i_0, \\ \hat{\sigma}_{\text{CM}}^{(i_M)} & \text{if } i > i_0. \end{cases}$$

For a given sample $\hat{\sigma}_{\text{CM}}^{(i)}$ can take at most 3 different values. It follows from this that,

$$P\left(\max_{1 \leq i \leq n_t} |\hat{\sigma}_{\text{CM}}^{(i)} - \sigma_{\text{CM}}| > \epsilon_0\right) \leq P(|\hat{\sigma}_{\text{CM}}^{(i_m)} - \sigma_{\text{CM}}| > \epsilon_0) + P(|\hat{\sigma}_{\text{CM}}^{(i_M)} - \sigma_{\text{CM}}| > \epsilon_0).$$

We will only prove that

$$\lim_{t \rightarrow \infty} P(|\hat{\sigma}_{\text{CM}}^{(i_M)} - \sigma_{\text{CM}}| > \epsilon_0) = 0.$$

The proof that the other term on the right-hand side of the above equality goes to 0 is similar.

$$P(|\hat{\sigma}_{\text{CM}}^{(i_M)} - \sigma_{\text{CM}}| > \epsilon_0) = P(\hat{\sigma}_{\text{CM}}^{(i_M)} - \sigma_{\text{CM}} > \epsilon_0) + P(\hat{\sigma}_{\text{CM}}^{(i_M)} - \sigma_{\text{CM}} < -\epsilon_0)$$

Again we will only prove that

$$\lim_{t \rightarrow \infty} P(\hat{\sigma}_{\text{CM}}^{(i_M)} - \sigma_{\text{CM}} > \epsilon_0) = 0.$$

$$\begin{aligned} P(\hat{\sigma}_{\text{CM}}^{(i_M)} - \sigma_{\text{CM}} > \epsilon_0) &= P(\hat{\sigma}_{\text{CM}}^{(i_M)} > \sigma_{\text{CM}} + \epsilon_0), \\ &\leq P\left\{1/2 > \hat{F}_{\text{At}(i)}[(\sigma_{\text{CM}} + \epsilon_0)/\lambda_0]\right\}, \\ &= P\left\{1 - \hat{F}_{\text{At}(i)}[(\sigma_{\text{CM}} + \epsilon_0)/\lambda_0] > 1/2\right\} \end{aligned} \quad (4.18)$$

It follows from the definition of $\hat{F}_{t(i)}$ that

$$\begin{aligned} \hat{F}_{t(i)}[(\sigma_{\text{CM}} + \epsilon_0)/\lambda_0] &= \hat{F}_{\bullet t(i)}[(\sigma_{\text{CM}} + \epsilon_0)/\lambda_0 + \delta_{n_t}], \\ &= \frac{n_t}{n_t - 1} \hat{F}_{\bullet t}[(\sigma_{\text{CM}} + \epsilon_0)/\lambda_0 + \delta_{n_t}] - \frac{d_i^*}{n_t - 1}, \end{aligned}$$

where $\delta_{n_t} = \hat{B}_{\text{CM}} - B_{\text{CM}}$ and $d_i^* = \mathbf{I}[E_{\text{CM}i} \leq (\sigma_{\text{CM}} + \epsilon_0)/\lambda_0 + \delta_{n_t}]$. It is straightforward to show that if condition (C2) is satisfied then for all $r \in \mathbb{R}$,

$$\hat{F}_{\text{At}(i)}(x) = \hat{F}_{t(i)}(x) - \hat{F}_{t(i)}(-x) + h_t(x) + u_{n_t}$$

where

$$h_t(x) = \lim_{\epsilon \rightarrow 0} (\hat{F}_t(x) - \hat{F}_t(x - \epsilon))$$

and $u_{n_t} = o_p(n_t^{-1})$. Therefore,

$$\begin{aligned}\widehat{F}_{At(i)}\left((\sigma_{CM} + \epsilon_0)/\lambda_0\right) &= \widehat{F}_{At}\left((\sigma_{CM} + \epsilon_0)/\lambda_0 + \delta_{n_t}\right) \\ &\quad + \frac{1}{n_t} \left[d_i^{**} + d_i^* + \widehat{F}_{At}\left((\sigma_{CM} + \epsilon_0)/\lambda_0\right) \right] + u_{n_t}^*\end{aligned}$$

where $d_i^* = \mathbf{I}\left\{E_{CMi} \leq [(\sigma_{CM} + \epsilon_0)/\lambda_0 + \delta_{n_t}]\right\}$ and $u_{n_t}^* = o_p(1/n_t)$.

Let $\widehat{Q}_{CM}$ be defined by

$$\widehat{Q}_{CM} = \frac{1}{n_t} \sum_{i=1}^{n_t} Q_{CM,i}$$

where $Q_{CM,i}$ is defined by (4.11) and let the function $H_t(\cdot)$ be defined for all x by

$H_t(x) = g_t(x)\widehat{Q}_{CM}$ where $g_t(x) = f_t(x) + f_t(-x)$, f_t being the function defined in condition (C1).

After some simple algebra, we obtain

$$\widehat{F}_{At(i)}(s_0) = \widehat{F}_{\bullet At}(s_0) + H_t(s_0) + \Delta_i(s_0), \quad (4.19)$$

where $s_0 = (\sigma_{CM} + \epsilon_0)/\lambda_0$ and

$$\begin{aligned}\Delta_i(s_0) &= \frac{\widehat{F}_{\bullet At}(s_0 + \delta_{n_t}) - \widehat{F}_{\bullet At}(s_0)}{\delta_{n_t}} (\delta_{n_t} - \widehat{Q}_{CM}) \\ &\quad + \left[\frac{\widehat{F}_{\bullet At}(s_0 + \delta_{n_t}) - \widehat{F}_{\bullet At}(s_0)}{\delta_{n_t}} - g_t(s_0) \right] \widehat{Q}_{CM} \\ &\quad + \frac{d_i^* + d_i^{**} + \widehat{F}_{\bullet At}(s_0 + \delta_{n_t})}{n_t} + u_{n_t}^*.\end{aligned}$$

It follows from condition (ii) that there exists t_0 and $\eta_0 > 0$ that if $t > t_0$ then

$\widehat{F}_{\bullet At}(s_0) > 1/2 + \eta_0$. Using (4.18) and (4.19) one gets

$$\begin{aligned}P(\widehat{\sigma}_{CM}^{(i_M)} - \sigma_{CM}) &\leq P\left[1 - \left(\widehat{F}_{\bullet At}(s_0) + H_t(s_0) + \Delta_{i_M}(s_0)\right) > 1/2\right], \\ &\leq P\left[1 - \left(\widehat{F}_{\bullet At}(s_0) + H_t(s_0)\right) > 1/2 - \eta_0/2\right] + P(|\Delta_{i_M}(s_0)| > \eta_0/2)\end{aligned}$$

for $t > t_0$. Using condition (C1), it is straightforward to show that

$$\lim_{t \rightarrow \infty} P(|\Delta_{i_M}(s_0)| > \eta_0/2) = 0.$$

Now we have that

$$\begin{aligned} P\left\{1 - \left(\widehat{F}_{\bullet At}(s_0) + H_t(s_0)\right) > 1/2 - \eta_0/2\right\} &\leq P\left\{1 - \widehat{F}_{\bullet At}(s_0) > 1/2 - \eta_0\right\} \\ &\quad + P\left(|H_t(s_0)| > \eta_0/2\right). \end{aligned}$$

It follows from condition (i) that

$$\lim_{t \rightarrow \infty} P\left(|H_t(s_0)| > \eta_0/2\right) = 0$$

and from Chebychev's inequality that

$$\begin{aligned} P\left\{1 - \widehat{F}_{\bullet At}(s_0) > 1/2 - \eta_0\right\} &= P\left\{[1 - \widehat{F}_{\bullet At}(s_0)] - [1 - F_{\bullet At}(s_0)] > F_{\bullet At}(s_0) - (1/2 + \eta_0)\right\} \\ &\leq \frac{V\left(1 - \widehat{F}_{\bullet At}(s_0)\right)}{F_{\bullet At}(s_0) - (1/2 + \eta_0)} = O(1/n_t). \end{aligned}$$

■

Normally, $\widehat{B}_{CM(-i)}$ is solution to the following equation:

$$\sum_{k \neq i} \psi_{\bullet H} \left(\frac{y_k - \widehat{B}_{CM(-i)}}{\widehat{\sigma}_{CM}^{(i)}} \right) = 0,$$

where $\widehat{\sigma}_{CM^{(i)}}$ is given by (4.8). However, theorem 4.1 implies that $\widehat{B}_{CM(-i)}$ has the same asymptotic distribution as the solution of the equation

$$\widehat{g}_{t(-i)}(t) = \frac{1}{n_t - 1} \sum_{k \neq i} \psi_{\bullet H} \left(\frac{y_k - b}{\sigma_{CM}} \right).$$

Another condition that is assumed to be satisfied is

$$C4. \hat{\sigma}_{CM}/\sigma_{CM} \xrightarrow{P} 1.$$

Because $\hat{\sigma}_{CM}$ and $\hat{B}_{CM}$ are estimated simultaneously, the consistency proof for $\hat{\sigma}_{CM}$ is difficult. However empirical results support this fact (see Huber, 1981, page 176). If the variability of $\hat{B}_{CM}$ is ignored, a consistency proof of $\hat{\sigma}_{CM}$ can be obtained along the lines of proposition 2.1 of Shao and Wu (1992).

It follows from assumption (C4) that $\hat{B}_{CM}$ have the same asymptotic distribution as the solution of the equation $\hat{g}_t(b) = 0$ where

$$\hat{g}_t(b) = \frac{1}{n_t} \sum_{k=1}^{n_t} \psi_{\bullet_H} \left(\frac{y_k - b}{\sigma_{CM}} \right),$$

Therefore, assuming that $\hat{B}_{CM}$ and $\hat{B}_{CM(-i)}$ are implicitly defined by the equations $\hat{g}_t(\hat{B}_{CM}) = 0$ and $\hat{g}_{t(-i)}(\hat{B}_{CM(-i)}) = 0$ will not change the asymptotic results.

We recall that B_{CM} is implicitly defined as solution to the equation $g_t(b) = 0$ where $g_t(\cdot)$ is given for any real number b by

$$g_t(b) = \frac{1}{N_t} \sum_{i=1}^{N_t} \psi_{\bullet_H} \left(\frac{y_i - b}{\sigma_{CM}} \right).$$

We assume as in chapter 2 the following superpopulation model:

$$\xi_0 : \begin{cases} y_i = \mu + \epsilon_i, \\ E_{\xi_0}(\epsilon_i) = 0, V_{\xi_0}(\epsilon_i) = \sigma^2, \\ E_{\xi_0}(\epsilon_i \epsilon_j) = 0 \text{ if } i \neq j. \end{cases}$$

The distribution function of the ϵ_i 's is denoted by F_{ξ_0} . Some general conditions that must be satisfied by $\psi_{\bullet_H}$ are also given in the next results allowing for more flexibility in the choice of the influence functions.

Lemma 4.1. (Serfling, 1980, page 14) Suppose that a sequence of random variables X_n is such that $X_n \xrightarrow{d} X$ and the sequence $\{X_n^r\}$ is uniformly integrable, where $r > 0$. Then $E|X|^r < \infty$ and $\lim_{t \rightarrow \infty} E|X_{n_t}|^r = E|X|^r$.

Let $\mathcal{N}(\mu)$ be the neighborhood of μ . We make further assumptions that will be used regularly:

$$\begin{aligned}
 C5 \quad & \lim_{t \rightarrow \infty} \sigma_{\text{CM},t} = \sigma_0, \\
 C6 \quad & \lim_{t \rightarrow \infty} B_{\text{CM},t} = \lim_{t \rightarrow \infty} \bar{Y}_t = \mu, \\
 C7 \quad & \frac{1}{N_t} \sum_{i=1}^{N_t} |y_i|^{2+\delta} = O(1), \text{ for some } \delta > 0 \\
 C8 \quad & |\psi_{\bullet_H}(y - b)| \leq |y| \quad \forall b \in \mathcal{N}(\mu)
 \end{aligned} \tag{4.20}$$

All these assumptions are standard regularity conditions that are satisfied by most influence functions proposed in the literature. We need the following lemma:

Lemma 4.2. Suppose that the following conditions are satisfied:

- (i) C5 to C8.
 - (ii) $\forall b \in \mathbb{R}$, $\lim_{t \rightarrow \infty} g_t(b) = g_0(b)$ where $g_0(\cdot)$ is a continuous non-increasing function satisfying the condition $g_0(\mu) = 0$ and being strictly decreasing in the neighborhood of 0.
 - (iii) $\sqrt{n_t}(\hat{g}_t(b) - g_t(b)) \xrightarrow{d} N(0, \sigma_0^{*2}(b))$ for all b in the neighborhood of μ and some function $\sigma_0^{*2}(\cdot) > 0$,
 - (iv) $\sup_{t \geq 1} E|\sqrt{n_t}(\hat{g}_t(b) - g_t(b))|^{2+\delta_0} < \infty \quad \forall b \in \mathcal{N}(\mu)$ and some $\delta_0 > 0$,
- Then $\max_{1 \leq i \leq n_t} |\hat{B}_{\text{CM}(-i)} - B_{\text{CM}}| = o_p(1)$.
-

For the CM-estimator using Huber's influence function, the function g_0 in assumption (ii) is given by $g_0(b) = (\mu - b)/\sigma_{\text{CM}}$ and well-known standard conditions will guarantee (iii) is satisfied. Assumption (iv) is a kind of uniform integrability of the sequence of random variables $\{|\sqrt{n_t}(\hat{g}_t(b) - g_t(b))|^2\}$. Note also that the functions $\hat{g}_t(\cdot)$, $\hat{g}_{t(-i)}(\cdot)$, $g_t(\cdot)$ and $g_{t(-i)}(\cdot)$ are supposed to be increasing in the neighborhood of B_{CM} .

Proof.

We have to prove that for all $\epsilon > 0$,

$$\lim_{t \rightarrow \infty} P\left(\max_{1 \leq i \leq n_t} |\hat{B}_{\text{CM}(-i)} - B_{\text{CM}}| > \epsilon\right) = 0.$$

Let $\epsilon_0 > 0$. We have that

$$P\left(\max_{1 \leq i \leq n_t} |\hat{B}_{\text{CM}(-i)} - B_{\text{CM}}| > \epsilon_0\right) \leq \sum_{i=1}^{N_t} P\left(|\hat{B}_{\text{CM}(-i)} - B_{\text{CM}}| > \epsilon_0\right)$$

where $\hat{B}_{\text{CM}(-i)} = B_{\text{CM}}$ if the population unit i is not in the sample. For any population unit i ,

$$P\left(|\hat{B}_{\text{CM}(-i)} - B_{\text{CM}}| > \epsilon_0\right) \leq P\left(\hat{B}_{\text{CM}(-i)} > B_{\text{CM}} + \epsilon_0\right) + P\left(\hat{B}_{\text{CM}(-i)} < B_{\text{CM}} - \epsilon_0\right).$$

Therefore,

$$\begin{aligned} \sum_{i=1}^{N_t} P\left(|\hat{B}_{\text{CM}(-i)} - B_{\text{CM}}| > \epsilon_0\right) &= \sum_{i=1}^{N_t} P\left(\hat{B}_{\text{CM}(-i)} > B_{\text{CM}} + \epsilon_0\right) \\ &\quad + \sum_{i=1}^{N_t} P\left(\hat{B}_{\text{CM}(-i)} < B_{\text{CM}} - \epsilon_0\right). \end{aligned} \tag{4.21}$$

We will only prove that

$$\lim_{t \rightarrow \infty} \sum_{i=1}^{N_t} P\left(\widehat{B}_{\text{CM}(-i)} > B_{\text{CM}} + \epsilon_0\right) = 0$$

the proof that the other term on the right-hand side of (4.21) also goes to 0 being similar. For any population unit i and any sample, the function $\widehat{g}_{t(-i)}(\cdot)$ is increasing. Therefore,

$$P\left(\widehat{B}_{\text{CM}(-i)} > B_{\text{CM}} + \epsilon_0\right) \leq P\left\{\widehat{g}_{t(-i)}(\widehat{B}_{\text{CM}(-i)}) < \widehat{g}_{t(-i)}(B_{\text{CM}} + \epsilon_0)\right\}.$$

But $\widehat{g}_{t(-i)}(\widehat{B}_{\text{CM}(-i)}) = 0$ by definition and

$$\widehat{g}_{t(-i)}(B_{\text{CM}} + \epsilon_0) = \frac{1}{n_t - \xi_i} \left\{ n_t \widehat{g}_t(B_{\text{CM}} + \epsilon_0) - \alpha_i(\epsilon_0) \right\}$$

where

$$\alpha_i(\epsilon_0) = \begin{cases} \psi_{\bullet H} \left[(y_i - (B_{\text{CM}} + \epsilon_0)) / \sigma_{\text{CM}} \right] & \text{if unit } i \text{ is in the sample.} \\ 0 & \text{otherwise} \end{cases}$$

and $\xi_i = 1$ if unit i is in the sample and is 0 otherwise. This implies that

$$\begin{aligned} P\left(\widehat{B}_{\text{CM}(-i)} > B_{\text{CM}} + \epsilon_0\right) &\leq P\left\{\widehat{\alpha}_t(\epsilon_0) > \alpha_i(\epsilon_0)/n_t\right\}, \\ &= P\left\{\widehat{\alpha}_t(\epsilon_0) - \alpha_t(\epsilon_0) - \alpha_i(\epsilon_0)/n_t > \alpha_t(\epsilon_0)\right\}, \\ &\leq \left(\frac{2}{\alpha_t(\epsilon_0)}\right)^{2+\delta_0/2} \left\{ E|\widehat{\alpha}_t(\epsilon_0) - \alpha_t(\epsilon_0)|^{2+\delta_0/2} + n_t^{-(2+\delta_0/2)} E|\alpha_i(\epsilon_0)|^{2+\delta_0/2} \right\} \end{aligned}$$

where $\widehat{\alpha}_t(\epsilon_0) = -\widehat{g}_t(B_{\text{CM}} + \epsilon_0)$, $\alpha_t(\epsilon_0) = -g_t(B_{\text{CM}} + \epsilon_0)$. It follows from assumptions (iii), (iv) and from lemma 4.1 that

$$\lim_{t \rightarrow \infty} E|\sqrt{n_t}(\widehat{g}_t(b) - g_t(b))|^{2+\delta_0/2} = E|Z|^{2+\delta_0/2}$$

where Z is a random variable that follows the standard normal distribution. There exists t_0 such that if $t > t_0$ then

$$E|\widehat{g}_t(b) - g_t(b)|^{2+\delta_0/2} \leq \frac{E|Z|^{2+\delta_0/2} + \epsilon_0}{n_t^{1+\delta_0/4}}.$$

Therefore,

$$\begin{aligned} \sum_{i=1}^{N_t} P\left(\widehat{B}_{\text{CM}(-i)} > B_{\text{CM}} + \epsilon_0\right) &\leq (2\alpha_t^{-1}(\epsilon_0))^{2+\delta_0/2} \frac{E|Z|^{2+\delta_0/2} + \epsilon_0}{n_t^{\delta_0/4}} \\ &\quad + \frac{1}{n_t^{1+\delta_0/2}} \frac{1}{N_t} \sum_{i=1}^{N_t} \left| \psi_{\bullet H} \left[(y_i - (B_{\text{CM}} + \epsilon_0)) / \sigma_{\text{CM}} \right] \right|^{2+\delta_0/2}, \\ &\leq (2\alpha_t^{-1}(\epsilon_0))^{2+\delta_0/2} \left\{ \frac{E|Z|^{2+\delta_0/2} + \epsilon_0}{n_t^{\delta_0/4}} + \frac{1}{n_t^{1+\delta_0/2}} \left(\frac{1}{N_t} \sum_{i=1}^{N_t} |y_i|^{2+\delta_0/2} + \epsilon_0^{2+\delta_0/2} \right) \right\}. \end{aligned}$$

Using conditions C5, C6, C7 and assumption (ii) and taking the limit to infinity, one gets the result.

■

Lemma 4.3. Suppose that

- (i) ψ_H has a bounded first-order derivative $\psi'_H(\cdot)$
- (ii) $\widehat{g}_t(\cdot)$ is always differentiable and has a continuous first-order derivative in the neighborhood of B_{CM} .
- (iii) The conditions of lemma 4.2 are satisfied.

Define for any population unit i two functions $b_i(\cdot)$ and $m_{it}(\cdot)$ by

$$\forall \theta \in (0, 1) \quad b_i(\theta) = \begin{cases} \theta \widehat{B}_{\text{CM}(-i)} + (1 - \theta) \widehat{B}_{\text{CM}} & \text{if } i \text{ is in the sample,} \\ \widehat{B}_{\text{CM}} & \text{otherwise} \end{cases}$$

and $m_{it}(b) = \widehat{g}'_{t(-i)}(b) - g'_t(B_{\text{CM}}) \quad \forall b \in \mathbb{R}$. Then,

$$\max_{1 \leq i \leq n_t} |m_{it}(b_i(\theta))| = o_p(1), \quad \forall \theta \in (0, 1).$$

Proof of lemma 4.3

For all $\theta \in (0, 1)$ and $i = 1, \dots, n_t$ we have that

$$m_{it}(b_i(\theta)) = m_{it}^*(b_i(\theta)) + m_t^*(b_i(\theta)) + m_t^{**}(\widehat{B}_{\text{CM}}) + m_t(B_{\text{CM}}),$$

where

$$m_{it}^*(b) = \widehat{g}_{t(-i)}(b) - \widehat{g}_t'(b); \quad m_t^*(b) = \widehat{g}_t'(b) - \widehat{g}_t'(\widehat{B}_{\text{CM}}),$$

$$m_t^{**}(b) = \widehat{g}_t'(b) - \widehat{g}_t'(B_{\text{CM}}); \quad m_t(b) = \widehat{g}_t'(b) - g_t'(B_{\text{CM}}).$$

Recall that for any real number b we have that

$$\widehat{g}_t(b) = \frac{1}{n_t} \sum_{i=1}^{n_t} \psi_{\bullet_H} \left(\frac{y_i - b}{\sigma_{\text{CM}}} \right).$$

Therefore, for $i = 1, \dots, n_t$,

$$\begin{aligned} \widehat{g}_{t(-i)}(b) &= -\frac{1}{\sigma_{\text{CM},t}(n_t - 1)} \sum_{k \neq i} \psi'_{\bullet_H} \left(\frac{y_k - b}{\sigma_{\text{CM}}} \right), \\ &= \widehat{g}_t'(b) + \frac{1}{n_t - 1} \left[\widehat{g}_t'(b) + (1/\sigma_{\text{CM}}) \psi'_{\bullet_H}((y_i - b)/\sigma_{\text{CM}}) \right]. \end{aligned}$$

This implies that

$$\begin{aligned} m_{it}^*(b_i(\theta)) &= \frac{1}{n_t - 1} \left\{ \widehat{g}_t'(b_i(\theta)) + (1/\sigma_{\text{CM}}) \psi'_{\bullet_H}((y_i - b_i(\theta))/\sigma_{\text{CM}}) \right\}, \\ &= O_p(1/n_t). \end{aligned}$$

The last equality follows from the boundedness of $\psi'_H(\cdot)$.

The next step is to prove that for all $\theta \in (0, 1)$.

$$\lim_{t \rightarrow \infty} P\left(\max_{1 \leq i \leq n_t} |m_t^*(b_i(\theta))| > \epsilon\right) = 0.$$

Let $\epsilon_0 > 0$. It follows from the continuity of $\hat{g}'_t(\cdot)$ that there exists $\eta_{\epsilon_0} > 0$ such that.

$$|b - B_{\text{CM}}| \leq \eta_0 \implies |\hat{g}'_t(b) - \hat{g}'_t(B_{\text{CM}})| \leq \epsilon_0/2.$$

It follows from this that

$$\begin{aligned} P\left(\max_{1 \leq i \leq n_t} |m_t^*(b_i(\theta))| > \epsilon\right) &= P\left(\max_{1 \leq i \leq n_t} |\hat{g}'_t(b_i(\theta)) - \hat{g}'_t(\hat{B}_{\text{CM}})| > \epsilon_0\right), \\ &\leq P\left(\max_{1 \leq i \leq n_t} |\hat{g}'_t(b_i(\theta)) - \hat{g}'_t(B_{\text{CM}})| > \epsilon_0/2\right) \\ &\quad + P\left(|\hat{g}'_t(\hat{B}_{\text{CM}}) - \hat{g}'_t(B_{\text{CM}})| > \epsilon_0/2\right), \\ &\leq P\left(\max_{1 \leq i \leq n_t} |b_i(\theta) - B_{\text{CM}}| > \eta_{\epsilon_0}\right) + P\left(|\hat{B}_{\text{CM}} - B_{\text{CM}}| > \eta_{\epsilon_0}\right), \\ &\leq P\left(\max_{1 \leq i \leq n_t} |\hat{B}_{\text{CM}(-i)} - B_{\text{CM}}| > (2\theta)^{-1}\eta_{\epsilon_0}\right) \\ &\quad + P\left(|\hat{B}_{\text{CM}} - B_{\text{CM}}| > (2(1-\theta))^{-1}\eta_{\epsilon_0}\right). \end{aligned}$$

The last inequality follows from the fact that

$$b_i(\theta) - B_{\text{CM}} = \theta(B_{\text{CM}(-i)} - B_{\text{CM}}) + (1-\theta)(\hat{B}_{\text{CM}} - B_{\text{CM}}).$$

It follows from lemma 4.2 that

$$\lim_{t \rightarrow \infty} P\left(\max_{1 \leq i \leq n_t} |\hat{B}_{\text{CM}(-i)} - B_{\text{CM}}| > (2\theta)^{-1}\eta_{\epsilon_0}\right) = 0$$

and

$$\lim_{t \rightarrow \infty} P\left(|\hat{B}_{\text{CM}} - B_{\text{CM}}| > (2(1-\theta))^{-1}\eta_{\epsilon_0}\right) = 0.$$

The fact that

$$\lim_{t \rightarrow \infty} P\left(|m_t^{**}(\hat{B}_{\text{CM}})| > \epsilon_0\right) = \lim_{t \rightarrow \infty} P\left(|m_t(B_{\text{CM}})| > \epsilon_0\right) = 0$$

is trivial.

■

We now prove the second lemma that is needed for proving theorem 4.2. Let $S_{Y,t}^k$ be defined by

$$S_{pt}^2 = \frac{1}{N_t} \sum_{i=1}^{N_t} |y_i - \bar{Y}_t|^p \quad (4.22)$$

Lemma 4.4. *Suppose that*

- (i) $S_{4t}^2 = O(1)$.
- (ii) $\lim_{t \rightarrow \infty} S_{2t}^2 = \sigma_0^2$.
- (iii) *There exists $\alpha_0 > 0$ such that $\lim_{t \rightarrow \infty} P\left(|\hat{g}'_t(b)| \geq \alpha_0 / b \in \mathcal{N}(B_{\text{CM}})\right) = 1$.*
where $\mathcal{N}(B_{\text{CM}})$ is the neighborhood of B_{CM} .
- (iv) *Conditions of lemma 4.2 are satisfied. Then*

$$\sum_{i=1}^{n_t} (\hat{B}_{\text{CM}(-i)} - \hat{B}_{\text{CM}})^2 = O_p(1/n_t).$$

Proof of lemma 4.4

Let $\epsilon_0 > 0$. It is sufficient to prove that there exists $t_{\epsilon_0} > 0$ and $\eta_{\epsilon_0} > 0$ such that

$$P\left(\sum_{i=1}^{n_t} (n_t - 1) (\hat{B}_{\text{CM}(-i)} - \hat{B}_{\text{CM}})^2 \geq \eta_{\epsilon_0}\right) \leq \epsilon_0.$$

It follows from the mean value theorem applied to $\widehat{g}_t$ that there exists b_i between $\widehat{B}_{\text{CM}(-i)}$ and $\widehat{B}_{\text{CM}}$ such that

$$(n_t - 1)(\widehat{B}_{\text{CM}(-i)} - \widehat{B}_{\text{CM}})\widehat{g}'_t(b_i) = \psi_{\bullet H}\left((y_i - \widehat{B}_{\text{CM}})/\sigma_{\text{CM}}\right).$$

Moreover, it follows from the assumptions (i) and (ii) that $\widehat{S}_{2t}^2 = O_p(1)$ where

$$\widehat{S}_{2t}^2 = \frac{1}{n_t - 1} \sum_{i=1}^{n_t} (y_i - \bar{Y}_t)^2.$$

Therefore, there exists $\eta_{\epsilon_0}^* > 0$ and $t_{\epsilon_0}^*$ such that $P(\widehat{S}_{2t}^2/\sigma_{\text{CM}}^2 \geq \eta_{\epsilon_0}^*) \leq \epsilon_0$ if $t > t_{\epsilon_0}^*$.

Let $\eta_{\epsilon_0} = 2\eta_{\epsilon_0}^*/\alpha_0^2$. We have that

$$\begin{aligned} & P\left(\sum_{i=1}^{n_t} (n_t - 1)(\widehat{B}_{\text{CM}(-i)} - \widehat{B}_{\text{CM}})^2 \geq \eta_{\epsilon_0}\right) \\ & \leq P\left(\sum_{i=1}^{n_t} (n_t - 1)(\widehat{B}_{\text{CM}(-i)} - \widehat{B}_{\text{CM}})^2 \widehat{g}_t'^2(b_i) \geq \eta_{\epsilon_0} \alpha_0^2\right) \\ & \quad + P\left(\inf_{1 \leq i \leq n_t} \widehat{g}_t'^2(b_i) < \alpha_0^2\right) \\ & = P\left(\frac{1}{n_t - 1} \sum_{i=1}^{n_t} \psi_{\bullet H}^2\left((y_i - \widehat{B}_{\text{CM}})/\sigma_{\text{CM}}\right) \geq \eta_{\epsilon_0} \alpha_0^2\right) \\ & \quad + P\left(\inf_{1 \leq i \leq n_t} \widehat{g}_t'^2(b_i) < \alpha_0^2\right) \end{aligned}$$

It is not difficult to see that

$$\begin{aligned} \psi_{\bullet H}^2\left((y_i - \widehat{B}_{\text{CM}})/\sigma_{\text{CM}}\right) & \leq (y_i - \widehat{B}_{\text{CM}})^2/\sigma_{\text{CM}}^2, \\ & = (y_i - \bar{Y}_t)^2/\sigma_{\text{CM}}^2 + \delta_{n_t} \end{aligned}$$

where $\delta_{n_t} = o_p(1)$. This implies that

$$\begin{aligned} P\left(\sum_{i=1}^{n_t} (n_t - 1)(\widehat{B}_{\text{CM}(-i)} - \widehat{B}_{\text{CM}})^2 \geq \eta_{\epsilon_0}\right) & \leq P\left(\widehat{S}_{2t}^2/\sigma_{\text{CM}}^2 > \eta_{\epsilon_0} \alpha_0^2/2\right) \\ & \quad + P(|\delta_{n_t}| > \eta_{\epsilon_0} \alpha_0^2/2) + P\left(\inf_{1 \leq i \leq n_t} \widehat{g}_t'^2(b_i) < \alpha_0^2\right) \end{aligned}$$

To complete the proof, we only need to show that

$$\lim_{t \rightarrow \infty} P(|\delta_{n_t}| > \eta_{\epsilon_0} \alpha_0^2/2) + P\left(\inf_{1 \leq i \leq n_t} \widehat{g}_t'^2(b_i) < \alpha_0^2\right) = 0.$$

It follows from assumption (iii) that there exists $\delta_0 > 0$ such that

$$\lim_{t \rightarrow \infty} P\left(|\widehat{g}_t'(b)| \geq \alpha_0/b \mid b \in \mathcal{I}(B_{\text{CM}}, \delta_0)\right) = 1$$

where $\mathcal{I}(B_{\text{CM}}, \delta_0) = (B_{\text{CM}} - \delta_0, B_{\text{CM}} + \delta_0)$. Let us define the following event:

$$\mathcal{B}_t = \left\{ \forall i = 1, \dots, n_t, b_i \in \mathcal{I}(B_{\text{CM}}, \delta_0) \right\}$$

and $\overline{\mathcal{B}}_t$ the corresponding complement. We have that

$$\begin{aligned} P\left(\inf_{1 \leq i \leq n_t} \widehat{g}_t'^2(b_i) \geq \alpha_0^2\right) &= P\left(\inf_{1 \leq i \leq n_t} \widehat{g}_t'^2(b_i) \geq \alpha_0^2 \mid \mathcal{B}_t\right) P(\mathcal{B}_t) \\ &\quad + P\left(\inf_{1 \leq i \leq n_t} \widehat{g}_t'^2(b_i) \geq \alpha_0^2 \mid \overline{\mathcal{B}}_t\right) P(\overline{\mathcal{B}}_t), \\ &\geq P\left(\inf_{1 \leq i \leq n_t} \widehat{g}_t'^2(b_i) \geq \alpha_0^2 \mid \mathcal{B}_t\right) P(\mathcal{B}_t) \end{aligned}$$

It follows from lemma 4.2 that $\lim_{t \rightarrow \infty} P(\mathcal{B}_t) = 1$ and from assumption (iii) that

$$\lim_{t \rightarrow \infty} P\left(\inf_{1 \leq i \leq n_t} \widehat{g}_t'^2(b_i) \geq \alpha_0^2 \mid \mathcal{B}_t\right) = 1.$$

■

Theorem 4.2. Suppose that the conditions of lemma 4.2 are satisfied as well as the following assumptions:

- (i) $|g'_t(b)| > 0 \forall b \in \mathcal{N}(B_{\text{CM}})$ where $\mathcal{N}(B_{\text{CM}})$ is the neighborhood of B_{CM} .
- (ii) $\widehat{B}_{\text{CM}} - B_{\text{CM}} = O_p(n_t^{-1/2})$.

(iii) $\sup_{t \geq 1} \sum_{i=1}^{N_t} |y_i - \bar{Y}_t|^{1+\delta} / N_t$ for some $\delta > 0$.

(iv) Conditions of lemmas 4.3 and 4.4 are all satisfied.

Then the jackknife estimator $\hat{V}_{JK}(\hat{\bar{Y}}_{CM})$ given in (4.3) is consistent for $V_a(\hat{\bar{Y}}_{CM})$ given in (4.10), that is

$$\hat{V}_{JK}(\hat{\bar{Y}}_{CM}) / V_a(\hat{\bar{Y}}_{CM}) \xrightarrow{p} 1.$$

Proof of theorem 4.2 .

It follows from the mean value theorem that for any sample unit i ,

$$\hat{g}_{t(-i)}(\hat{B}_{CM(-i)}) = \hat{g}_{t(-i)}(\hat{B}_{CM}) + (\hat{B}_{CM(-i)} - \hat{B}_{CM})\hat{g}'_{t(-i)}(b_i(\theta_0)),$$

for some $\theta_0 \in (0, 1)$ where $b_i(\theta)$ is given by

$$b_i(\theta) = \theta \hat{B}_{CM(-i)} + (1 - \theta) \hat{B}_{CM}$$

But $\hat{g}_{t(-i)}(\hat{B}_{CM(-i)}) = 0$ and

$$\begin{aligned} \hat{g}_{t(-i)}(\hat{B}_{CM}) &= \frac{1}{n_t - 1} \left\{ n_t \hat{g}_t(\hat{B}_{CM}) - \psi_{\bullet H} \left((y_i - \hat{B}_{CM}) / \sigma_{CM} \right) \right\}, \\ &= -\frac{1}{n_t - 1} \psi_{\bullet H} \left((y_i - \hat{B}_{CM}) / \sigma_{CM} \right) \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{1}{n_t - 1} \psi_{\bullet H} \left((y_i - \hat{B}_{CM}) / \sigma_{CM} \right) &= (\hat{B}_{CM(-i)} - \hat{B}_{CM}) \hat{g}'_{t(-i)}(b_i(\theta_0)), \\ &= (\hat{B}_{CM(-i)} - \hat{B}_{CM}) \left[m_{it}(b_i(\theta_0)) + g'_t(B_{CM}) \right] \end{aligned}$$

where

$$m_{it}(b) = \widehat{g}'_{t(-i)}(b) - g'_t(B_{CM}), \quad \forall b \in \mathbb{R}.$$

Applying the mean value theorem again on the term on the left-hand side of the equality above, one obtains

$$(\widehat{B}_{CM(-i)} - \widehat{B}_{CM})g'_t(B_{CM}) = \frac{1}{n_t - 1}\psi_{\bullet H}\left((y_i - B_{CM})/\sigma_{CM}\right) + \Delta_i(\theta_0, \theta_0^*),$$

for some $\theta_0^* \in (0, 1)$ where

$$\Delta_i(\alpha, \beta) = -\left\{ (\widehat{B}_{CM(-i)} - \widehat{B}_{CM})m_{it}(b_i(\alpha)) + \frac{(\widehat{B}_{CM} - B_{CM})}{(n_t - 1)\sigma_{CM}}\psi'_{\bullet H}\left((y_i - b_i(\beta))/\sigma_{CM}\right) \right\}$$

It follows from this that

$$(\widehat{B}_{CM(-i)} - \widehat{B}_{CM}) = \frac{1}{n_t - 1}Q_{CM,i} + \frac{\Delta_i(\theta_0, \theta_0^*)}{g'_t(B_{CM})}$$

where

$$Q_{CM,i} = \frac{1}{(n_t - 1)g'_t(B_{CM})}\psi_{\bullet H}\left((y_i - B_{CM})/\sigma_{CM}\right).$$

We have that,

$$\begin{aligned} \widehat{V}_{JK}(\widehat{B}_{CM}) &= \frac{1-f}{n_t} \sum_{i=1}^{n_t} (n_t - 1)(\widehat{B}_{CM(-i)} - \widehat{B}_{CM})^2, \\ &= \frac{1-f}{n_t} \frac{1}{n_t - 1} \sum_{i=1}^{n_t} Q_{CM,i}^2 + \lambda_{1t}(\theta_0, \theta_0^*) + \lambda_{2t}(\theta_0, \theta_0^*) \end{aligned} \quad (4.23)$$

where $\lambda_{1t}(\theta_0, \theta_0^*)$ and $\lambda_{2t}(\theta_0, \theta_0^*)$ are respectively given by

$$\lambda_{1t}(\theta_0, \theta_0^*) = \frac{1-f}{n_t} \sum_{i=1}^{n_t} (n_t - 1) \frac{\Delta_i^2(\theta_0, \theta_0^*)}{g_t'^2(B_{CM})}$$

and

$$\lambda_{2t}(\theta_0, \theta_0^*) = 2 \frac{1-f}{n_t} \sum_{i=1}^{n_t} \frac{\Delta_i(\theta_0, \theta_0^*)}{g_t'(B_{CM})} Q_{CM,i}.$$

It follows from condition (iii) (see Scott and Wu, 1981) that the first term is $O_p[V_a(\hat{B}_{CM})]$. It remains to show that $\lambda_{1t}(\theta_0, \theta_0^*) = o_p(1/n_t)$ and $\lambda_{2t}(\theta_0, \theta_0^*) = o_p(1/n_t)$ as $O_p[V_a(\hat{B}_{CM})] = O_p(1/n_t)$.

We have that

$$\begin{aligned} \lambda_{1t}(\theta_0, \theta_0^*) &= \frac{1}{g_t'^2(B_{CM})} \left\{ \frac{1-f}{n_t} \sum_{i=1}^{n_t} (n_t - 1) (\hat{B}_{CM(-i)} - \hat{B}_{CM})^2 m_{it}^2(b_i(\theta_0)) + \right. \\ &\quad + (\hat{B}_{CM} - B_{CM})^2 \frac{1-f}{n_t} \frac{1}{(n_t - 1)\sigma_{CM}^2} \sum_{i=1}^{n_t} \psi_{\bullet H}'^2((y_i - b_i(\theta_0^*)) / \sigma_{CM}) + \\ &\quad + (\hat{B}_{CM} - B_{CM}) \frac{2(1-f)}{n_t \sigma_{CM}} \sum_{i=1}^{n_t} m_{it}(b_i(\theta_0)) (\hat{B}_{CM(-i)} - \hat{B}_{CM}) \\ &\quad \left. \times \psi_{\bullet H}'((y_i - b_i(\theta_0^*)) / \sigma_{CM}) \right\} \\ &\leq \frac{1}{g_t'^2(B_{CM})} \left\{ \left[\max_{1 \leq i \leq n_t} m_{it}^2(b_i(\theta_0)) \right] \times \frac{1-f}{n_t} \sum_{i=1}^{n_t} (n_t - 1) (\hat{B}_{CM(-i)} - \hat{B}_{CM})^2 \right. \\ &\quad + \left[(\hat{B}_{CM} - B_{CM})^2 / \sigma_{CM}^2 \right] \times \frac{1-f}{n_t} \frac{1}{n_t - 1} \sum_{i=1}^{n_t} \psi_{\bullet H}'^2((y_i - b_i(\theta_0^*)) / \sigma_{CM}) \\ &\quad + 2 \left(|\hat{B}_{CM} - B_{CM}| / \sigma_{CM} \right) \left(\max_{1 \leq i \leq n_t} |m_{it}(b_i(\theta_0))| \right) \left(\frac{1}{n_t} \sum_{i=1}^{n_t} (\hat{B}_{CM(-i)} - \hat{B}_{CM})^2 \right)^{1/2} \\ &\quad \left. \times \left(\frac{1}{n_t} \sum_{i=1}^{n_t} \psi_{\bullet H}'^2((y_i - b_i(\theta_0^*)) / \sigma_{CM}) \right)^{1/2} \right\} \end{aligned}$$

Since the ψ_H function is supposed to have a bounded first-order derivative, it follows from assumption (ii) that $\lambda_{1t}(\theta_0, \theta_0^*) = o_p(1/n_t)$ if the following two conditions are satisfied:

$$\max_{1 \leq i \leq n_t} |m_{it}(b_i(\theta_0))| = o_p(1)$$

and

$$\sum_{i=1}^{n_t} (\hat{B}_{\text{CM}(-i)} - \hat{B}_{\text{CM}})^2 = O_p(1/n_t).$$

But this follows from the lemmas 4.3 and 4.4 .

■

The next theorem deals with the consistency of the jackknife MSE estimator of $\hat{\bar{Y}}_{\text{CM}}$.

We will prove that for large samples, the jackknife MSE estimator approximates the asymptotic MSE of the CM-estimator, the neglected terms being of order $o_p(1/n_t)$.

Theorem 4.3.

Suppose that the conditions of theorem 4.2 are all satisfied, that $\forall r \in \mathbb{R}$.

$|\psi_H(r, K_n) - r| \leq r^2/K_n$ and $|1 - \psi'_H(r, K_n)| < 2|r|/K_n$. Assume also that

(i) $\sum_{i=1}^{N_t} |y_i - \bar{Y}_t|^4 / N_t = O(1)$.

(ii) Suppose that for some $\delta_0 > 0$ and $\delta_0^* > 0$, we have

$$\sup_{t \geq 1} E |\sqrt{n_t}(\hat{B}_{\text{CM}} - B_{\text{CM}})|^{12+\delta_0} < \infty \text{ and } \sup_{t \geq 1} E |\sqrt{n_t}(\hat{\bar{Y}}_{\text{G}} - \bar{Y}_t)|^{6+\delta_0^*} < \infty$$

(iii) $\bar{Y}_t = B_{\text{CM}} + O(n_t^{-1/2})$.

Then,

$$mse_{\text{JK}}(\hat{\bar{Y}}_{\text{CM}}) = MSE_A(\hat{\bar{Y}}_{\text{CM}}) + o_p(1/n_t),$$

where $mse_{\text{JK}}(\hat{\bar{Y}}_{\text{CM}})$ is given by (4.14) and $MSE_A(\hat{\bar{Y}}_{\text{CM}})$ given by (4.13).

The consistency of the jackknife estimator of $MSE_A(\hat{\bar{Y}}_{\text{CM}})$ follows from this theorem. We need the following two lemmas to prove this theorem.

Lemma 4.5. (*Bhat (1981), lemma 5.2, pp. 85*)

Let $\{X_n\}_{n \geq 1}$ be a sequence of random variables. If $\sup_{n \geq 1} E|X_n| < \infty$ then $X_n = O_p(1)$.

Lemma 4.6. Suppose the conditions of theorem 4.3 are all satisfied. Then we have that $b_{2JK} \mathbf{I}(b_{2JK} < 0) = o_p(1/n_t)$.

The proof of this lemma can be found in the appendix B.

Proof of theorem 4.3 .

We know from (4.14) that

$$mse_{JK}(\widehat{Y}_{CM}) = \widehat{V}_{JK}(\widehat{Y}_{CM}) + b_{2JK}^*(\widehat{Y}_{CM}).$$

It was proved in theorem 4.2 that $\widehat{V}_{JK}(\widehat{Y}_{CM}) = V_a(\widehat{Y}_{CM}) + o_p(1/n_t)$. Therefore, it remains to prove that $b_{2JK}^*(\widehat{Y}_{CM}) = B_a^2(\widehat{Y}_{CM}) + o_p(1/n_t)$. For simplicity, $b_{2JK}^*(\widehat{Y}_{CM})$ and $b_{2JK}(\widehat{Y}_{CM})$ will sometimes be denoted b_{2JK}^* and b_{2JK} respectively. We know that

$$b_{2JK}^* = b_{2JK} \mathbf{I}(b_{2JK} > 0),$$

where

$$b_{2JK} = (\widehat{B}_{CM} - \widehat{Y}_G)^2 - \widehat{V}_{JK}(\widehat{B}_{CM} - \widehat{Y}_G).$$

It follows from lemma 4.6 that $b_{2JK} \mathbf{I}(b_{2JK} < 0) = o_p(1/n_t)$. Therefore it remains to

show that $b_{2JK} - B_a^2(\widehat{Y}_{CM}) = o_p(1/n_t)$. We have that,

$$\begin{aligned} b_{2JK} - B_a^2(\widehat{Y}_{CM}) &= (\widehat{Y}_{CM} - \widehat{Y}_G)^2 - E(\widehat{Y}_{CM} - \widehat{Y}_G)^2 \\ &\quad - \left[\widehat{V}_{JK}(\widehat{Y}_{CM} - \widehat{Y}_G) - V_a(\widehat{Y}_{CM} - \widehat{Y}_G) \right] \end{aligned}$$

Using the same techniques as in the proof of theorem 4.2 , it is possible to prove that

$$\widehat{V}_{JK}(\widehat{Y}_{CM} - \widehat{Y}_G) - V_a(\widehat{Y}_{CM} - \widehat{Y}_G) = o_p(1/n_t).$$

Therefore, it remains to prove that

$$(\widehat{Y}_{CM} - \widehat{Y}_G)^2 - E(\widehat{Y}_{CM} - \widehat{Y}_G)^2 = o_p(1/n_t).$$

Let $\delta_1 \in (0, 1)$. We have that

$$\begin{aligned} \sup_{t \geq 1} n_t^{1+\delta_1} E \left\{ \left| (\widehat{Y}_{CM} - \widehat{Y}_G)^2 - E(\widehat{Y}_{CM} - \widehat{Y}_G)^2 \right| \right\}^{1+\delta_1} &\leq E \left| \sqrt{n_t} (\widehat{Y}_{CM} - \widehat{Y}_G) \right|^{2(1+\delta_1)} \\ &\leq 2^{2(1+\delta_1)} E \left| \sqrt{n_t} (\widehat{B}_{CM} - B_{CM}) \right|^{2(1+\delta_1)} + 2^{4(1+\delta_1)} E \left| \sqrt{n_t} (\widehat{Y}_G - \bar{Y}_t) \right|^{2(1+\delta_1)} \\ &\quad + 2^{4(1+\delta_1)} E \left| \sqrt{n_t} (B_{CM} - \bar{Y}_t) \right|^{2(1+\delta_1)} \end{aligned}$$

It follows from the assumptions (ii) and (iii) that this quantity is finite and from lemma 4.5 that

$$(\widehat{Y}_{CM} - \widehat{Y}_G)^2 - E(\widehat{Y}_{CM} - \widehat{Y}_G)^2 = O_p(n_t^{-\delta_1})$$

which completes the proof.

■

4.4.2. The LM-Estimator

Throughout this section, we will need the following four distribution functions F_M .

$\hat{F}_M$, $\hat{F}_{M(-i)}$ and $\hat{F}_{*M(-i)}$. These are defined for any real number r by

$$F_M(r) = \frac{1}{N_t} \sum_{k=1}^{N_t} I(E_{Mk} \leq r), \quad \hat{F}_M(r) = \frac{1}{n_t} \sum_{k=1}^{n_t} I(e_{Mk} \leq r),$$

$$\hat{F}_{M(-i)}(r) = \frac{1}{n_t - 1} \sum_{k \neq i} I(e_{Mk} \leq r), \quad \hat{F}_{*M(-i)}(r) = \frac{1}{n_t - 1} \sum_{k \neq i} I(e_{Mk}^{(i)} \leq r).$$

where $E_{Mk} = y_k - B_M$, $e_{Mk} = y_k - \hat{B}_M$, and $e_{Mk}^{(i)} = y_k - \hat{B}_{M(-i)}$.

To prove the consistency of the jackknife estimator of the LM-estimator asymptotic variance, we need the following two conditions:

C9: The sequence of distribution functions $\{F_{Mt}\}_{t \geq 1}$ is tight.

C10: There exists a positive real number M such that $\lim_{t \rightarrow \infty} \max_{1 \leq i \leq N_t} |y_i - B_{Mt}| \leq M$.

In addition to the conditions introduced above, we will need three conditions $C1'$, $C2'$ and $C3'$ identical to $C1$, $C2$ and $C3$ but applied to the M-estimator and not to the CM-estimator. These conditions guarantee the consistency of the jackknife variance estimator of $\hat{B}_M$.

We start by proving the consistency of the variance estimator (4.16) obtained by the linearization series method. This result is important in proving the consistency of the jackknife variance estimator. In chapters 2 and 3, we have assumed that this estimator is consistent for estimating the asymptotic variance without proving it.

That proof will be given now in a more general setting where the J_n -function is not necessarily bounded.

We recall that a function f is Lipschitz continuous if there exists a constant C such that $\forall t, s \in \mathbb{R}$,

$$|f(t) - f(s)| \leq C|t - s|.$$

Theorem 4.4. *It assumed that*

- (i) *The J_n -function is Lipschitz continuous in $[0, 1]$.*
- (ii) $\sup_{t \geq 1} \int |r| dF_{Mt}(r) < \infty$.
- (iii) *There exists $\sigma_U^2 > 0$ such that $\lim_{t \rightarrow \infty} \frac{1}{N_t - 1} \sum_{i=1}^{N_t} U_i^2 = \sigma_U^2$.*

Then

$$\widehat{V}_{TL}(\widehat{Y}_{LM}) - V_a(\widehat{Y}_{LM}) = o_p(1/n_t).$$

Proof.

We know from (4.16) that

$$\begin{aligned} u_i &= w_n(e_{Mi} - \bar{e}_M) + (1 - w_n)z_i, \\ &= w_n(E_{Mi} - \bar{E}_M) + (1 - w_n)Z_i + (1 - w_n)(z_i - Z_i), \\ &= U_i + (1 - w_n)(z_i - Z_i). \end{aligned}$$

It follows from this that

$$\widehat{V}_{JK}(\widehat{Y}_{LM}) = \frac{1-f}{n_t} \frac{1}{n_t - 1} \sum_{i=1}^{n_t} U_i^2 + m_{1t} + m_{2t}.$$

$$\text{where } \begin{cases} m_{1t} = \frac{(1-f)(1-w_n)^2}{n_t} \frac{1}{n_t-1} \sum_{i=1}^{n_t} (z_i - Z_i)^2, \\ m_{2t} = 2 \frac{(1-f)(1-w_n)}{n_t} \frac{1}{n_t-1} \sum_{i=1}^{n_t} U_i(z_i - Z_i). \end{cases}$$

Let z_{0i} be defined as follows,

$$z_{0i} = - \int \left(I(E_{Mi} \leq r) - \hat{F}_{0M}(r) \right) J_n(\hat{F}_{0M}(r)) dr, \text{ where } \hat{F}_{0M}(r) = \frac{1}{n_t} \sum_{i=1}^{n_t} \delta_i(r).$$

where $\delta_i(r) = I(E_{Mi} \leq r)$ and let m_{1t}^* and m_{2t}^* be defined by

$$\begin{aligned} m_{1t}^* &= \frac{(1-f)(1-w_n)^2}{n_t} \frac{1}{n_t-1} \sum_{i=1}^{n_t} (z_{0i} - Z_i)^2, \\ m_{2t}^* &= 2 \frac{(1-f)(1-w_n)}{n_t} \frac{1}{n_t-1} \sum_{i=1}^{n_t} U_i(z_{0i} - Z_i). \end{aligned}$$

Since $\hat{B}_M - B_M \xrightarrow{p} 0$, m_{1t}^* and m_{2t}^* have respectively the same asymptotic distributions as m_{1t} and m_{2t} . It is proved in appendix B that $m_{1t}^* = o_p(1/n_t)$ and $m_{2t}^* = o_p(1/n_t)$. These two results along with assumption (iii) lead to the conclusion.

■

Theorem 4.5. Suppose that the conditions $C1'$, $C2'$ and $C3'$ as well as $C9$ and $C10$ hold. Then the jackknife variance estimator $\hat{V}_{JK}(\hat{\bar{Y}}_{LM})$ in (4.17) is consistent for $V_a(\hat{\bar{Y}}_{LM})$ in (4.15), that is

$$\hat{V}_{JK}(\hat{\bar{Y}}_{LM}) / V_a(\hat{\bar{Y}}_{LM}) \xrightarrow{p} 1.$$

Proof.

The LM-estimator of $\bar{Y}$ is given by $\hat{\bar{Y}}_{LM} = \hat{B}_M + w_n \bar{e}_M + (1 - w_n) \bar{e}_M^{(\alpha_n)}$, where

$$\bar{e}_M^{(\alpha_n)} = \frac{1}{n} \sum_{k=1}^n e_{Mk}^*, \text{ and } e_{Mk}^* = \begin{cases} e_{Mk}/(1 - 2\alpha_n) & \text{if } \alpha_n < \hat{F}_M(e_{Mk}) \leq 1 - \alpha_n, \\ 0, & \text{otherwise.} \end{cases}$$

If the i -th sample unit is omitted, then the LM-estimator becomes

$$\hat{\bar{Y}}_{LM(-i)} = \hat{B}_{M(-i)} + w_n \bar{e}_{M(-i)} + (1 - w_n) \bar{e}_{M(-i)}^{(\alpha_{n-1})}, \text{ where } \bar{e}_{M(-i)}^{(\alpha_{n-1})} = \frac{1}{n-1} \sum_{k \neq i} e_{Mk}^{*(i)}.$$

and

$$e_{Mk}^{*(i)} = \begin{cases} e_{Mk}^{(i)}/(1 - 2\alpha_{n-1}) & \text{if } \alpha_{n-1} < \hat{F}_{*M(-i)}(e_{Mk}^{(i)}) \leq 1 - \alpha_{n-1}, \\ 0, & \text{otherwise.} \end{cases}$$

Since $e_{Mk}^{(i)} = y_k - \hat{B}_{M(-i)}$, it is straightforward to see that

$$\bar{e}_{M(-i)} - \bar{e}_M = \left(\hat{B}_M - \hat{B}_{M(-i)} \right) - \frac{1}{n-1} (e_{Mi} - \bar{e}_M).$$

The trimmed mean of the empirical residuals can be rewritten as

$$\begin{aligned} \bar{e}_M^{(\alpha_n)} &= \int r J_n(\hat{F}_M(r)) d\hat{F}_M(r), \\ &= \int r \left(\frac{dK_n(\delta)}{d\delta} \Big|_{\delta=\hat{F}_M(r)} \right) d\hat{F}_M(r), \\ &= \int r dK_n(\hat{F}_M(r)), \text{ where } K_n(t) = \int_0^t J_n(s) ds. \end{aligned}$$

Let $\lambda_{*k}^{(i)} = J_{(n-1)}(\hat{F}_{*M(-i)}(e_{Mk}^{(i)}))$. We have that

$$\begin{aligned} \bar{e}_{M(-i)}^{(\alpha_{n-1})} &= \frac{1}{n-1} \sum_{k \neq i} \lambda_{*k}^{(i)} e_{Mk}^{(i)} = \frac{1}{n-1} \sum_{k \neq i} \lambda_{*k}^{(i)} (y_k - \hat{B}_{M(-i)}), \\ &= \frac{1}{n-1} \sum_{k \neq i} \lambda_{*k}^{(i)} (e_{Mk} - (\hat{B}_{M(-i)} - \hat{B}_M)), \\ &= \frac{1}{n-1} \sum_{k \neq i} \lambda_{*k}^{(i)} e_{Mk} - (\hat{B}_{M(-i)} - \hat{B}_M) \left(\frac{1}{n-1} \sum_{k \neq i} \lambda_{*k}^{(i)} \right). \end{aligned}$$

Note that $\widehat{F}_{*M(-i)}(e_{Mk}^{(i)}) = \widehat{F}_{M(-i)}(e_{Mk})$. Therefore,

$$\forall i = 1, \dots, n_t, \quad \lambda_{*k}^{(i)} = \lambda_k^{(i)} = J_{(n-1)}\left(\widehat{F}_{M(-i)}(e_{Mk})\right),$$

and,

$$\begin{aligned} \bar{e}_{M(-i)}^{(\alpha_{n-1})} &= \frac{1}{n-1} \sum_{k \neq i} \lambda_k^{(i)} e_{Mk} - (\widehat{B}_{M(-i)} - \widehat{B}_M) \left(\frac{1}{n-1} \sum_{k \neq i} \lambda_k^{(i)} \right), \\ &= \int r J_{(n-1)}\left(\widehat{F}_{M(-i)}(r)\right) d\widehat{F}_{M(-i)}(r) - (\widehat{B}_{M(-i)} - \widehat{B}_M) + \\ &\quad + (\widehat{B}_{M(-i)} - \widehat{B}_M) \left(1 - \frac{1}{n-1} \sum_{k \neq i} \lambda_k^{(i)} \right). \end{aligned}$$

Consequently,

$$\begin{aligned} \bar{e}_{M(-i)}^{(\alpha_{n-1})} &= \int r dK_{(n-1)}\left(\widehat{F}_{M(-i)}(r)\right) - (\widehat{B}_{M(-i)} - \widehat{B}_M) + \\ &\quad + (\widehat{B}_{M(-i)} - \widehat{B}_M) \left(1 - \frac{1}{n-1} \sum_{k \neq i} \lambda_k^{(i)} \right). \end{aligned}$$

It follows from all this that

$$\begin{aligned} \bar{e}_{M(-i)}^{(\alpha_{n-1})} - \bar{e}_M^{(\alpha_n)} &= \int r d\left\{ K_{(n-1)}\left(\widehat{F}_{M(-i)}(r)\right) - K_n\left(\widehat{F}_M(r)\right) \right\} - (\widehat{B}_{M(-i)} - \widehat{B}_M) + \\ &\quad + (\widehat{B}_{M(-i)} - \widehat{B}_M) \left(1 - \frac{1}{n-1} \sum_{k \neq i} \lambda_k^{(i)} \right). \end{aligned}$$

By integrating by parts, we obtain,

$$\begin{aligned} \int r d\left\{ K_{(n-1)}\left(\widehat{F}_{M(-i)}(r)\right) - K_n\left(\widehat{F}_M(r)\right) \right\} &= \\ &= \lim_{a \rightarrow \infty} [h_{n(i)}(a) - h_{n(i)}(-a)] - \int [K_{(n-1)}\left(\widehat{F}_{M(-i)}(r)\right) - K_n\left(\widehat{F}_M(r)\right)] dr. \end{aligned}$$

where,

$$h_{n(i)}(r) = r \left[K_{(n-1)}\left(\widehat{F}_{M(-i)}(r)\right) - K_n\left(\widehat{F}_M(r)\right) \right].$$

But for an arbitrary i , $\lim_{a \rightarrow \infty} h_{n(i)}(a) = 0$. This follows from the fact that if we have

$a > \max_{1 \leq k \leq n_t} |e_{Mk}|$ then $h_{n(i)}(a) = 0$. Therefore,

$$\bar{e}_{M(-i)}^{(\alpha_{n-1})} - \bar{e}_M^{(\alpha_n)} = - \int [K_{(n-1)}\left(\widehat{F}_{M(-i)}(r)\right) - K_n\left(\widehat{F}_M(r)\right)] dr + \Delta_{(i)}.$$

where,

$$\Delta_{(i)} = -(\hat{B}_{M(-i)} - \hat{B}_M) + (\hat{B}_{M(-i)} - \hat{B}_M) \left(1 - \frac{1}{n-1} \sum_{k \neq i} \lambda_k^{(i)} \right).$$

Let z_i be given by

$$z_i = - \int \left(I(e_{Mi} \leq r) - \hat{F}_M(r) \right) J_n \left(\hat{F}_M(r) \right) dr.$$

It is straightforward to see that $\bar{z} = 0$. This implies that

$$\bar{z}_{(-i)} = \frac{1}{n-1} \sum_{k \neq i} z_k = -\frac{z_i}{n-1}.$$

Now we have that

$$\bar{e}_{M(-i)}^{(\alpha_{n-1})} - \bar{e}_M^{(\alpha_n)} - \bar{z}_{(-i)} = - \int W_{(-i)}(r) \left(\hat{F}_{M(-i)}(r) - \hat{F}_M(r) \right) dr + \Delta_{(i)},$$

where,

$$W_{(-i)}(r) = \begin{cases} 0, & \text{if } \hat{F}_M(r) = \hat{F}_{M(-i)}(r) \\ \frac{K_{(n-1)} \left(\hat{F}_{M(-i)}(r) \right) - K_n \left(\hat{F}_M(r) \right)}{\hat{F}_{M(-i)}(r) - \hat{F}_M(r)} - J_n \left(\hat{F}_M(r) \right), & \text{otherwise.} \end{cases}$$

Therefore,

$$\bar{e}_{M(-i)}^{(\alpha_{n-1})} - \bar{e}_M^{(\alpha_n)} = -\frac{z_i}{n-1} - \int W_{(-i)}(r) \left(\hat{F}_{M(-i)}(r) - \hat{F}_M(r) \right) dr + \Delta_{(i)}.$$

Now we have that

$$\begin{aligned} \hat{Y}_{LM(-i)} - \hat{Y}_{LM} &= \left(\hat{B}_{M(-i)} - \hat{B}_M \right) + w_n (\bar{e}_{M(-i)} - \bar{e}_M) + (1 - w_n) \left(\bar{e}_{M(-i)}^{(\alpha_{n-1})} - \bar{e}_M^{(\alpha_n)} \right) \\ &= -\frac{u_i}{n-1} - (1 - w_n) \int W_{(-i)}(r) \left(\hat{F}_{M(-i)}(r) - \hat{F}_M(r) \right) dr + \Delta_{(i)} + \\ &\quad + (1 - w_n) \left(\hat{B}_{M(-i)} - \hat{B}_M \right) \left(1 - \frac{1}{n-1} \sum_{k \neq i} \lambda_k^{(i)} \right), \end{aligned}$$

where $u_i = w_n(e_{Mi} - \bar{e}_M) + (1 - w_n)z_i$. Let,

$$\varphi_{n(i)} = \int W_{(-i)}(r) \left(\hat{F}_{M(-i)}(r) - \hat{F}_M(r) \right) dr \text{ and } \eta_{n(i)} = 1 - \frac{1}{n-1} \sum_{k \neq i} \lambda_k^{(-i)}.$$

It follows that

$$\begin{aligned} & \left(\hat{\bar{Y}}_{LM(-i)} - \hat{\bar{Y}}_{LM} \right)^2 = \\ &= \frac{u_i^2}{(n-1)^2} + (1 - w_n)^2 \varphi_{n(i)}^2 + (1 - w_n)^2 (\hat{B}_{M(-i)} - \hat{B}_M)^2 \eta_{n(i)}^2 + \\ &+ \frac{2(1 - w_n)}{n-1} u_i \varphi_{n(i)} - \frac{2(1 - w_n)}{n-1} u_i (\hat{B}_{M(-i)} - \hat{B}_M) \eta_{n(i)} - \\ &- 2(1 - w_n)^2 \varphi_{n(i)} \eta_{n(i)} (\hat{B}_{M(-i)} - \hat{B}_M). \end{aligned}$$

It follows from all this that,

$$\begin{aligned} \hat{V}_{JK}(\hat{\bar{Y}}_{LM}) &= \frac{1-f}{n} (n-1) \sum_{i=1}^n (\hat{\bar{Y}}_{LM(-i)} - \hat{\bar{Y}}_{LM})^2, \\ &= \frac{1-f}{n} \frac{1}{n-1} \sum_{i=1}^n u_i^2 + d_{1t} + d_{2t} + d_{3t}, \end{aligned}$$

where,

$$\begin{aligned} d_{1t} &= \frac{1-f}{n} (n-1) (1 - w_n)^2 \sum_{i=1}^n \varphi_{n(i)}^2, \\ d_{2t} &= \frac{1-f}{n} (n-1) (1 - w_n)^2 \sum_{i=1}^n (\hat{B}_{M(-i)} - \hat{B}_M)^2 \eta_{n(i)}^2, \\ d_{3t} &= 2 \frac{1-f}{n} (1 - w_n) \left\{ \sum_{i=1}^n u_i \varphi_{n(i)} - \sum_{i=1}^n u_i \eta_{n(i)} (\hat{B}_{M(-i)} - \hat{B}_M) - \right. \\ &\quad \left. - \sum_{i=1}^n (1 - w_n) (n-1) \varphi_{n(i)} \eta_{n(i)} (\hat{B}_{M(-i)} - \hat{B}_M) \right\}. \end{aligned}$$

Note that

$$\begin{aligned} |d_{3t}| &\leq 2 \left\{ \left(\frac{1}{n} \sum_{i=1}^n u_i^2 \right)^{1/2} \left(\frac{1}{n} \sum_{i=1}^n \varphi_{n(i)}^2 \right)^{1/2} + \right. \\ &\quad + \left(\frac{1}{n} \sum_{i=1}^n u_i^2 \right)^{1/2} \left(\frac{1}{n} \sum_{i=1}^n (\hat{B}_{M(-i)} - \hat{B}_M)^2 \right)^{1/2} \times \max_{1 \leq i \leq n} |\eta_{n(i)}| + \\ &\quad \left. + \left(\sum_{i=1}^n \varphi_{n(i)}^2 \right)^{1/2} \left(\sum_{i=1}^n (\hat{B}_{M(-i)} - \hat{B}_M)^2 \right)^{1/2} \times \max_{1 \leq i \leq n} |\eta_{n(i)}| \right\}. \end{aligned}$$

Note that $\widehat{V}_{JK}(\widehat{Y}_{LM}) = \widehat{V}_{TL}(\widehat{Y}_{LM}) + d_{1t} + d_{2t} + d_{3t}$. It follows from this that

$$\widehat{V}_{JK}(\widehat{Y}_{LM}) / \widehat{V}_{TL}(\widehat{Y}_{LM}) \approx \widehat{V}_{TL}(\widehat{Y}_{LM}) / V_a(\widehat{Y}_{LM}) + (d_{1t} + d_{2t} + d_{3t}) / V_a(\widehat{Y}_{LM})$$

It follows from theorem 4.4 that $\widehat{V}_{TL}(\widehat{Y}_{LM}) / V_a(\widehat{Y}_{LM}) - 1 = o_p(1)$ and from assumption (iii) that $V_a(\widehat{Y}_{LM}) = O(1/n_t)$. Therefore, to complete the proof it is enough to prove that $d_{1t} + d_{2t} + d_{3t} = o_p(1/n_t)$. But the inequality above, involving $|d_{3t}|$ shows that if $d_{1t} = o_p(1/n_t)$ and $d_{2t} = o_p(1/n_t)$ then $d_{3t} = o_p(1/n_t)$.

- $d_{1t} = o_p(1/n_t)$.

It follows from condition C9 that there exists two constants c_1 and c_2 such that $\forall t \geq 1$, $F_{M_t}(c_1) = 0$ and $F_{M_t}(c_2) = 0$. Since $\widehat{F}_{M_t}(c_1) - F_{M_t}(c_1) = o_p(1)$ and $\widehat{F}_{M_t}(c_2) - F_{M_t}(c_2) = o_p(1)$, it is enough to prove that $\lambda_{1t}^* = o_p(1/n_t)$, where

$$\lambda_{1t}^* = \frac{1-f}{n_t} (n_t - 1) (1 - w_n)^2 \sum_{i=1}^{n_t} \varphi_{n(i)}^{*2},$$

where

$$\varphi_{n(i)}^* = \int_{c_1}^{c_2} W_{t(-i)}(r) \left(\widehat{F}_{M(-i)}(r) - \widehat{F}_M(r) \right)^2 dr.$$

We have that

$$\varphi_{n(i)}^{*2} \leq \left(\int_{c_1}^{c_2} W_{t(-i)}^2(r) dr \right) \left(\int_{c_1}^{c_2} \left(\widehat{F}_{M(-i)}(r) - \widehat{F}_M(r) \right)^2 dr \right).$$

It follows from the proof of lemma 2.5 in chapter 2 that

$$\int_{c_1}^{c_2} W_{t(-i)}^2(r) dr = o_p(1), \text{ uniformly in } i.$$

Furthermore, we have that

$$\widehat{F}_{M(-i)}(r) = \widehat{F}_M(r) + \frac{\widehat{F}_M(r) - I(e_{Mi} \leq r)}{n_t - 1}.$$

It follows from this that

$$\left(\widehat{F}_{M(-i)}(r) - \widehat{F}_M(r) \right)^2 \leq \frac{1}{(n_t - 1)^2} (c_2 - c_1),$$

which implies

$$\lambda_{1t}^* \leq o_p(1) \cdot \frac{1-f}{n_t} \cdot (n_t - 1) \frac{n_t}{(n_t - 1)^2} = o_p(1/n_t).$$

- $d_{2t} = o_p(1/n_t).$

It follows from the definition of d_{2t} that

$$|d_{2t}| \leq \left(\frac{1-f}{n_t} (n_t - 1) \sum_{i=1}^{n_t} \left(\widehat{B}_{M(-i)} - \widehat{B}_M \right)^2 \right) \times \left(\max_{1 \leq i \leq n_t} \eta_{n(i)}^2 \right).$$

The first term of the above inequality is nothing else that the jackknife variance estimator of $\widehat{B}_M$. It follows from the conditions $C1'$, $C2'$, $C3'$ and the proof of theorem 4.2 that this jackknife estimator is of order $O_p(1/n_t)$. Furthermore the last term on the right-hand side of the above inequality is $o_p(1)$. This follows from lemma 2.4 in chapter 2.

■

4.5. Monte Carlo Study

In order to investigate the small sample performance of the jackknife variance estimators proposed in this chapter, we have undertaken a small Monte Carlo study. This section is entirely devoted to the presentation of this simulation study.

4.5.1. Objectives.

Some consistency results were proved in section 4.3. One objective of this simulation study is to support these findings with some empirical results. The second objective of the study is to confirm the superiority of the jackknife estimator for estimating the variance of the CM-estimator. This is a conjecture made earlier and due to the fact that in the Taylor series method, $\hat{\sigma}_{\text{CM}}$ is treated as a constant. This is likely to yield an underestimation of the variance very often. The third objective is to investigate whether the jackknife method works as well as the Taylor series method for estimating the variance of the LM-estimator.

All these three objectives can be achieved by calculating the percentage relative bias of the MSE (resp. variance) estimators, with respect to the “true” MSE (resp. variance). Let us assume the estimator under study is the CM-estimator. Since the “true” MSE is difficult to obtain, it will be replaced with its Monte Carlo

approximation given by

$$\text{MSE}_{\text{MC}}(\hat{\bar{Y}}_{\text{CM}}) = \frac{1}{m} \sum_{s=1}^m (\hat{\bar{Y}}_{\text{CM}}(s) - \bar{Y})^2,$$

where m is the total number of Monte Carlo samples, $\bar{Y}$ the parameter of interest and $\hat{\bar{Y}}_{\text{CM}}(s)$ the estimate of $\bar{Y}$ based on the particular Monte Carlo sample s . Likewise, the “true” variance is given by

$$V_{\text{MC}}(\hat{\bar{Y}}_{\text{CM}}) = \frac{1}{m} \sum_{s=1}^m [\hat{\bar{Y}}_{\text{CM}}(s) - E_{\text{MC}}(\hat{\bar{Y}}_{\text{CM}})]^2, \text{ where } E_{\text{MC}}(\hat{\bar{Y}}_{\text{CM}}) = \frac{1}{m} \sum_{s=1}^m \hat{\bar{Y}}_{\text{CM}}(s).$$

Suppose that $\text{mse}(\hat{\bar{Y}}_{\text{CM}})$ is an estimator of the MSE to be investigated. The percentage relative bias of that estimator is given by

$$RB(\text{mse}(\hat{\bar{Y}}_{\text{CM}})) = 100 \times \left(\frac{E_{\text{MC}}[\text{mse}(\hat{\bar{Y}}_{\text{CM}})]}{\text{MSE}_{\text{MC}}(\hat{\bar{Y}}_{\text{CM}})} - 1 \right).$$

Similarly the relative bias of $\hat{V}(\hat{\bar{Y}}_{\text{CM}})$ as an estimator of the CM-estimator variance, is given by

$$RB(\hat{V}(\hat{\bar{Y}}_{\text{CM}})) = 100 \times \left(\frac{E_{\text{MC}}[\hat{V}(\hat{\bar{Y}}_{\text{CM}})]}{V_{\text{MC}}(\hat{\bar{Y}}_{\text{CM}})} - 1 \right).$$

In fact, these two quantities will be calculated for the variance estimators of $\hat{\bar{Y}}_{\text{CM}}$ and $\hat{\bar{Y}}_{\text{LM}}$ based on the Taylor series method as well as the jackknife method. The performances of the variance estimators are evaluated only in terms of the bias.

4.5.2. Description of the Experiment.

The simulation results reported here are based on a real population provided by Särndal et al. (1992). This population consists of the annual revenues of the 284 municipalities in Sweden. The revenues are from the 1985 municipal taxation. This population has the particularity of being extremely skewed because of the presence of three large municipalities.

The population mean which is the parameter of interest is $\bar{Y} = 244.99$. The population standard deviation is $S_y = 596.31$ where

$$S_y^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{Y})^2.$$

The parameter B_M is equal to 151.71. The difference between B_m and $\bar{Y}$ confirms the skewness of the population.

For the CM-estimator, we have taken $\lambda_0 = 10$ again. As mentioned in chapter 2, it is also possible to determine that number in a two-step procedure using the most recent census. The first step will be to determine the smallest sample size n_0 necessary for the GREG estimator to have an acceptable low CV. The second step is to determine the value of λ_0 so as to control the efficiency of the CM-estimator at $n = n_0$ with respect to the GREG estimator.

This parameter does not have to be determined with great precision as it cannot have a great impact on the efficiency of the robust estimators.

Our results are based on 10,000 simple random samples. In selecting the samples, we have used the following implementation of the SRSWOR design:

- For a population of size N we have selected N independent uniform random numbers $\varepsilon_1, \dots, \varepsilon_N$ in the interval $(0,1)$.
- These numbers are then sorted in ascending order and let the ε_k^* 's be the ordered numbers and r_k the rank of ε_k^* in the original sample. Then our simple random sample is given by the population units $(r_1, \dots, r_n)$.

4.5.3. Simulation Results.

Table 4.1 indicates that for this particular population under study, the robust estimators have a substantial bias. As mentioned previously, this is due to the fact that the population is heavily skewed. Because the bias is important, this study provides a good opportunity to see how well the squared bias can be estimated and how reliable MSE estimators are.

The third column of the table indicates the relative bias of the estimators obtained by dividing the bias by the population mean. The relative bias of the CM-estimator ranges from -20.84% for a sample size of 20 to -15.82% for a sample size of 60, while that of the LM-estimator ranges from -26.56% to -16.54%. Therefore, the bias of the LM-estimator is a little higher than that of the CM-estimator. However, in spite of these large biases, the two robust estimators substantially outperform the

GREG estimator in terms of the MSE. For a sample of size 20, the MSE of the GREG estimator is 16,639.70 while the CM and the KM estimators have MSE's of 6,716.08 and 7,480.55 respectively. The MSE of the GREG estimator is about 2.5 times that of the CM-estimator and about 2.2 times that of the LM-estimator. For larger sample sizes as 60, the MSE of the GREG estimator can be as high as 1.5 times those of the robust estimators. This means that on average, the robust estimates will be closer to the true value than the GREG estimates. On the other hand, the confidence intervals constructed with the GREG estimator will be more reliable in terms of coverage than those constructed with the two robust estimators.

Turning to table 4.2, we see immediately a confirmation of the fact the jackknife method provides a much better estimator of the asymptotic variance of the CM-estimator. The R.B. of $\hat{V}_{JK}(\hat{\bar{Y}}_{CM})$ ranges from 4.10 to 1.18 while that of $\hat{V}_{TL}(\hat{\bar{Y}}_{CM})$ ranges from -12.24 to -11.04 when the sample size goes from 20 to 60. This clearly shows that the Taylor series method leads to a substantial underestimation of the asymptotic variance of the CM-estimator. As mentioned before, for the Taylor series approach to be tractable, we have to assume that the scale parameter is constant and known, which is not true. These results also confirm the consistency of the jackknife variance estimator as the RB goes to 0 as the sample size n increases.

Concerning the MSE estimators of the CM-estimator, we notice from table 4.2 that the RB of the MSE estimator based on the jackknife method has a bias that is negligible. The estimator based on the Taylor series method however has a

nonnegligible bias. This again stems from the poor variance estimator previously mentioned.

The problem raised earlier about the scale parameter is not relevant for the LM-estimator. Therefore, there is no reason to believe that the jackknife method will outperform the linearization method. In fact table 4.2 shows that the two methods lead to very similar results. Both methods yield an R.B. for the variance estimator ranging roughly from 13 to 10. Concerning the MSE estimator, the R.B. is almost negligible. These simulation results also confirm the consistency of the proposed MSE estimators.

Table 4.1 Monte Carlo Biases, Variances and Mean Squared Errors of the
 GREG, CM and LM estimators based on 10,000 replicates.

Estimator	Sample Size	Relative Bias(%)	Variance	MSE
GREG	20	-0.30	16,639.70	16,639.70
	40	-0.22	7,624.92	7,624.92
	60	-0.22	4,640.93	4,640.93
CM	20	-20.84	4,110.12	6,716.08
	40	-18.29	2,158.52	4,167.05
	60	-15.82	1,579.12	3,081.14
LM	20	-26.56	3,245.48	7,480.55
	40	-20.39	2,223.78	4,719.19
	60	-16.54	1,797.52	3,440.12

Table 4.2 Percentage Relative Biases of the Variance Estimators Based on the Jackknife and the Taylor Series Methods. Results Based on 10,000 Replicates.

Percentage Relative Bias	Sample Size		
	20	40	60
$RB[\hat{V}_{JK}(\hat{Y}_{CM})]$	4.10	2.62	1.18
$RB[\hat{V}_{TL}(\hat{Y}_{CM})]$	-12.24	-11.64	-11.04
$RB[mse_{JK}(\hat{Y}_{CM})]$	0.31	-0.52	-1.21
$RB[mse_{TL}(\hat{Y}_{CM})]$	-10.63	-10.92	-10.29
$RB[\hat{V}_{JK}(\hat{Y}_{LM})]$	13.19	13.31	9.92
$RB[\hat{V}_{TL}(\hat{Y}_{LM})]$	12.69	13.44	10.16
$RB[mse_{JK}(\hat{Y}_{LM})]$	-4.40	-1.54	-1.26
$RB[mse_{TL}(\hat{Y}_{LM})]$	-3.72	-0.98	-0.81

CHAPTER 5

Robust Estimation of Small Area Means Under the One-Fold Nested Error Regression Model

5.1. Introduction

The demand for reliable statistics related to small areas from the public and the private sectors has grown substantially during the past few years. The term “small area” is used to designate a small geographical area such as a county, a municipality or a census division. A “small domain” usually designates a small subpopulation such as a specific age-sex group of people within a large geographical area. However, the estimation methods developed for small area estimation are applicable to small domain estimation as well. For that reason, the two terms “small area” and “small domain” are used interchangeably in this chapter.

The problem addressed in this chapter is to estimate the small area means. Because of the smallness of the sample size in areas, standard survey procedures are likely to yield estimates with unduly large variances.

Statisticians have been aware of this problem for a long time and have proposed various alternative procedures. Almost all of these procedures rely upon the idea of “borrowing strength” from related domains to obtain more stability to the estimators. Borrowing strength from other domains is usually done through modeling where the use of the small area model is implicit or explicit.

Prasad and Rao (1986) have studied extensively the one-fold nested-error regression model under the assumption of constant error variance. Stukel (1991) extended that model to the case where the error variance is proportional to a known auxiliary variable and derived the best linear unbiased predictor (BLUP) of the domain mean. In practice however none of these models can be expected to be fully satisfactory. A more realistic assumption would be that an unknown proportion of population units follow a model with constant error variance, while the remaining units follow a model with an error variance with a complex structure.

Our goal in this chapter is to propose an estimator that will be robust to ill-specified structures for the error variance. The efficiency of the proposed estimators will be close to that of the BLUP estimator under a model with a given structure for the error variance. But the mean squared error (MSE) of the robust estimators will be much smaller than that of the BLUP estimators and the other existing procedures

if the error variance is unknown or misspecified.

It will be assumed throughout this chapter that we are dealing with a population $\mathcal{U}$ of size N that is made up of m domains some of which are fairly small. Each domain i is denoted by $\mathcal{U}_i$ and contains N_i population units. Therefore,

$$\mathcal{U} = \bigcup_{i=1}^m \mathcal{U}_i \text{ and } N = \sum_{i=1}^m N_i.$$

A sample s of size n is drawn according to a certain design. The subsample s_i of sample units that fall into $\mathcal{U}_i$ is defined by $s_i = s \cap \mathcal{U}_i$ and contains n_i sample units. Therefore,

$$s = \bigcup_{i=1}^m s_i \text{ and } n = \sum_{i=1}^m n_i.$$

Although a number of approaches are available for estimating small domain means, we will focus on the model-based approach that relies entirely on a small area model. The parameter of interest, which is the domain mean, is expressed as a function of the unknown model parameters. An estimator of the domain mean is then obtained by replacing the unknown parameters by the corresponding estimators.

Many small area models in the literature are special cases of the following general mixed linear model:

$$(\xi_g) \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \sum_{k=1}^r \mathbf{Z}_k \mathbf{v}_k + \mathbf{e},$$

where the $\mathbf{v}_k$'s are the random effects, r the number of random effects and the $\mathbf{Z}_k$'s.

are known $n \times q_k$ -dimensional matrices defined by

$$\mathbf{Z}_k = \begin{pmatrix} \mathbf{z}_1^{(k)} & \mathbf{0}_{n_1} & \cdots & \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} & \ddots & & \\ \vdots & & \ddots & \\ \mathbf{0}_{n_m} & \cdots & \mathbf{0}_{n_m} & \mathbf{z}_{q_k}^{(k)} \end{pmatrix}$$

and

$$\mathbf{y} = \underset{1 \leq i \leq m}{\text{col}} \underset{1 \leq j \leq n_i}{\text{col}} (y_{ij}), \quad \mathbf{X} = \underset{1 \leq i \leq m}{\text{col}} (\mathbf{X}_i), \quad \mathbf{X}_i = \underset{1 \leq j \leq n_i}{\text{col}} (\mathbf{x}_{ij}^\top),$$

$$\mathbf{z}_i^{(k)} = \underset{1 \leq j \leq n_i}{\text{col}} (z_{ij}^{(k)}), \quad E(\mathbf{v}) = \mathbf{0}, \quad V(\mathbf{v}) = \mathbf{G}, \quad E(\mathbf{e}) = \mathbf{0},$$

$$V(\mathbf{e}) = \mathbf{R}.$$

The vectors $\mathbf{v}_i$'s and $\mathbf{e}$ are distributed independently, $\mathbf{X}$ and $\mathbf{Z}$ are known matrices ($\mathbf{Z} = \underset{1 \leq i \leq r}{\text{row}} (\mathbf{z}_i)$).

As mentioned by Robinson (1991), the outliers can occur on the random effects $\mathbf{v}_i$'s as well as on the error $\mathbf{e}$. The robust procedures developed by Fellner (1986) are designed to deal with both types of outliers. However, his main concern was to estimate the variance components in the particular case where

$$\mathbf{G} = \text{diag}(\sigma_{v_i}^2, i = 1, \dots, m)$$

and $\mathbf{R} = \sigma_e^2 \mathbf{I}_n$. In such a case and under any model derived from (ξ_g) , modifying the realization of a random effect will not have a direct impact on the performance of the resulting robust estimator as the other random effects are involved in the estimation as well. In small area estimation, however, where the parameters of interest are sometimes specific to a given small area, an outlier on a random effect is likely to have a significant influence on the true value of the parameter. That influence should be reflected in the estimators as well.

In section 5.2, the small area model that will be the focus of this study is introduced. It is in fact the well-known two-fold nested error regression model. In section 5.3, under the assumption of known variance components, we propose a robust procedure for estimating a small area mean. The proposed procedure is based on the M-estimation techniques of Huber (1964), Hampel et al. (1986). Because of the type of model we will be focusing on and the outlier problem on random effects mentioned earlier, the proposed procedure is expected to perform very well if the outliers occur mainly in the errors. In section 5.4, we assume that the variance components are unknown and propose an extension of the theory developed in section 5.3. Section 5.5 presents the results of a simulation study.

We introduce the following notation that will be used in the next section:

$$\mathbf{X}_P = \text{col}_{1 \leq i \leq m} (\mathbf{X}_{Pi}), \text{ where } \mathbf{X}_{Pi} = \text{col}_{1 \leq j \leq N_i} (\mathbf{x}_{ij}^\top).$$

5.2. The Small Area Model

We confine ourselves to the nested-error regression model of Battese, Harter and Fuller (1988). It is assumed that there are m small areas under investigation, each small area i contains N_i population units. The number of sample units belonging to the i th domain is n_i . Suppose that for each small domain i and each individual j belonging to that domain, there exists a known p -dimensional vector $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp})^\top$ to be used as auxiliary information. The model suggested

by these authors includes a random variable v_i which represents the i -th domain effect. That model is given by

$$(\xi_0) \quad \begin{cases} y_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\beta} + v_i + e_{ij}, & i = 1, \dots, m, j = 1, \dots, n_i, \text{ where} \\ E_{\xi_0}(v_i) = 0, V_{\xi_0}(v_i) = \sigma_v^2, \\ E_{\xi_0}(e_{ij}) = 0, V_{\xi_0}(e_{ij}) = \sigma_e^2, \text{ and} \\ E_{\xi_0}(v_i e_{ij}) = 0. \end{cases}$$

In the above model, y_{ij} is the variable of interest for the sample unit j belonging to the i th domain and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ is the p -dimensional vector of unknown parameters. The random variables v_i 's and e_{ij} 's are assumed to be independent and identically distributed (i.i.d.).

It is clear that for each realization of the variable y , there are corresponding realizations of v and e that will still be denoted by v and e for simplicity. Our goal will be to derive a robust estimator of the domain mean when the variance components are known. An approximation of the mean squared error (MSE) will be derived as well and an estimator of that approximation proposed.

5.3. Robust Estimation of the Domain Mean Under the Assumption of Known Variance Components

In the next section, two robust alternatives to the customary ξ_0 -BLUP estimator of the domain mean will be introduced for the infinite population case. The extension to the finite population case will be discussed in the section after.

5.3.1. The Infinite Population Case

If the domain size N_i is large, then it follows from model (ξ_0) and the law of large numbers that the mean of the i th domain is given by

$$\mu_i \approx \bar{\mathbf{X}}_i^\top \boldsymbol{\beta} + v_i,$$

where

$$\bar{\mathbf{X}}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \mathbf{x}_{ij}.$$

This follows from the fact that

$$\bar{E}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} e_{ij} \xrightarrow{P} 0$$

for large values of N_i . Henderson (1963) used a generalization of the Gauss-Markov theorem and proposed the Best Linear Prediction (BLUP) estimator of μ_i . It is given by

$$\hat{\mu}_{Bi} = \bar{\mathbf{X}}_i^\top \hat{\boldsymbol{\beta}}_0 + \hat{v}_{0i},$$

where $\hat{\boldsymbol{\beta}}_0$ is the Best Linear Unbiased Estimator (BLUE) of $\boldsymbol{\beta}$ and $\hat{v}_{0i}$ the BLUP estimator of v_i . These two estimators are given by:

$$\begin{cases} \hat{\boldsymbol{\beta}}_0 = \left[\sum_{i=1}^m \left(\sum_{j=1}^{n_i} \mathbf{x}_{ij} \mathbf{x}_{ij}^\top - n_i \alpha_i \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^\top \right) \right]^{-1} \sum_{i=1}^m \left(\sum_{j=1}^{n_i} \mathbf{x}_{ij} y_{ij} - n_i \alpha_i \bar{\mathbf{x}}_i \bar{y}_i \right), \\ \hat{v}_{0i} = \alpha_i (\bar{y}_i - \bar{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}_0), \end{cases}$$

where

$$\alpha_i = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_e^2/n_i}, \text{ and } \bar{\mathbf{x}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{x}_{ij}.$$

These two estimators are solutions to the following “normal equations”:

$$\begin{cases} \sum_{i=1}^m \sum_{j=1}^{n_i} \sigma_e^{-1} \mathbf{x}_{ij} \left(\frac{y_{ij} - (\mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_0 + \hat{v}_{0i})}{\sigma_e} \right) = 0, \\ \hat{v}_{0i} = \alpha_i (\bar{y}_i - \bar{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}_0) \text{ for } i = 1, \dots, m. \end{cases} \quad (5.1)$$

These equations are not the normal equations in a strict sense, but following Fellner (1986) we will keep the same terminology.

The vector of variance components $\boldsymbol{\sigma}^{(2)} = (\sigma_v^2, \sigma_e^2)$ is assumed here to be known. But in practice, it must be estimated and the resulting small area mean estimator is called the Empirical Best Linear Unbiased Prediction (EBLUP) estimator.

Because the random component v_i cannot be estimated consistently, the performance of small area estimators is measured by the mean squared error (MSE). Using Prasad-Rao notation (see Prasad and Rao, 1990), the MSE of $\hat{\mu}_{Bi}(\boldsymbol{\sigma}^{(2)})$ is given by

$$MSE\left\{\hat{\mu}_{Bi}(\boldsymbol{\sigma}^{(2)})\right\} = g_{1i}(\boldsymbol{\sigma}^{(2)}) + g_{2i}(\boldsymbol{\sigma}^{(2)})$$

where $g_{1i}(\boldsymbol{\sigma}^{(2)}) = (1 - \alpha_i)\sigma_v^2$,

$$g_{2i}(\boldsymbol{\sigma}^{(2)}) = (\bar{\mathbf{X}}_i - \alpha_i \bar{\mathbf{x}}_i)^\top \mathbf{Q}_b^{-1} (\bar{\mathbf{X}}_i - \alpha_i \bar{\mathbf{x}}_i) / n$$

and

$$\mathbf{Q}_b = \frac{1}{n} \sum_{i=1}^m \left(\sum_{j=1}^{n_i} \mathbf{x}_{ij} \mathbf{x}_{ij}^\top - n_i \alpha_i \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^\top \right).$$

Let $\hat{e}_{0ij}$ be defined by $\hat{e}_{0ij} = y_{ij} - (\mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_0 + \hat{v}_{0i})$. For a good-fitting model, one would expect the $\hat{e}_{0ij}$'s as well as the $\hat{v}_{0i}$'s to be fairly close to 0. But after solving the above normal equations, one may find that some of these quantities are much greater than 0 in absolute value. The outliers among the v_i 's define a new type of outlying units mentioned by Fellner (1986) and Robinson (1991), that do not exist in standard regression models.

Because v_i represents the i th domain effect, an outlying value for $\hat{v}_i$ might suggest that the entire domain is outlying with respect to the other domains. This problem was pointed out by Robinson (1991) in a more general setting and suggested to look at the $\hat{v}_i$'s instead of just looking at the residuals $\hat{e}_{ij}$. Moreover, an outlying value for one $\hat{v}_i$ is likely to affect the other estimates as well as the estimator $\hat{\boldsymbol{\beta}}_0$ of $\boldsymbol{\beta}$. It is then essential to reduce their influence.

In order to develop robust estimators of $\boldsymbol{\beta}$ and the v_i 's, the theory of M-statistics (see Huber, 1964) suggests to replace the least squares principle by the bounded influence principle. We propose two robust alternatives to the BLUP estimator $\hat{\mu}_{Bi}$ of μ_i . The first alternative can be used if the practitioner does not expect high leverage points (outliers in the design space) and the second alternative is to be used if there is some concern about such points.

The first robust estimator of μ_i , which will be denoted by $\hat{\mu}_{Mi}$, is the M-estimator

given by,

$$\hat{\mu}_{Mi} = \bar{\mathbf{X}}_i^\top \hat{\boldsymbol{\beta}}_M + \alpha_i (\bar{y}_i - \bar{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}_M), \quad (5.2)$$

where $\hat{\boldsymbol{\beta}}_M$ is solution to the following system of equations:

$$\begin{cases} (m1) & \sum_{i=1}^m \sum_{j=1}^{n_i} \sigma_e^{-1} \mathbf{x}_{ij} \psi_H \left(\frac{y_{ij} - (\mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_M + \hat{v}_{Mi})}{\sigma_e} \right) = 0, \\ (m2) & \hat{v}_{Mi} = \alpha_i \bar{\sigma}_i \psi_H [(\bar{y}_i - \bar{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}_M) / \bar{\sigma}_i], \text{ for } i = 1, \dots, m. \end{cases} \quad (5.3)$$

where $\bar{\sigma}_i^2 = \sigma_v^2 + \sigma_e^2/n_i$. The influence function ψ_H is the Huber function defined for any real number t by.

$$\psi_H(t) = \text{sgn}(t) \min(K_H, |t|),$$

where K_H is the robustness constant. Another possibility that can be considered as well is the “fair” function ψ_F (see Rey, 1983) defined for any real number t by

$$\psi_F(t) = \frac{t}{1 + |t|/K_F},$$

where K_F is the robustness constant. To achieve approximately a 95% efficiency under normality, it is usually suggested to take K_H between 1 and 2. We can use $K_H = 2$ as a general-purpose value.

It is clear that the proposed M-estimator is likely to be vulnerable to high leverage points. In order to solve this problem we will use the generalized M-estimation (GME) technique and propose the GM-estimator $\hat{\mu}_{Gi}$ of μ_i given by

$$\hat{\mu}_{Gi} = \bar{\mathbf{X}}_i^\top \hat{\boldsymbol{\beta}}_G + \alpha_i (\bar{y}_i - \bar{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}_G), \quad (5.4)$$

where $\hat{\boldsymbol{\beta}}_G$ is solution to the following system of equations:

$$\begin{cases} (g1) & \sum_{i=1}^m \sum_{j=1}^{n_i} \sigma_e^{-1} d_{ij} \mathbf{x}_{ij} \psi_H \left(\frac{y_{ij} - (\mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_G + \hat{v}_{Gi})}{d_{ij} \sigma_e} \right) = 0, \\ (g2) & \hat{v}_{Gi} = \alpha_i \bar{\sigma}_i \psi_H [(\bar{y}_i - \bar{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}_G) / \bar{\sigma}_i], \forall i. \end{cases} \quad (5.5)$$

Note that d_{ij} is a function that depends on $\mathbf{x}_{ij}$. Its role is to dampen the impact of high leverage points on the M-estimates. It is therefore natural for d_{ij} to be a decreasing function of a measure of leverage. We propose the following function:

$$d_{ij}^2 = \min \left(1, \frac{1 + \chi_{p-\lambda, 0.95}^2}{NH_{ij,ij}} \right)$$

where

$$H_{ij,ij} = \mathbf{x}_{ij}^\top (\mathbf{X}_P^\top \mathbf{X}_P)^{-1} \mathbf{x}_{ij} \quad (5.6)$$

and

$$\mathbf{X}_P = \text{col}_{1 \leq i \leq m} \text{col}_{1 \leq j \leq N_i} \mathbf{x}_{ij}^\top.$$

We have $\lambda = 1$ or $\lambda = 0$ according as the model ξ_0 has an intercept or not. To motivate this choice, we note that if the population is reasonably large, $NH_{ij,ij} - 1$ is equivalent to the squared Mahalanobis distance of the j th unit in the i th small area which is usually compared to the quantiles of the chi-squared distribution with $p - \lambda$ degrees of freedom.

The robustness constant used for GM-estimators are usually slightly greater than those used for standard M-estimators. This is necessary as the standardized residuals for GM-estimators are further divided by a d-factor d_{ij} which lies between 0 and

1. Its effect is to increase the standardized residual which might become an artificial outlier if the robustness constant remains the same. All robustness constants proposed in the literature on robustness are between 2.0 and 3.0. The GM-estimator we propose here is close to the Schweppe-type GM-estimator (see Hampel et al., 1986). An important feature of this estimator is that the influence of the design vector $\mathbf{x}_{ij}$ is reduced only when the residual is large, otherwise it remains unchanged.

5.3.2. The Finite Population Case

We assume in this section that N_i can be small. In this case the population mean $\bar{Y}_i$ can no longer be approximated by μ_i as in the previous section. Estimators of $\bar{Y}_i$ are usually constructed by noting that

$$\bar{Y}_i = f_i \bar{y}_i + (1 - f_i) \bar{y}_i^* \text{ where } \begin{cases} f_i = n_i/N_i, \\ \bar{y}_i^* = \frac{1}{N_i - n_i} \sum_{j \in s'_i} y_{ij}, \\ s'_i = \mathcal{U}_i \setminus s_i. \end{cases}$$

Therefore, finding a robust estimator of $\bar{Y}_i$ amounts to finding a robust estimator of $E_{\xi_0}(\bar{y}_i^* | v_i)$. Note that

$$E_{\xi_0}(\bar{y}_i^* | v_i) = \bar{\mathbf{X}}_i^{*\top} \boldsymbol{\beta} + v_i \text{ where } \bar{\mathbf{X}}_i^* = (N_i - n_i)^{-1} \sum_{j \in s'_i} \mathbf{x}_{ij},$$

which is fairly close to the domain mean in the infinite population case. The proposed robust estimator of $\bar{Y}_i$ is given by:

$$\widehat{\bar{Y}}_{Gi} = f_i \bar{y}_i + (1 - f_i) [\bar{\mathbf{X}}_i^{*\top} \widehat{\boldsymbol{\beta}}_G + \alpha_i (\bar{y}_i - \bar{\mathbf{x}}_i^{*\top} \widehat{\boldsymbol{\beta}}_G)],$$

where $\hat{\boldsymbol{\beta}}_G$ and $\hat{v}_{Gi}$ are defined as in (5.5). Thus, the finite population case and the infinite population case are similar in nature but for simplicity, only the infinite population case will be treated.

5.3.3. Computation of the Robust Estimators

In this section, we present the algorithm for computing the robust estimators introduced in the two previous sections. Whether we are dealing with the finite population case or with the infinite population case, the estimation problem amounts to calculating the robust estimate of the parameter $\boldsymbol{\beta}$.

Since the GM-estimator is a generalization of the M-estimator, the algorithm will be presented for the GM-estimator only. The corresponding algorithm for the M-estimator is easily obtained by replacing d_{ij} by 1. Our problem is to estimate $\boldsymbol{\beta}$.

Let the robust residuals e_{gij} and their standardized version $\bar{e}_{gij}$ be denoted by

$$e_{gij} = y_{ij} - (\mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_G + \hat{v}_{Gi}) \text{ and } \bar{e}_{gij} = \frac{e_{gij}}{\sigma_e d_{ij}}.$$

It follows from (g1) of (5.5) that

$$\hat{\boldsymbol{\beta}}_G = \left(\sum_{i=1}^m \sum_{j=1}^{n_i} w_{ij} \mathbf{x}_{ij} \mathbf{x}_{ij}^\top \right)^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} w_{ij} \mathbf{x}_{ij} (y_{ij} - \hat{v}_{Gi}), \quad (5.7)$$

where

$$\hat{v}_{Gi} = \alpha_i \bar{\sigma}_i \psi_H \left[(\bar{y}_i - \bar{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}_G) / \bar{\sigma}_i \right] \text{ and } w_{ij} = \begin{cases} 1, & \text{if } \bar{e}_{gij} = 0, \\ \frac{\psi_H(\bar{e}_{gij})}{\bar{e}_{gij}} & \text{otherwise} \end{cases}.$$

In order to calculate the GM-estimator, we propose the following algorithm:

- (1) Obtain the values d_{ij} .
- (2) Obtain the least-squares estimators $\hat{\boldsymbol{\beta}}_0$ and $\hat{v}_{0i}$, $i = 1, \dots, m$.
- (3) Using the least-squares estimators of the previous step, calculate the initial residuals $\hat{e}_{ij}^{(0)} = y_{ij} - (\mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_0 + \hat{v}_{0i})$ and the weights $w_{ij}^{*(0)}$ using the formulas defined previously.
- (4) In order to calculate the main estimates $\hat{\boldsymbol{\beta}}_G$ and $\hat{v}_{Gi}$, proceed as follows:
Calculate from $g2$ of (5.5)

$$\hat{v}_{Gi}^{(t+1)} = \alpha_i \bar{\sigma}_i \psi_H [(\bar{y}_i - \bar{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}_G^{(t)}) / \bar{\sigma}_i]$$

and $\hat{\boldsymbol{\beta}}_G^{(t+1)}$ from formula (5.7) by replacing w_{ij} by $w_{ij}^{(t)}$ and $\hat{v}_{Gi}$ by $\hat{v}_{Gi}^{(t)}$. Continue the process until the following two conditions are satisfied:

$$\|\hat{\boldsymbol{\beta}}_G^{(t+1)} - \hat{\boldsymbol{\beta}}_G^{(t)}\| < \epsilon_0$$

for some desired precision ϵ_0 . The norm $\|\cdot\|$ might be the euclidean norm or any other equivalent norm. For our simulations, we will use the sup norm $\|\cdot\|_\infty$. Therefore,

$$\|\hat{\boldsymbol{\beta}}_G^{(t+1)} - \hat{\boldsymbol{\beta}}_G^{(t)}\|_\infty = \max_{1 \leq k \leq p} |\hat{\beta}_{gk}^{(t+1)} - \hat{\beta}_{gk}^{(t)}|.$$

5.3.4. Asymptotic Properties of the GM-estimators

Since the GM-estimator defined by (5.4) is not a linear function, it is difficult to derive exact bias and variance under a given model. Instead, we will adopt the asymptotic approach. In small domain estimation it is customarily assumed that the number m of small areas goes to infinity while the domain sample remains small. Therefore, our theorems will be proved for large values of m .

The problem to be solved in this section is to derive the first-order approximation to the MSE of the proposed estimator as well as an estimator of that approximation. In the infinite population case, the GM-estimator $\hat{\mu}_{Gi}$ of the i -th domain mean as defined in (5.4) is built from a robust estimator $\hat{\boldsymbol{\beta}}_G$ of $\boldsymbol{\beta}$. The approach adopted here is to use the asymptotic properties of $\hat{\boldsymbol{\beta}}_G$ to study those of $\hat{\mu}_{Gi}$.

It is assumed throughout this section that the working model ξ is given by:

$$(\xi) \begin{cases} y_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\beta} + v_i + e_{ij}, & i = 1, \dots, m, j = 1, \dots, n_i, \text{ where,} \\ E_\xi(v_i) = 0, V_\xi(v_i) = \sigma_v^2, \\ E_\xi(e_{ij}) = 0, V_\xi(e_{ij}) = \sigma_e^2, \text{ and ,} \\ E_\xi(v_i e_{ij}) = 0, \end{cases}$$

where $\boldsymbol{\beta}$ is an arbitrary parameter. The random variables v_i 's and e_{ij} 's also satisfy the following two conditions:

C1: The random variables v_i 's and e_{ij} 's are symmetrically distributed (normality is not required).

C2: The true parameter $\boldsymbol{\beta}$ satisfies the condition $\mathbf{g}(\boldsymbol{\beta}) = 0$ where the function $\mathbf{g}$ is defined below.

In the first theorem of this section, we will prove that there exists a GM-estimator $\hat{\boldsymbol{\beta}}_G$ of $\boldsymbol{\beta}$ such that the stochastic order of $\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}$ is $O_p(m^{-1/2})$. Let us introduce the following notations:

$$\Gamma_i(\mathbf{b}) = \alpha_i \bar{\sigma}_i \psi_H[(\bar{y}_i - \bar{\mathbf{x}}_i^\top \mathbf{b})/\bar{\sigma}_i], \quad \xi_{ij}(\mathbf{b}) = \frac{y_{ij} - (\mathbf{x}_{ij}^\top \mathbf{b} + \Gamma_i(\mathbf{b}))}{d_{ij}\sigma_e} \quad (5.8)$$

where $\mathbf{b} \in \mathbb{R}^p$ (arbitrary). Note that $\xi_{ij}(\mathbf{b})$ can be rewritten as

$$\xi_{ij}(\mathbf{b}) = u_{ij}(\mathbf{b}) - \Delta_{ij}(\mathbf{b})$$

where

$$\Delta_{ij}(\mathbf{b}) = \frac{\mathbf{x}_{ij}^\top (\mathbf{b} - \boldsymbol{\beta})}{d_{ij}\sigma_e} \quad (5.9)$$

and

$$u_{ij}(\mathbf{b}) = \frac{(v_i - \Gamma_i(\mathbf{b})) + e_{ij}}{d_{ij}\sigma_e} \quad (5.10)$$

Let us define the following two functions that will be used regularly. For any number $\mathbf{b} \in \mathbb{R}^p$,

$$\begin{aligned} \hat{\mathbf{g}}(\mathbf{b}) &= \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^{n_i} \sigma_e^{-1} d_{ij} \mathbf{x}_{ij} \psi_H \left(\frac{y_{ij} - [\mathbf{x}_{ij}^\top \mathbf{b} + \Gamma_i(\mathbf{b})]}{d_{ij}\sigma_e} \right), \\ \mathbf{g}(\mathbf{b}) &= \lim_{m \rightarrow \infty} E_\xi [\hat{\mathbf{g}}(\mathbf{b})]. \end{aligned} \quad (5.11)$$

Note also that ψ_H is treated here as an arbitrary influence function that must satisfy the regularity conditions given in the theorems.

The theorem establishing the existence and the stochastic order of the GM-estimator is the following:

Theorem 5.1.

Assume that conditions of lemmas 5.3 and 5.4 are satisfied and that

(i) $m = O(n)$ where n is the sample size and m the number of domains.

(ii) $\frac{1}{m} \sum_{i=1}^m n_i^2 = O(1)$.

(iii) $\max_{1 \leq i \leq m} \left\{ \frac{1}{n_i} \sum_{j=1}^{n_i} \|\mathbf{x}_{ij}\| \right\} = O(1)$.

Then there exists a GM-estimator $\hat{\boldsymbol{\beta}}_G$ of $\boldsymbol{\beta}$ satisfying

$$\hat{\boldsymbol{\beta}}_G = \boldsymbol{\beta} + O_p(m^{-1/2}).$$

Remarks.

- Assumption (i) prevents n from going to infinity at much faster than m which is trivial in the context of small area estimation.
- Assumption (ii) can be replaced with the stronger assumption $\sup_{i \geq 1} n_i < \infty$ used by Prasad and Rao (1990). Small area sizes are simply assumed to remain small as the number of domains increases.
- Assumption (iii) can be replaced with a slightly stronger assumption that the elements of the matrix $\mathbf{X}$ are uniformly bounded (see Prasad and Rao (1990)).

Four lemmas are needed in order to prove this theorem. The first lemma is a particular case of result 6.3.4 of Ortega and Rheinbolt (1970, page 163).

Lemma 5.1. Let U be a bounded neighborhood of a p -dimensional real vector $\boldsymbol{\beta}$. Suppose that $\mathbf{f} : \bar{U} \rightarrow \mathbb{R}^p$ is a function defined on the closure $\bar{U}$ of U . If $\mathbf{f}$ is continuous and satisfies the condition $(\mathbf{x} - \boldsymbol{\beta})^\top \mathbf{f}(\mathbf{x}) < 0 \ \forall \mathbf{x} \in \dot{U}$ where $\dot{U}$ is the boundary of U , then the equation $\mathbf{f}(\mathbf{x}) = \mathbf{0}$ has a solution in $\bar{U}$.

Lemma 5.2. (*Markov's Theorem*)

If the sequence of random variables $Y_1, \dots, Y_n, \dots$ is such that

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} V \left(\sum_{k=1}^n Y_k \right) = 0,$$

then for any positive constant ϵ ,

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{1}{n} \sum_{k=1}^n Y_k - \frac{1}{n} \sum_{k=1}^n E(Y_k) \right| < \epsilon \right\} = 1.$$

This lemma is a direct consequence of Chebychev's inequality (see Gnedenko. 1962. page 232).

Now, let the function $\bar{S}(\cdot, \cdot)$ be defined for all $(\mathbf{t}, \mathbf{b}) \in \mathbb{R}^p \times \mathbb{R}^p$ by

$$\bar{S}(\mathbf{t}, \mathbf{b}) = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^{n_i} (\mathbf{x}_{ij}^\top \mathbf{t})^2 \psi'_H[\xi_{ij}(\mathbf{b})].$$

Lemma 5.3. Suppose that the following conditions are satisfied:

(i) Let n_i denote the number of sample units belonging to domain i . Then

$$\max_{1 \leq i \leq m} \left\{ \frac{1}{n_i} \sum_{j=1}^{n_i} \|\mathbf{x}_{ij}\|^4 \right\} = O(1)$$

(ii) As the number of domains increases, the domain sizes remain small.

$$\sum_{i=1}^m n_i^2 = O(m \log m).$$

Then, for any $(\mathbf{t}, \mathbf{b}) \in \mathbb{R}^p \times \mathbb{R}^p$ such that $\|\mathbf{t}\| = 1$,

$$\lim_{m \rightarrow \infty} V_\xi[\bar{S}(\mathbf{t}, \mathbf{b})] = 0.$$

Note that assumptions (i) and (ii) are similar to the assumptions (iii) and (ii) of theorem 5.1 and have the same motivation.

Proof.

It is clear that

$$\begin{aligned} V_{\xi}[\bar{S}(\mathbf{t}, \mathbf{b})] &\leq \frac{1}{n^2} \sum_{i=1}^m E_{\xi} \left(\sum_{j=1}^{n_i} \sigma_e^{-2} (\mathbf{x}_{ij}^{\top} \mathbf{t})^2 \psi'_H[\xi_{ij}(\mathbf{b})] \right)^2, \\ &= \frac{1}{n^2} \sum_{i=1}^m E_{\xi} \left(\sum_{j=1}^{n_i} \sigma_e^{-4} (\mathbf{x}_{ij}^{\top} \mathbf{t})^4 \psi_H'^2[\xi_{ij}(\mathbf{b})] + \sum_{j \neq l}^{n_i} \sum_{l=1}^{n_i} \sigma_e^{-4} (\mathbf{x}_{ij}^{\top} \mathbf{t})^2 (\mathbf{x}_{il}^{\top} \mathbf{t})^2 \right. \\ &\quad \left. \psi'_H[\xi_{ij}(\mathbf{b})] \psi'_H[\xi_{il}(\mathbf{b})] \right) \end{aligned}$$

Note that the function $\psi'_H(\cdot)$ is bounded by 1. Therefore,

$$\begin{aligned} V_{\xi}(\bar{S}(\mathbf{t}, \mathbf{b})) &\leq \frac{1}{n^2} \sum_{i=1}^m \sum_{j=1}^{n_i} \sigma_e^{-4} (\mathbf{x}_{ij}^{\top} \mathbf{t})^4 + \frac{1}{n^2} \sum_{i=1}^m \left(\sum_{j=1}^{n_i} \sigma_e^{-2} (\mathbf{x}_{ij}^{\top} \mathbf{t})^2 \right)^2, \\ &\leq \sigma_e^{-4} \left\{ \frac{1}{n^2} \sum_{i=1}^m \sum_{j=1}^{n_i} \|\mathbf{x}_{ij}\|^4 + \frac{1}{n^2} \sum_{i=1}^m \left(\sum_{j=1}^{n_i} \|\mathbf{x}_{ij}\|^2 \right)^2 \right\}. \end{aligned}$$

The rightmost term on the right-hand side of the above inequality is bounded by

$$\left(\max_{1 \leq i \leq m} \frac{1}{n_i} \sum_{j=1}^{n_i} \|\mathbf{x}_{ij}\|^2 \right)^2 \cdot \frac{1}{n^2} \sum_{i=1}^m n_i^2,$$

and the result follows from assumptions (i) and (ii).

■

For any $\theta \in (0, 1)$ and any p -dimensional real-valued vector $\mathbf{b}$ define $\varphi_{ij}(\theta, \mathbf{b})$ by $\varphi_{ij}(\theta, \mathbf{b}) = u_{ij}(\mathbf{b}) - \theta \Delta_{ij}(\mathbf{b})$. We have the following lemma:

Lemma 5.4. *Assume the following conditions:*

(i)

$$\lim_{m \rightarrow \infty} \inf_{\|\mathbf{t}\|=1} \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^{n_i} (\mathbf{x}_{ij}^\top \mathbf{t})^2 = l_0 > 0$$

(ii) $\max_{i,j} H_{ij,ij} = O(1/m)$.

(iii) *The influence function $\psi_H(\cdot)$ is bounded with a bounded first-order derivative.*

Then, there exists $\delta_0 > 0$ and $\eta > 0$, such that $\forall \theta \in (0, 1)$

$$\lim_{m \rightarrow \infty} P \left\{ \inf_{\|\mathbf{t}\|=1} \left[\frac{1}{n} \sum_{i=1}^m \sum_{j=1}^{n_i} (\mathbf{x}_{ij}^\top \mathbf{t})^2 \psi'_H[\varphi_{ij}(\theta, \mathbf{b})] / \sigma_e^2, \|\mathbf{b} - \boldsymbol{\beta}\| \leq \delta_0 \right] \geq \eta \right\} = 1$$

Remarks:

- Condition (i) is trivial. In fact the only where (i) is not satisfied is when all $\mathbf{x}_{ij}$ are equal to 0.
- Condition (ii) is a standard regularity condition (see (5.6) for the definition of $H_{ij,ij}$). This condition is needed here because it ensures that for all i and j , $E_\xi \psi'_H[\varphi_{ij}(\theta, \mathbf{b})] > 0$ if $\mathbf{b}$ is in the neighborhood of $\boldsymbol{\beta}$. For Some influence functions, $\psi'(\cdot)$ can be 0 if its argument is very large. Under (ii), $\inf_{i,j} d_{ij} > 0$, which will guarantee that $\Delta_{ij}(\mathbf{b})$ (see (5.9)) will be close to 0 in probability.

Proof:

Let $\theta \in (0, 1)$ and $\bar{S}_\theta(\mathbf{t}, \mathbf{b})$ defined for all $(\mathbf{t}, \mathbf{b}) \in \mathbb{R}^p \times \mathbb{R}^p$ by

$$\bar{S}_\theta(\mathbf{t}, \mathbf{b}) = \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^{n_i} (\mathbf{x}_{ij}^\top \mathbf{t})^2 \psi'_H[\varphi_{ij}(\theta, \mathbf{b})] / \sigma_e^2.$$

It is sufficient to prove that there exist two fixed positive real numbers δ_0 and η such that for all $(\mathbf{t}, \mathbf{b}) \in \mathbb{R}^p \times \mathbb{R}^p$ where $\|\mathbf{t}\| = 1$ and $\|\mathbf{b} - \boldsymbol{\beta}\| < \delta_0$, $\lim_{m \rightarrow \infty} P[\bar{S}_\theta(\mathbf{t}, \mathbf{b}) \geq \eta] = 1$. In fact, $\forall \epsilon > 0$, $\exists \mathbf{t}_\epsilon \in \mathbb{R}^p$ with $\|\mathbf{t}_\epsilon\| = 1$ such that

$$\bar{S}_\theta(\mathbf{t}_\epsilon, \mathbf{b}) < \inf_{\|\mathbf{t}\|=1} \bar{S}_\theta(\mathbf{t}, \mathbf{b}) + \epsilon.$$

Therefore,

$$1 = \lim_{m \rightarrow \infty} P[\bar{S}_\theta(\mathbf{t}_\epsilon, \mathbf{b}) > \eta] \leq \lim_{m \rightarrow \infty} P\left[\inf_{\|\mathbf{t}\|=1} \bar{S}_\theta(\mathbf{t}, \mathbf{b}) + \epsilon > \eta\right]$$

It suffices to take $\epsilon = \eta/2$ to prove the theorem.

It follows from assumption (ii) that there exists $\delta_0 > 0$ such that if $\|\mathbf{b} - \boldsymbol{\beta}\| < \delta_0$ then

$$\lim_{m \rightarrow \infty} \inf_{i,j} E_\xi \psi'_H[\varphi_{ij}(\theta, \mathbf{b})] = \mu(\theta, \mathbf{b}) > 0.$$

Let $\mathbf{t} \in \mathbb{R}^p$ be such that $\|\mathbf{t}\| = 1$ and $\mathbf{b}$ satisfying $\|\mathbf{b} - \boldsymbol{\beta}\| \leq \delta_0$. Set

$$\mu_0 = \sigma_e^{-2} \min_{\theta \in (0,1)} \mu(\theta, \mathbf{b}) \lim_{m \rightarrow \infty} \inf_{\|\mathbf{t}\|=1} \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^{n_i} (\mathbf{x}_{ij}^\top \mathbf{t})^2.$$

It follows from assumption (i) that $\mu_0 > 0$. Define $\eta = \mu_0/2$. It follows from Markov's theorem and lemma 5.3 that

$$\lim_{m \rightarrow \infty} P(|\bar{S}_\theta(\mathbf{t}, \mathbf{b}) - E_\xi \bar{S}_\theta(\mathbf{t}, \mathbf{b})| < \eta) = 1.$$

Therefore,

$$\forall \epsilon > 0 \exists N_\epsilon \text{ such that } n > N_\epsilon \implies P[\bar{S}_\theta(\mathbf{t}, \mathbf{b}) > E_\xi \bar{S}_\theta(\mathbf{t}, \mathbf{b}) - \eta] > 1 - \epsilon,$$

$$\implies P[\bar{S}_\theta(\mathbf{t}, \mathbf{b}) > \mu_0 - \eta] > 1 - \epsilon$$

Note that $\eta = \mu_0/2$ and the proof is completed.

■

Proof of theorem 5.1

We have to prove the existence of $\hat{\boldsymbol{\beta}}_G \in \mathbb{R}^p$ satisfying the following condition:

$$\forall \epsilon > 0, \exists M_\epsilon \mid m \geq M_\epsilon \implies P\left(\|\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}\|^2 \leq \frac{M_\epsilon}{m}\right) \geq 1 - \epsilon.$$

Let us define the following set:

$$U(\boldsymbol{\beta}, r) = \{\mathbf{x} \in \mathbb{R}^p : \|\mathbf{x} - \boldsymbol{\beta}\|^2 < r\}.$$

Note that the closure $\bar{U}(\boldsymbol{\beta}, r)$ and the boundary $\dot{U}(\boldsymbol{\beta}, r)$ of this set are respectively given by:

$$\bar{U}(\boldsymbol{\beta}, r) = \{\mathbf{x} \in \mathbb{R}^p : \|\mathbf{x} - \boldsymbol{\beta}\|^2 \leq r\}, \text{ and } \dot{U}(\boldsymbol{\beta}, r) = \{\mathbf{x} \in \mathbb{R}^p : \|\mathbf{x} - \boldsymbol{\beta}\|^2 = r\}.$$

Let $\epsilon_0 > 0$. It follows from lemma 5.1 that it suffices to prove that there exists M_{ϵ_0} for which

$$P\left(\left\{(\mathbf{b} - \boldsymbol{\beta})^\top \hat{\mathbf{g}}(\mathbf{b}) < 0, \forall \mathbf{b} \in \dot{U}(\boldsymbol{\beta}, M_{\epsilon_0}/m)\right\}\right) \geq 1 - \epsilon_0.$$

We have for all $\mathbf{b} \in \mathbb{R}^p$ that

$$\begin{aligned} (\mathbf{b} - \boldsymbol{\beta})^\top \hat{\mathbf{g}}(\mathbf{b}) &= \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^{n_i} \sigma_e^{-1} d_{ij} \mathbf{x}_{ij}^\top (\mathbf{b} - \boldsymbol{\beta}) \psi_H[\xi_{ij}(\mathbf{b})], \\ &= \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^{n_i} \sigma_e^{-1} d_{ij} \mathbf{x}_{ij}^\top (\mathbf{b} - \boldsymbol{\beta}) \psi_H[u_{ij}(\mathbf{b}) - \Delta_{ij}(\mathbf{b})]. \end{aligned}$$

It follows from the mean value theorem that $\forall i, j$ we have

$$\psi_H[u_{ij}(\mathbf{b}) - \Delta_{ij}(\mathbf{b})] = \psi_H[u_{ij}(\mathbf{b})] - \Delta_{ij}(\mathbf{b}) \psi'_H[\theta_{ij}(\mathbf{b})],$$

where $\theta_{ij}(\mathbf{b}) = u_{ij} - \theta \Delta_{ij}(\mathbf{b})$ and $\theta \in (0, 1)$. Therefore,

$$\begin{aligned} (\mathbf{b} - \boldsymbol{\beta})^\top \hat{\mathbf{g}}(\mathbf{b}) &= \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^{n_i} \sigma_e^{-1} d_{ij} \mathbf{x}_{ij}^\top (\mathbf{b} - \boldsymbol{\beta}) \psi_H[u_{ij}(\mathbf{b})] - \\ &\quad - \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^{n_i} \sigma_e^{-2} [\mathbf{x}_{ij}^\top (\mathbf{b} - \boldsymbol{\beta})]^2 \psi'_H[\theta_{ij}(\mathbf{b})] = G_1(\mathbf{b}) - G_2(\mathbf{b}). \end{aligned}$$

It follows from Markov's inequality that

$$P\left(G_1(\mathbf{b}) \leq \sqrt{2\epsilon_0^{-1} E_\xi |G_1(\mathbf{b})|^2}\right) \geq 1 - \epsilon_0/2.$$

But,

$$\begin{aligned} |G_1(\mathbf{b})|^2 &\leq \|\mathbf{b} - \boldsymbol{\beta}\|^2 \left\| \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^{n_i} \sigma_e^{-1} d_{ij} \mathbf{x}_{ij} \psi_H[u_{ij}(\mathbf{b})] \right\|^2, \\ &= \|\mathbf{b} - \boldsymbol{\beta}\|^2 \frac{1}{n^2} \sum_{i=1}^m \sum_{k=1}^m \sum_{j=1}^{n_i} \sum_{l=1}^{n_k} \sigma_e^{-2} d_{ij} d_{kl} \psi_H[u_{ij}(\mathbf{b})] \psi_H[u_{kl}(\mathbf{b})] \mathbf{x}_{ij}^\top \mathbf{x}_{kl}. \end{aligned}$$

Note that if $i \neq k$ then for all $j = 1, \dots, n_i$ and $l = 1, \dots, n_k$

$$\begin{aligned} E_\xi \psi_H[u_{ij}(\mathbf{b})] \psi_H[u_{kl}(\mathbf{b})] &= E_\xi \psi_H[u_{ij}(\mathbf{b})] \cdot E_\xi \psi_H[u_{kl}(\mathbf{b})], \\ &= \alpha_i \alpha_k \bar{\mathbf{x}}_i^\top (\mathbf{b} - \boldsymbol{\beta}) \bar{\mathbf{x}}_k^\top (\mathbf{b} - \boldsymbol{\beta}) / (d_{ij} d_{kl} \sigma_e^2). \end{aligned}$$

This result follows from the assumption of symmetry made on the variables v_i 's and

e_{ij} 's. In fact, $u_{ij}(\mathbf{b})$ is given by (see (5.10))

$$u_{ij}(\mathbf{b}) = \frac{(v_i - \Gamma_i(\mathbf{b})) + e_{ij}}{d_{ij} \sigma_e}$$

and

$$\begin{aligned} \Gamma_i(\mathbf{b}) &= \alpha_i \bar{\sigma}_i \psi_H \left\{ (\bar{y}_i - \bar{\mathbf{x}}_i^\top \mathbf{b}) / \bar{\sigma}_i \right\} \\ &= \alpha_i \bar{\sigma}_i \psi_H \left\{ \frac{(\bar{y}_i - \bar{\mathbf{x}}_i^\top \boldsymbol{\beta}) - \bar{\mathbf{x}}_i^\top (\mathbf{b} - \boldsymbol{\beta})}{\bar{\sigma}_i} \right\}, \\ &= \alpha_i \bar{\sigma}_i \psi_H \left\{ \frac{(v_i + \bar{e}_i) - \bar{\mathbf{x}}_i^\top (\mathbf{b} - \boldsymbol{\beta})}{\bar{\sigma}_i} \right\}, \end{aligned}$$

where $\bar{e}_i = (1/n_i) \sum_{j=1}^{n_i} e_{ij}$. Since ψ_H is an odd function and the $v_i + \bar{e}_i$'s are symmetrically distributed, $\Gamma_i(\mathbf{b})$ is a random variable that is symmetrically distributed around $-\alpha_i \bar{\mathbf{x}}_i^\top (\mathbf{b} - \boldsymbol{\beta})$. Therefore, the $u_{ij}(\mathbf{b})$'s are symmetrically distributed around $\alpha_i \bar{\mathbf{x}}_i^\top (\mathbf{b} - \boldsymbol{\beta}) / (d_{ij} \sigma_e)$.

Now, we have that:

$$\begin{aligned} E_\xi |G_1(\mathbf{b})|^2 &\leq \|\mathbf{b} - \boldsymbol{\beta}\|^2 \left\{ \frac{1}{n^2} \sum_{i=1}^m \sum_{j=1}^{n_i} \sum_{l=1}^{n_i} \sigma_e^{-2} d_{ij} d_{il} \mathbf{x}_{ij}^\top \mathbf{x}_{il} E_\xi \psi_H[u_{ij}(\mathbf{b})] \psi_H[u_{il}^*(\mathbf{b})] + \right. \\ &\quad \left. + \frac{1}{n^2} \sum_{i=1}^m \sum_{k \neq i}^m \sum_{j=1}^{n_i} \sum_{l=1}^{n_k} \sigma_e^{-4} \alpha_i \alpha_k \bar{\mathbf{x}}_i^\top (\mathbf{b} - \boldsymbol{\beta}) \bar{\mathbf{x}}_k^\top (\mathbf{b} - \boldsymbol{\beta}) \mathbf{x}_{ij}^\top \mathbf{x}_{kl} \right\}. \end{aligned}$$

Since the ψ_H function is bounded and the d_{ij} 's are all smaller than 1, there exists a positive constant B such that

$$\begin{aligned} E_\xi |G_1(\mathbf{b})|^2 &\leq B \|\mathbf{b} - \boldsymbol{\beta}\|^2 \left(\frac{1}{n^2} \sum_{i=1}^m \sum_{j=1}^{n_i} \sum_{l=1}^{n_i} \mathbf{x}_{ij}^\top \mathbf{x}_{il} \right) + \\ &\quad + \sigma_e^{-2} l_0^2 \|\mathbf{b} - \boldsymbol{\beta}\|^4 \left(\frac{1}{n^2} \sum_{i=1}^m \sum_{k \neq i}^m \sum_{j=1}^{n_i} \sum_{l=1}^{n_k} \mathbf{x}_{ij}^\top \mathbf{x}_{kl} \right), \end{aligned}$$

where,

$$l_0 = \lim_{m \rightarrow \infty} \max_{1 \leq i \leq m} \left\{ \frac{1}{n_i} \sum_{j=1}^{n_i} \|\mathbf{x}_{ij}\| \right\}.$$

Note that

$$\begin{aligned} \sum_{i=1}^m \sum_{j=1}^{n_i} \sum_{l=1}^{n_i} \mathbf{x}_{ij}^\top \mathbf{x}_{il} &= \sum_{i=1}^m n_i^2 \bar{\mathbf{x}}_i^2, \\ &\leq \left\{ \max_{1 \leq i \leq m} \frac{1}{n_i} \sum_{j=1}^{n_i} \|\mathbf{x}_{ij}\| \right\}^2 \sum_{i=1}^m n_i^2. \end{aligned}$$

This quantity is $O(m)$ because of (ii) and (iii). Moreover,

$$\frac{1}{n^2} \sum_{i=1}^m \sum_{k \neq i}^m \sum_{j=1}^{n_i} \sum_{l=1}^{n_k} \mathbf{x}_{ij}^\top \mathbf{x}_{kl} \leq \left(\frac{1}{n} \sum_{i=1}^m \sum_{j=1}^{n_i} \mathbf{x}_{ij}^\top \right) \left(\frac{1}{n} \sum_{k=1}^m \sum_{l=1}^{n_k} \mathbf{x}_{kl} \right).$$

It follows from (ii) and (iii) that the above quantity is $O(1)$. Therefore, there exist two positive constants $B_{1\epsilon_0}$ and $B_{2\epsilon_0}$ such that if m is large enough,

$$2\epsilon_0^{-1} E_\xi |G_1(\mathbf{b})|^2 \leq \|\mathbf{b} - \boldsymbol{\beta}\|^2 \frac{B_{1\epsilon_0}^2}{m} + B_{2\epsilon_0}^2 \|\mathbf{b} - \boldsymbol{\beta}\|^4.$$

Thus, if $\mathbf{b}$ is such that $\|\mathbf{b} - \boldsymbol{\beta}\| < \delta_0$ for an arbitrary positive value δ_0 ,

$$P\left(G_1(\mathbf{b}) \leq \|\mathbf{b} - \boldsymbol{\beta}\| \frac{B_{1\epsilon_0}}{m^{1/2}} + B_{2\epsilon_0} \|\mathbf{b} - \boldsymbol{\beta}\|^2\right) \geq 1 - \epsilon_0/2. \quad (5.12)$$

Let us now solve the problem of $G_2(\mathbf{b})$. It follows from the definition of G_2 that

$$G_2(\mathbf{b}) \geq \|\mathbf{b} - \boldsymbol{\beta}\|^2 \inf_{\|\mathbf{t}\|=1} \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^{n_i} \sigma_e^{-2} (\mathbf{x}_{ij}^\top \mathbf{t})^2 \psi'_H[\varphi_{ij}(\theta, \mathbf{b})].$$

It follows from lemma 5.4 that

$$\lim_{m \rightarrow \infty} P\left(G_2(\mathbf{b}) \geq \eta \|\mathbf{b} - \boldsymbol{\beta}\|^2, \|\mathbf{b} - \boldsymbol{\beta}\| \leq \delta_1\right) = 1, \quad (5.13)$$

for some constant values $\eta > 0$ and $\delta_1 > 0$.

Let us define the following events:

$$E_1 = \left\{ G_1(\mathbf{b}) \leq \|\mathbf{b} - \boldsymbol{\beta}\| \frac{B_{1\epsilon_0}}{m^{1/2}} + B_{2\epsilon_0} \|\mathbf{b} - \boldsymbol{\beta}\|^2, \|\mathbf{b} - \boldsymbol{\beta}\| < \delta \right\},$$

$$E_2 = \left\{ G_2(\mathbf{b}) \geq \eta \|\mathbf{b} - \boldsymbol{\beta}\|^2, \|\mathbf{b} - \boldsymbol{\beta}\| < \delta_1 \right\},$$

$$F = \left\{ G_1(\mathbf{b}) - G_2(\mathbf{b}) \leq \|\mathbf{b} - \boldsymbol{\beta}\| \frac{B_{1\epsilon_0}}{m^{1/2}} + (B_{2\epsilon_0} - \eta) \|\mathbf{b} - \boldsymbol{\beta}\|^2, \|\mathbf{b} - \boldsymbol{\beta}\| \leq \delta_{01} \right\}.$$

where $\delta_{01} = \min(\delta_0, \delta_1)$. It is clear that

$$\begin{aligned} P(F) &\geq P(E_1 E_2) = P(E_1) - P(E_1 E_2^c), \\ &\geq P(E_1) - P(E_2^c). \end{aligned}$$

It follows from (5.13) that $P(E_2^c) \rightarrow 0$ as $m \rightarrow \infty$. Thus, $\exists M_0$ such that if $m \geq M_0$ then $P(E_2^c) \leq \epsilon_0/2$ and $P(F) \geq 1 - \epsilon_0$. It follows again from (5.13) that there is no problem assuming that $\eta > B_{2\epsilon_0}$. Now let

$$K_{\epsilon_0} = \frac{4B_{1\epsilon_0}^2}{(\eta - B_{2\epsilon_0})^2}.$$

Clearly there exists $M_0^* > 0$ such that if $m \geq M_0^*$ then $K_{\epsilon_0}/m \leq \min(\delta_0, \delta_1)$. Thus.

$$m \geq M_0^* \implies P\left(G_1(\mathbf{b}) - G_2(\mathbf{b}) \leq -2\frac{B_{1\epsilon_0}}{\eta - B_{2\epsilon_0}}, \|\mathbf{b} - \boldsymbol{\beta}\| = \frac{K_{\epsilon_0}}{m^{1/2}}\right) \geq 1 - \epsilon_0.$$

Since $-2\frac{B_{1\epsilon_0}}{\eta - B_{2\epsilon_0}} < 0$, the event F implies that $(\mathbf{b} - \boldsymbol{\beta})^\top \mathbf{g}(\mathbf{b}) < 0$ whenever $\|\mathbf{b} - \boldsymbol{\beta}\| = K_{\epsilon_0}/m^{1/2}$. The ball in $\mathbb{R}^p$ originally sought is given by $U(\boldsymbol{\beta}, K_{\epsilon_0}^2/m)$.

■

Remark:

We have just proved the existence of $\hat{\boldsymbol{\beta}}_G$ such that $O_p(m^{-1/2})$ is the stochastic order of magnitude of the difference $\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}$. The next result states that this solution is unique.

Corollary 5.1. *Assume the conditions of theorem 5.1 and that $\mathbf{X}$ is a full column rank matrix. Then the solution $\hat{\boldsymbol{\beta}}_G$ is unique.*

Lemma 5.5. *For any $i = 1, \dots, m$ let $\mathbf{D}_i = \text{diag}(\psi_{1ij}, j = 1, \dots, n_i)$ where $\psi_{1ij} = \psi'_H[\xi_{ij}(\boldsymbol{\beta})]$. The matrix $\mathbf{1}_n, \mathbf{1}_n^\top \mathbf{D}_i^{-1}$ is positive definite.*

Proof:

It suffices to show that all the eigenvalues of this matrix are strictly positive. Let λ_i be an eigenvalue of $\mathbf{1}_{n_i} \mathbf{1}_{n_i}^\top D_i^{-1}$. Then there exists an associated nonzero eigenvector $\mathbf{u}_i$ such that $\mathbf{1}_{n_i} \mathbf{1}_{n_i}^\top D_i^{-1} \mathbf{u}_i = \lambda_i \mathbf{u}_i$. The left-hand side of the above equality is a vector multiplied by a scalar. Thus, for $j = 1, \dots, n_i$, $\lambda_i u_{ij} = \mathbf{1}_{n_i}' D_i^{-1} \mathbf{u}_i$. Therefore, there exists a nonzero number a_i such that $u_{ij} = a_i$. As a result,

$$\begin{aligned} \lambda_i &= \mathbf{1}_{n_i}' D_i^{-1} \mathbf{1}_{n_i} \\ &= \sum_{j=1}^{n_i} \frac{1}{\psi'_H[\xi_{ij}(\boldsymbol{\beta})]} > 0 \end{aligned}$$

■

Proof of corollary 5.1 :

It suffices to prove that the matrix $\partial \hat{\mathbf{g}}(\mathbf{b}) / \partial \mathbf{b}$ is negative definite for any $\mathbf{b} \in \mathbb{R}^p$.

Recall that for any $\mathbf{b} \in \mathbb{R}^p$,

$$\hat{\mathbf{g}}(\mathbf{b}) = \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^{n_i} \sigma_e^{-1} d_{ij} \mathbf{x}_{ij} \psi_H \left(\frac{y_{ij} - [\mathbf{x}_{ij}^\top \mathbf{b} + \Gamma_i(\mathbf{b})]}{d_{ij} \sigma_e} \right).$$

The derivative of this function is given by:

$$\partial \hat{\mathbf{g}}(\mathbf{b}) / \partial \mathbf{b} \Big|_{\mathbf{b}=\boldsymbol{\beta}} = -\frac{1}{n} \sum_{i=1}^m \sum_{j=1}^{n_i} \sigma_e^{-2} \mathbf{x}_{ij} \psi'_H[\xi_{ij}(\boldsymbol{\beta})] [\mathbf{x}_{ij}^\top + \Gamma'_i(\boldsymbol{\beta})].$$

Because $d\Gamma_i(\mathbf{b})/d\mathbf{b} = -\alpha_i \psi'_H[t_i(\mathbf{b})] \bar{\mathbf{x}}_i^\top$, we have that

$$\partial \hat{\mathbf{g}}(\mathbf{b}) / \partial \mathbf{b} \Big|_{\mathbf{b}=\boldsymbol{\beta}} = -(n\sigma_e^2)^{-1} \sum_{i=1}^m \left[\sum_{j=1}^{n_i} \mathbf{x}_{ij} \psi_{1ij} \mathbf{x}_{ij}^\top - n_i \alpha_{Gi} \psi_{1i} \bar{\mathbf{x}}_{1i} \bar{\mathbf{x}}_i^\top \right],$$

where $\alpha_{Gi} = \alpha_i \psi'_H[t_i(\boldsymbol{\beta})]$ and $\psi_{1ij} = \psi'_H(\xi_{ij}(\boldsymbol{\beta}))$. Note that

$$\partial \hat{\mathbf{g}}(\mathbf{b}) / \partial \mathbf{b} \Big|_{\mathbf{b}=\boldsymbol{\beta}} = -(n\sigma_e^2)^{-1} \mathbf{X} \mathbf{V}_G^{-1} \mathbf{X}$$

where

$$\begin{cases} \mathbf{V}_G = \text{block diag}(\mathbf{V}_{Gi}, i = 1, \dots, m), \\ \mathbf{V}_{Gi} = \sigma_e^2 \mathbf{D}_i^{-1} + \sigma_v^2 \mathbf{1}_{n_i} \mathbf{1}_{n_i}^\top \mathbf{D}_i^{-1} \end{cases}$$

We have to show that the matrix $\partial \widehat{\mathbf{g}}(\mathbf{b}) / \partial \mathbf{b} \big|_{\mathbf{b}=\boldsymbol{\beta}}$ is positive-definite. Since $\mathbf{X}$ is full column rank, it is enough to prove that $\mathbf{V}_G$ is positive-definite as the inverse of a positive definite matrix is also positive-definite.

Let $\mathbf{y}$ be an arbitrary nonzero n -dimensional vector. Suppose that $\mathbf{y}^\top = (\mathbf{y}_1, \dots, \mathbf{y}_m)$ where $\mathbf{y}_i^\top = (y_{i1}, \dots, y_{in_i})$. We have to show that $\mathbf{y}^\top \mathbf{V}_G \mathbf{y} > 0$.

$$\begin{aligned} \mathbf{y}^\top \mathbf{V}_G \mathbf{y} &= \sum_{i=1}^m \mathbf{y}_i^\top \mathbf{V}_{Gi} \mathbf{y}_i, \\ &= \sum_{i=1}^m \sum_{j=1}^{n_i} \sigma_e^2 \psi_{1ij}^{-1} y_{ij}^2 + \sigma_v^2 \sum_{i=1}^m \sum_{j=1}^{n_i} \mathbf{y}_i^\top \mathbf{1}_{n_i} \mathbf{1}_{n_i}^\top \mathbf{D}_i^{-1} \mathbf{y}_i. \end{aligned}$$

It follows from lemma 5.5 that all summands in the last equation are positive. Since $\mathbf{y} \neq \mathbf{0}$, there exists i_0 and $j_0 \in \{1, \dots, n_{i_0}\}$ such that $y_{i_0 j_0} \neq 0$. We have that

$$\begin{aligned} \mathbf{y}^\top \mathbf{V}_G \mathbf{y} &\geq \sigma_e^2 \psi_{1i_0 j_0}^{-1} y_{i_0 j_0}^2, \\ &> 0 \end{aligned}$$

■

Clearly in our theoretical investigation, we will need a linear representation of $\widehat{\boldsymbol{\beta}}_G$ and $\widehat{v}_{Gi} = \Gamma_i(\widehat{\boldsymbol{\beta}})$. The following notations are also needed:

$$\begin{aligned} e_{Gij} &= \sigma_e d_{ij} \xi_{ij}(\widehat{\boldsymbol{\beta}}_G), \quad e_{Gij}^* = \sigma_e d_{ij} \xi_{ij}(\boldsymbol{\beta}), \quad \bar{e}_{Gij} = \xi_{ij}(\widehat{\boldsymbol{\beta}}_G), \quad \bar{e}_{Gij}^* = \xi_{ij}(\boldsymbol{\beta}), \\ \psi_{1ij} &= \psi'_H(\bar{e}_{Gij}^*), \quad \psi_{1i\cdot} = \sum_{j=1}^{n_i} \psi_{1ij} \quad \text{and} \quad \bar{\psi}_{Gij} = \psi_H(\bar{e}_{Gij}^*). \end{aligned} \tag{5.14}$$

ψ_{1ij} will sometimes be used as weights to calculate certain means. For example $\bar{\mathbf{x}}_i^*$ is a mean using the weights ψ_{1ij} . It is given by:

$$\bar{\mathbf{x}}_i^* = \frac{1}{n_i} \sum_{j=1}^{n_i} \psi_{1ij} \mathbf{x}_{ij}.$$

Now we can state the next theorem:

Theorem 5.2. *Let us assume that the following conditions are satisfied:*

- (i) $\hat{\boldsymbol{\beta}}_G = \boldsymbol{\beta} + O_p(m^{-1/2})$
- (ii) *The design vectors $\mathbf{x}_{ij}$'s are such that*

$$\max_{1 \leq i \leq m} \left\{ \frac{1}{n_i} \sum_{j=1}^{n_i} \|\mathbf{x}_{ij}\|^3 \right\} = O(1).$$

Then the following linear representations hold:

$$\begin{aligned} \hat{\boldsymbol{\beta}}_G &= \boldsymbol{\beta} + \mathbf{Q}_m^{-1} \hat{\mathbf{g}}(\boldsymbol{\beta}) + o_p(m^{-1/2}), \\ \hat{v}_{Gi} &= v_{Gi} - \alpha_{Gi} \bar{\mathbf{x}}_i^\top (\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}) + o_p(m^{-1/2}). \end{aligned}$$

where

$$\begin{aligned} \sigma_e^2 \mathbf{Q}_m &= \frac{1}{n} \sum_{i=1}^m \left[\sum_{j=1}^{n_i} \psi'_H(\bar{e}_{Gi}^*) \mathbf{x}_{ij} \mathbf{x}_{ij}^\top - \alpha_{Gi} n_i \bar{\mathbf{x}}_i^* \bar{\mathbf{x}}_i'^* \right], \\ \alpha_{Gi} &= \alpha_i \psi'_H[(\bar{y}_i - \bar{\mathbf{x}}_i^\top \boldsymbol{\beta}) / \bar{\sigma}_i] \text{ and } v_{Gi} = \alpha_i \bar{\sigma}_i \psi_H[(\bar{y}_i - \bar{\mathbf{x}}_i^\top \boldsymbol{\beta}) / \bar{\sigma}_i]. \end{aligned}$$

Proof.

Let

$$t_i(\mathbf{b}) = \frac{\bar{y}_i - \bar{\mathbf{x}}_i^\top \mathbf{b}}{\bar{\sigma}_i}, \quad \mathbf{b} \in \mathbb{R}^p, \quad \Gamma_{1i}(\mathbf{b}) = \left. \frac{d\Gamma_i(\mathbf{b})}{d\mathbf{b}} \right|_{\mathbf{b}}, \text{ and } \Gamma_{2i}(\mathbf{b}) = \left. \frac{d^2\Gamma_i(\mathbf{b})}{d\mathbf{b}^2} \right|_{\mathbf{b}}.$$

where $\Gamma_i(\mathbf{b}) = \alpha_i \bar{\sigma}_i \psi_H[t_i(\mathbf{b})]$. The expansion of $\hat{v}_{Gi}$ will be proved first as it is needed to prove that of $\hat{\boldsymbol{\beta}}_G$.

Since the ψ functions we use are twice differentiable everywhere, we can apply Taylor's formula to obtain the following expansion:

$$\Gamma_i(\hat{\boldsymbol{\beta}}_G) = \Gamma_i(\boldsymbol{\beta}) + \Gamma_{1i}(\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}) + \frac{1}{2}(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})^\top \Gamma_{2i}(\hat{\boldsymbol{\beta}}_\theta)(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}).$$

where $\boldsymbol{\beta}_\theta = \theta \hat{\boldsymbol{\beta}}_G + (1 - \theta)\boldsymbol{\beta}$ for some $\theta \in (0, 1)$. It follows from the definition of the function Γ_i that,

$$\Gamma_{1i}(\boldsymbol{\beta}) = -\alpha_i \psi'_H[t_i(\boldsymbol{\beta})] \bar{\mathbf{x}}_i^\top \text{ and } \Gamma_{2i}(\boldsymbol{\beta}) = \sigma_v^{-1} \alpha_i^{3/2} \psi''_H[t_i(\boldsymbol{\beta})].$$

Thus,

$$\hat{v}_{Gi} = v_{Gi} - \alpha_i \psi'_H[t_i(\boldsymbol{\beta})] \bar{\mathbf{x}}_i^\top (\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}) + (2\sigma_v)^{-1} \alpha_i^{3/2} \psi''_H[t_i(\boldsymbol{\beta}_\theta)] (\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})^\top \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^\top (\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}) \quad (5.15).$$

The result for $\hat{v}_{Gi}$ now follows from the hypotheses (i) and (ii) and the use of the Schwartz inequality.

To prove the linear representation of $\hat{\boldsymbol{\beta}}$ we first obtain a Taylor's expansion of function $\hat{\mathbf{g}}$ in the neighborhood of $\boldsymbol{\beta}$. Let

$$\hat{\mathbf{g}}_1(\mathbf{b}) = \left. \frac{\partial \hat{\mathbf{g}}(\mathbf{b})}{\partial \mathbf{b}} \right|_{\mathbf{b}} \text{ and } \hat{\mathbf{g}}_2(\mathbf{b}) = \left. \frac{\partial^2 \hat{\mathbf{g}}(\mathbf{b})}{\partial \mathbf{b}^2} \right|_{\mathbf{b}}.$$

We have that

$$\hat{\mathbf{g}}(\hat{\boldsymbol{\beta}}_G) = \hat{\mathbf{g}}(\boldsymbol{\beta}) + \hat{\mathbf{g}}_1(\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}) + (1/2)(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})^\top \hat{\mathbf{g}}_2(\boldsymbol{\beta}_\theta)(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}).$$

The function $\widehat{\mathbf{g}}_1(\cdot)$ is given by:

$$\begin{aligned}\widehat{\mathbf{g}}_1(\boldsymbol{\beta}) &= -\frac{1}{n} \sum_{i=1}^m \sum_{j=1}^{n_i} \sigma_e^{-2} \mathbf{x}_{ij} \psi'_H[\xi_{ij}(\boldsymbol{\beta})] \mathbf{x}_{ij}^\top - \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^{n_i} \sigma_e^{-2} \mathbf{x}_{ij} \psi'_H[\xi_{ij}(\boldsymbol{\beta})] \Gamma_{1i}(\boldsymbol{\beta}). \\ &= -\widehat{\mathbf{g}}_{1,1}(\boldsymbol{\beta}) - \widehat{\mathbf{g}}_{1,2}(\boldsymbol{\beta}).\end{aligned}$$

Using the formula of $\Gamma_{1i}(\boldsymbol{\beta})$ developed previously, it can be seen that $\widehat{\mathbf{g}}_1(\boldsymbol{\beta}) = -Q_m$

where Q_m is defined in the statement of the theorem. Therefore,

$$0 = \widehat{\mathbf{g}}(\boldsymbol{\beta}) - Q_m(\widehat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}) + \mathbf{r}_m(\theta),$$

where

$$\mathbf{r}_m(\theta) = (1/2)(\widehat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})^\top \widehat{\mathbf{g}}_2(\boldsymbol{\beta}_\theta)(\widehat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})$$

After some algebra we obtained that $\mathbf{r}_m(\theta) = \mathbf{r}_{m1}(\theta) + \mathbf{r}_{m2}(\theta) + \mathbf{r}_{m3}(\theta) + \mathbf{r}_{m4}(\theta)$

where $\mathbf{r}_{mi}(\theta) = (1/2)(\widehat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})^\top \widehat{\mathbf{g}}_{2,i}(\theta)(\widehat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})$ $i = 1, 2, 3, 4$ and

$$\begin{aligned}\widehat{\mathbf{g}}_{2,1}(\theta) &= \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^{n_i} (d_{ij} \sigma_e^3)^{-1} \psi''_H[\xi_{ij}(\boldsymbol{\beta}_\theta)] \mathbf{x}_{ij} \mathbf{x}_{ij}^\top \mathbf{x}_{ij}, \\ \widehat{\mathbf{g}}_{2,2}(\theta) &= -\frac{2}{n} \sum_{i=1}^m \sum_{j=1}^{n_i} (d_{ij} \sigma_e^3)^{-1} \alpha_i \psi'_H[t_i(\boldsymbol{\beta}_\theta)] \psi''_H[\xi_{ij}(\boldsymbol{\beta}_\theta)] \mathbf{x}_{ij} \mathbf{x}_{ij}^\top \bar{\mathbf{x}}_i, \\ \widehat{\mathbf{g}}_{2,3}(\theta) &= \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^{n_i} (d_{ij} \sigma_e^3)^{-1} \alpha_i^2 \psi_H'^2[t_i(\boldsymbol{\beta}_\theta)] \psi''_H[\xi_{ij}(\boldsymbol{\beta}_\theta)] \mathbf{x}_{ij} \bar{\mathbf{x}}_i^\top \bar{\mathbf{x}}_i, \\ \widehat{\mathbf{g}}_{2,4}(\theta) &= -\frac{1}{n} \sum_{i=1}^m \sum_{j=1}^{n_i} (\sigma_e^2 \sigma_v)^{-1} \alpha_i^{3/2} \psi_H''[t_i(\boldsymbol{\beta}_\theta)] \psi'_H[\xi_{ij}(\boldsymbol{\beta}_\theta)] \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^\top \mathbf{x}_{ij}.\end{aligned}$$

The next step is to use the relation just proved and then the result for $\widehat{\boldsymbol{\beta}}_G$ follows from the boundedness of the ψ function, the conditions (i), (ii), (iii) and the Chebychev inequality.

■

The theorem 5.2 just proved will now be used to derive the first-order approximation of the MSE of the estimator $\hat{\mu}_{Gi}$ of the i th domain mean μ_i in the infinite population case. Before stating the next theorem, it is important to note that the matrix $\mathbf{Q}_m$ can be rewritten as

$$\mathbf{Q}_m = (1/n) (\mathbf{X}^\top \mathbf{V}_G^{-1} \mathbf{X}),$$

where

$$\mathbf{V}_G = \text{diag}(\mathbf{V}_{Gi}, i = 1, \dots, m), \quad \mathbf{V}_{Gi} = \sigma_e^2 \mathbf{D}_{1,i}^{-1} + \sigma_v^2 \mathbf{1}_{n_i} \mathbf{1}_{n_i}' \mathbf{D}_{1,i}^{-1}.$$

$$\mathbf{D}_{1,i} = \text{diag}(\psi_{1ij}, j = 1, \dots, n_i).$$

It follows from this that

$$\mathbf{V}_G^{-1} = \text{diag}(\mathbf{V}_{Gi}^{-1}, i = 1, \dots, m) \text{ and } \mathbf{V}_{Gi}^{-1} = \sigma_e^{-2} \left(\mathbf{D}_{1,i} - \frac{\alpha_i}{n_i} \mathbf{D}_{1,i} \mathbf{1}_{n_i} \mathbf{1}_{n_i}' \right)$$

Let $\hat{\mathbf{g}}_k(\boldsymbol{\beta})$ ($k = 1, \dots, m$) be defined by

$$\hat{\mathbf{g}}_k(\boldsymbol{\beta}) = \frac{1}{n_k} \sum_{l=1}^{n_k} d_{kl} \mathbf{x}_{kl} \psi_H(\bar{e}_{Gkl}^*) / \sigma_e$$

and the matrix $\mathbf{U}_m$ by

$$\mathbf{U}_m = \frac{1}{n} \sum_{k=1}^m n_k^2 E_\xi (\hat{\mathbf{g}}_k(\boldsymbol{\beta}) [\hat{\mathbf{g}}_k(\boldsymbol{\beta})]^\top).$$

Note that,

$$\hat{\mathbf{g}}(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^m n_i \hat{\mathbf{g}}_i(\boldsymbol{\beta})$$

where

$$\begin{aligned} \hat{\mathbf{g}}_i(\boldsymbol{\beta}) &= \frac{1}{n_i} \sum_{j=1}^{n_i} (d_{ij} / \sigma_e) \mathbf{x}_{ij} \psi_H \left(\frac{y_{ij} - (\mathbf{x}_{ij}^\top \boldsymbol{\beta} + \Gamma_i(\boldsymbol{\beta}))}{d_{ij} \sigma_e} \right), \\ &= \frac{1}{n_i} \sum_{j=1}^{n_i} (d_{ij} / \sigma_e) \mathbf{x}_{ij} \psi_H \left(\frac{v_i + e_{ij} - \Gamma_i(\boldsymbol{\beta})}{d_{ij} \sigma_e} \right). \end{aligned}$$

and $\Gamma_i(\boldsymbol{\beta}) = \alpha_i \bar{\sigma}_i \psi_H \left\{ (\bar{e}_{i\bullet} + v_i) / \bar{\sigma}_i \right\}$. Let η_{ij} and λ_{ij} be respectively given by

$$\eta_{ij} = (1/\sigma_e) E_\xi \left\{ v_i d_{ij} \psi_H \left(\frac{v_i + e_{ij} - \Gamma_i(\boldsymbol{\beta})}{d_{ij} \sigma_e} \right) \right\}$$

and

$$\lambda_{ij} = (1/\sigma_e) E_\xi \left\{ \bar{e}_{i\bullet} d_{ij} \psi_H \left(\frac{v_i + e_{ij} - \Gamma_i(\boldsymbol{\beta})}{d_{ij} \sigma_e} \right) \right\}$$

Lemma 5.6. Assume that

- (i) The v_i 's ($i = 1, \dots, m$) are independent, identically and symmetrically distributed with mean 0 and variance σ_v^2 .
- (ii) The e_{ij} 's are independent, identically and symmetrically distributed with mean 0 and variance σ_e^2 for $j = 1, \dots, n_i$ and $i = 1, \dots, m$.
- (iii) The v_i 's and the e_{ij} 's are mutually independent.

Then $\eta_{ij} = \alpha_i/n_i$ and $\lambda_{ij} = (1 - \alpha_i)/n_i$.

Proof of lemma 5.6

We have that

$$\eta_{ij} = (1/\sigma_e) E_\xi \left\{ v_i E_\xi \left[d_{ij} \psi_H \left(\frac{v_i + e_{ij} - \Gamma_i(\boldsymbol{\beta})}{d_{ij} \sigma_e} \right) \middle| v_i \right] \right\}$$

When v_i is fixed, the distribution of $\Gamma_i(\boldsymbol{\beta})$ is symmetric with mean $\alpha_i v_i$. Moreover, the distribution of $(v_i + e_{ij} - \Gamma_i(\boldsymbol{\beta})) / (d_{ij} \sigma_e)$ is symmetric with mean $(1 - \alpha_i) v_i / (d_{ij} \sigma_e)$. Therefore, $\eta_{ij} = (1/\sigma_e^2) (1 - \alpha_i) E_\xi(v_i^2) = \alpha_i/n_i$

Similarly we have

$$\lambda_{ij} = (1/\sigma_e) E_\xi \left\{ \bar{e}_{i\bullet} E_\xi \left[d_{ij} \psi_H \left(\frac{v_i + e_{ij} - \Gamma_i(\boldsymbol{\beta})}{d_{ij} \sigma_e} \right) \middle| e_{i1}, \dots, e_{in_i} \right] \right\}$$

When the random variables $e_{i1}, \dots, e_{in_i}$ are all fixed, then the distribution of $\Gamma_i(\boldsymbol{\beta})$ is symmetric with mean $\alpha_i \bar{e}_{i\bullet}$. The distribution of $(v_i + e_{ij} - \Gamma_i(\boldsymbol{\beta})) / (d_{ij} \sigma_e)$ is symmetric with mean $(e_{ij} - \alpha_i \bar{e}_{i\bullet}) / (d_{ij} \sigma_e)$. Therefore,

$$\begin{aligned} \lambda_{ij} &= (1/\sigma_e^2) E_\xi [\bar{e}_{i\bullet} (e_{ij} - \alpha_i \bar{e}_{i\bullet})], \\ &= (1 - \alpha_i) / n_i. \end{aligned}$$

■

We define the two vectors $\bar{\mathbf{x}}_{i\eta}$ and $\bar{\mathbf{x}}_{i\lambda}$ by

$$\bar{\mathbf{x}}_{i\eta} = \frac{1}{n_i} \sum_{j=1}^{n_i} \eta_{ij} \mathbf{x}_{ij} \quad \text{and} \quad \bar{\mathbf{x}}_{i\lambda} = \frac{1}{n_i} \sum_{j=1}^{n_i} \lambda_{ij} \mathbf{x}_{ij}.$$

We have the following theorem that provides the first-order approximation to the MSE of the small area mean.

Theorem 5.3. Assume that the conditions of theorem 5.2 are satisfied and that

- (i) The matrices $\mathbf{Q}_m$ and $\mathbf{U}_m$ converge in probability elementwise to the two $p \times p$ positive-definite matrices $\mathbf{Q}$ and $\mathbf{U}$ respectively.
- (ii) $\sqrt{m} \hat{\mathbf{g}}(\boldsymbol{\beta})$ converges in distribution to the multivariate normal $\mathcal{N}_p(\mathbf{0}, \mathbf{U})$.
- (iii) There exists $\delta_0 > 0$ such that $\sup_{m \geq 1} E_\xi \|\sqrt{m}(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})\|^{2+\delta_0} = O(1)$
and $\sup_{m \geq 1} E_\xi \|\sqrt{m} \hat{\mathbf{g}}(\boldsymbol{\beta})\|^{2+\delta_0} = O(1)$.
- (iv) $\max_{1 \leq i \leq m} \left\{ \frac{1}{N_i} \sum_{j=1}^{N_i} \|\mathbf{x}_{ij}\|^3 \right\} = O(1)$.

Then under model ξ , the mean squared error of $\hat{\mu}_{Gi}$ is given by

$$\begin{aligned} \text{MSE}(\hat{\mu}_{Gi}) &= (1 - \alpha_i) \sigma_v^2 + \\ &\quad + \frac{1}{n} \left\{ (\bar{\mathbf{X}}_i^\top - \alpha_i \bar{\mathbf{x}}_i^\top) \mathbf{Q}^{-1} \mathbf{U} \mathbf{Q}^{-1} (\bar{\mathbf{X}}_i - \alpha_i \bar{\mathbf{x}}_i) \right\} + o(1/m). \end{aligned}$$

Remarks:

- (i) Assumption (i) ensures that for large samples, the matrices $\mathbf{Q}_m$ and $\mathbf{U}_m$ are positive-definite.
- (ii) Assumption (ii) is satisfied (Central Limit Theorem) under some standard regularity conditions.
- (iii) Conditions (iii) are equivalent to uniform integrability.
- (iv) It is important to note that if the ψ 's functions that are used here are all replaced by the identity function, the leading term of the above formula yields the customary formula of the MSE of the ξ -BLUP estimator of μ_i which is given by

$$MSE\{\hat{\mu}_{Bi}\} = g_{1i}(\boldsymbol{\sigma}^{(2)}) + g_{2i}(\boldsymbol{\sigma}^{(2)})$$

where $g_{1i}(\boldsymbol{\sigma}^{(2)}) = (1 - \alpha_i)\sigma_v^2$, $g_{2i}(\boldsymbol{\sigma}^{(2)}) = (\bar{\mathbf{X}}_i - \alpha_i\bar{\mathbf{x}}_i)^\top \mathbf{Q}_b^{-1}(\bar{\mathbf{X}}_i - \alpha_i\bar{\mathbf{x}}_i)/n$

and

$$\mathbf{Q}_b = (1/n) \sum_{i=1}^m \left(\sum_{j=1}^{n_i} \mathbf{x}_{ij} \mathbf{x}_{ij}^\top - n_i \alpha_i \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^\top \right).$$

This follows from the fact that with the identity function, the matrices $\mathbf{U}$ and $\mathbf{Q}$ are identical.

Proof of theorem 5.3 :

The MSE of $\hat{\mu}_{Gi}$ is given by:

$$MSE(\hat{\mu}_{Gi}) = E_\xi (\hat{\mu}_{Gi} - \mu_i)^2.$$

Since $\hat{\mu}_{Gi} = \bar{\mathbf{X}}_i^\top \hat{\boldsymbol{\beta}}_G + \alpha_i(\bar{y}_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_G)$, we have that

$$\begin{aligned}\hat{\mu}_{Gi} - \mu_i &= (\bar{\mathbf{X}}_i^\top - \alpha_i \bar{\mathbf{x}}_i^\top) (\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}) + \alpha_i(\bar{y}_i - \bar{\mathbf{x}}_i^\top \boldsymbol{\beta}) - v_i, \\ &= (\bar{\mathbf{X}}_i^\top - \alpha_i \bar{\mathbf{x}}_i^\top) (\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}) + \alpha_i \bar{e}_{i\bullet} - (1 - \alpha_i)v_i,\end{aligned}$$

where,

$$\bar{e}_{i\bullet} = \frac{1}{n_i} \sum_{j=1}^{n_i} e_{ij}.$$

Under the assumptions that the e_{ij} 's and the v_i 's are independently distributed, we have that.

$$\begin{aligned}\text{MSE}(\hat{\mu}_{Gi}) &= E_\xi (\hat{\mu}_{Gi} - \mu_i)^2, \\ &= (1 - \alpha_i)^2 \sigma_v^2 + \alpha_i^2 \sigma_e^2 / n_i \\ &\quad + (\bar{\mathbf{X}}_i^\top - \alpha_i \bar{\mathbf{x}}_i^\top) E_\xi (\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}) (\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})^\top (\bar{\mathbf{X}}_i - \alpha_i \bar{\mathbf{x}}_i) \\ &\quad - 2(1 - \alpha_i) E_\xi \left\{ v_i (\bar{\mathbf{X}}_i^\top - \alpha_i \bar{\mathbf{x}}_i^\top) (\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}) \right\} \\ &\quad + 2\alpha_i E_\xi \left\{ \bar{e}_{i\bullet} (\bar{\mathbf{X}}_i^\top - \alpha_i \bar{\mathbf{x}}_i^\top) (\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}) \right\}, \\ &= (1 - \alpha_i)^2 \sigma_v^2 + \alpha_i^2 \sigma_e^2 / n_i \\ &\quad + (\bar{\mathbf{X}}_i^\top - \alpha_i \bar{\mathbf{x}}_i^\top) E_\xi (\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}) (\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})^\top (\bar{\mathbf{X}}_i - \alpha_i \bar{\mathbf{x}}_i) \\ &\quad + 2(\bar{\mathbf{X}}_i^\top - \alpha_i \bar{\mathbf{x}}_i^\top) \left\{ \alpha_i E_\xi [\bar{e}_{i\bullet} (\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})] - (1 - \alpha_i) E_\xi [v_i (\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})] \right\}.\end{aligned}$$

It follows from the proof of theorem 5.2 that,

$$0 = \hat{\mathbf{g}}(\boldsymbol{\beta}) - \mathbf{Q}_m (\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}) + \mathbf{r}_m(\theta),$$

where $\mathbf{r}_m(\theta)$ is the remainder. Therefore,

$$\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta} = \mathbf{Q}_m^{-1} (\hat{\mathbf{g}}(\boldsymbol{\beta}) + \mathbf{r}_m(\theta)) \quad (5.16)$$

It follows from equation 5.16 that

$$E_{\xi}[v_i(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}_G)] = \mathbf{Q}_m^{-1} E_{\xi}[v_i \hat{\mathbf{g}}(\boldsymbol{\beta})] + \mathbf{Q}_m^{-1} E_{\xi}[v_i \mathbf{r}_m(\theta)],$$

$$E_{\xi}[\bar{e}_{i\bullet}(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}_G)] = \mathbf{Q}_m^{-1} E_{\xi}[\bar{e}_{i\bullet} \hat{\mathbf{g}}(\boldsymbol{\beta})] + \mathbf{Q}_m^{-1} E_{\xi}[\bar{e}_{i\bullet} \mathbf{r}_m(\theta)].$$

It follows from the independence between domains that

$$\begin{aligned} E_{\xi}[v_i \hat{\mathbf{g}}(\boldsymbol{\beta})] &= \frac{1}{n} \sum_{j=1}^{n_i} (1/\sigma_e) d_{ij} \mathbf{x}_{ij} E_{\xi} \left\{ v_i \psi_H \left(\frac{v_i + e_{ij} - \Gamma_i(\boldsymbol{\beta})}{d_{ij} \sigma_e} \right) \right\}, \\ &= (n_i/n) \bar{\mathbf{x}}_{i\eta} = \alpha_i \bar{\mathbf{x}}_i / n. \end{aligned}$$

and that

$$\begin{aligned} E_{\xi}[\bar{e}_{i\bullet} \hat{\mathbf{g}}(\boldsymbol{\beta})] &= \frac{1}{n} \sum_{j=1}^{n_i} (1/\sigma_e) d_{ij} \mathbf{x}_{ij} E_{\xi} \left\{ \bar{e}_{i\bullet} \psi_H \left(\frac{v_i + e_{ij} - \Gamma_i(\boldsymbol{\beta})}{d_{ij} \sigma_e} \right) \right\}, \\ &= (n_i/n) \bar{\mathbf{x}}_{i\lambda} = (1 - \alpha_i) \bar{\mathbf{x}}_i / n. \end{aligned}$$

Therefore,

$$\begin{aligned} &2(\bar{\mathbf{X}}_i^{\top} - \alpha_i \bar{\mathbf{x}}_i^{\top}) \left\{ \alpha_i E_{\xi}[\bar{e}_{i\bullet}(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})] - (1 - \alpha_i) E_{\xi}[v_i(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})] \right\} = \\ &2(\bar{\mathbf{X}}_i^{\top} - \alpha_i \bar{\mathbf{x}}_i^{\top}) \left\{ \alpha_i(1 - \alpha_i) \bar{\mathbf{x}}_i / n - \alpha_i(1 - \alpha_i) \bar{\mathbf{x}}_i / n + \right. \\ &\quad \left. + \mathbf{Q}_m^{-1} E_{\xi}[\alpha_i \bar{e}_{i\bullet} \mathbf{r}_m(\theta)] - \mathbf{Q}_m^{-1} E_{\xi}[(1 - \alpha_i) v_i \mathbf{r}_m(\theta)] \right\}, \\ &= 2(\bar{\mathbf{X}}_i^{\top} - \alpha_i \bar{\mathbf{x}}_i^{\top}) \mathbf{Q}_m^{-1} \left\{ E_{\xi}[\alpha_i \bar{e}_{i\bullet} \mathbf{r}_m(\theta)] - E_{\xi}[(1 - \alpha_i) v_i \mathbf{r}_m(\theta)] \right\}, \end{aligned}$$

assuming that the v_i 's and $\mathbf{Q}_m$ as well as the e_{ij} 's and $\mathbf{Q}_m$ are independent. The proof that $E_{\xi}[(1 - \alpha_i) v_i \mathbf{r}_m(\theta)] = o(1/m)$ and $E_{\xi}[\alpha_i \bar{e}_{i\bullet} \mathbf{r}_m(\theta)] = o(1/m)$ is obtained using techniques similar to those used to prove that $E_{\xi}[v_i \hat{\mathbf{g}}(\boldsymbol{\beta})] = O(1/m)$.

It follows again from the proof of theorem 5.2 that,

$$0 = \hat{\mathbf{g}}(\boldsymbol{\beta}) - \mathbf{Q}(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}) - (\mathbf{Q}_m - \mathbf{Q})(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}) + \mathbf{r}_m(\theta).$$

which implies that

$$\begin{aligned} (\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})^\top &= \mathbf{Q}^{-1}(\hat{\mathbf{g}}(\boldsymbol{\beta})\hat{\mathbf{g}}^\top(\boldsymbol{\beta}))\mathbf{Q}^{-1} + \\ &+ \mathbf{Q}^{-1}(\hat{\mathbf{g}}(\boldsymbol{\beta})\mathbf{d}_m^\top(\theta) + \mathbf{d}_m(\theta)\hat{\mathbf{g}}^\top(\boldsymbol{\beta}) + \mathbf{d}_m(\theta)\mathbf{d}_m^\top(\theta))\mathbf{Q}^{-1}. \end{aligned}$$

The MSE of $\hat{\mu}_{Gi}$ is then given by,

$$\begin{aligned} \text{MSE}(\hat{\mu}_{Gi}) &= (1 - \alpha_i)\sigma_v^2 + (\bar{\mathbf{X}}_i^\top - \alpha_i\bar{\mathbf{x}}_i^\top)\mathbf{Q}^{-1}\left[E_\xi(\hat{\mathbf{g}}(\boldsymbol{\beta})\hat{\mathbf{g}}^\top(\boldsymbol{\beta}))\right]\mathbf{Q}^{-1}(\bar{\mathbf{X}}_i - \alpha_i\bar{\mathbf{x}}_i) + \\ &+ A_i + o(1/m), \end{aligned}$$

where A_i is given by,

$$\begin{aligned} A_i &= (\bar{\mathbf{X}}_i^\top - \alpha_i\bar{\mathbf{x}}_i^\top)\mathbf{Q}^{-1}\left(E_\xi\hat{\mathbf{g}}(\boldsymbol{\beta})\mathbf{d}_m^\top(\theta) + E_\xi\mathbf{d}_m(\theta)\hat{\mathbf{g}}^\top(\boldsymbol{\beta}) + \right. \\ &\quad \left. + E_\xi\mathbf{d}_m(\theta)\mathbf{d}_m^\top(\theta)\right)\mathbf{Q}^{-1}(\bar{\mathbf{X}}_i - \alpha_i\bar{\mathbf{x}}_i). \end{aligned}$$

Because of the independence between domains, we have that,

$$\begin{aligned} E_\xi(\hat{\mathbf{g}}(\boldsymbol{\beta})\hat{\mathbf{g}}^\top(\boldsymbol{\beta})) &= \frac{1}{n^2} \sum_{i=1}^m \sum_{k=1}^m n_i n_k E(\hat{\mathbf{g}}_i(\boldsymbol{\beta})\hat{\mathbf{g}}_k^\top(\boldsymbol{\beta})), \\ &= \frac{1}{n^2} \sum_{i=1}^m n_i^2 E(\hat{\mathbf{g}}_i(\boldsymbol{\beta})\hat{\mathbf{g}}_i^\top(\boldsymbol{\beta})) + \frac{1}{n^2} \sum_{i \neq k} n_i n_k (E\hat{\mathbf{g}}_i(\boldsymbol{\beta}))(E\hat{\mathbf{g}}_k(\boldsymbol{\beta}))^\top. \end{aligned}$$

But, $\forall i = 1, \dots, m$,

$$E_\xi \hat{\mathbf{g}}_i(\boldsymbol{\beta}) = \frac{1}{n_i} \sum_{j=1}^{n_i} \sigma_e^{-1} d_{ij} \mathbf{x}_{ij} E\psi_H(\xi_{ij})$$

where

$$\xi_{ij} = \frac{v_i + e_{ij} - \alpha_i \bar{\sigma}_i \psi_H[(v_i + \bar{e}_{i\bullet})/\bar{\sigma}_i]}{d_{ij} \sigma_e}.$$

But

$$E_\xi \psi_H(\xi_{ij}) = E_\xi \left[E_\xi \psi_H(\xi_{ij}) \middle| v_i, e_{ik} (k = 1, \dots, n_i) \right] = E_\xi [(1 - \alpha_i) v_i / (d_{ij} \sigma_e)] = 0.$$

Thus.

$$E_{\xi}(\widehat{\mathbf{g}}(\boldsymbol{\beta})\widehat{\mathbf{g}}^{\top}(\boldsymbol{\beta})) = \mathbf{U}_m/n \text{ where } \mathbf{U}_m = \frac{1}{n} \sum_{i=1}^m n_i^2 E(\widehat{\mathbf{g}}_i(\boldsymbol{\beta})\widehat{\mathbf{g}}_i^{\top}(\boldsymbol{\beta})).$$

Or

$$E_{\xi}(\widehat{\mathbf{g}}(\boldsymbol{\beta})\widehat{\mathbf{g}}^{\top}(\boldsymbol{\beta})) = \mathbf{U}/n + o(1/m).$$

To complete the proof, we have to show that $A_i = o(1/m)$ for all $i = 1, \dots, m$ and this is done in appendix C.

■

For estimation purposes, it is important to have an estimator of the asymptotic MSE of $\widehat{\mu}_{Gi}$ developed in theorem 5.3 . That AMSE can be written as

$$\text{AMSE}(\widehat{\mu}_{Gi}) = g_{1i}(\boldsymbol{\sigma}) + g_{2i}(\boldsymbol{\sigma}), \quad (5.17)$$

where

$$\begin{cases} g_{1i}(\boldsymbol{\sigma}) = (1 - \alpha_i)\sigma_v^2, \\ g_{2i}(\boldsymbol{\sigma}) = (1/n)(\overline{\mathbf{X}}_i^{\top} - \alpha_i\overline{\mathbf{x}}_i^{\top})[\mathbf{Q}^{-1}\mathbf{U}\mathbf{Q}^{-1}](\overline{\mathbf{X}}_i^{\top} - \alpha_i\overline{\mathbf{x}}_i^{\top}). \end{cases}$$

We propose the following estimator of the MSE approximation:

$$\text{amse}(\widehat{\mu}_{Gi}) = g_{1i}(\boldsymbol{\sigma}) + \widehat{g}_{2i}(\boldsymbol{\sigma}), \quad (5.18)$$

where

$$\begin{cases} \widehat{g}_{2i}(\boldsymbol{\sigma}) = (1/n)(\overline{\mathbf{X}}_i^{\top} - \alpha_i\overline{\mathbf{x}}_i^{\top})[\widehat{\mathbf{Q}}_m^{-1}\widehat{\mathbf{U}}_m\widehat{\mathbf{Q}}_m^{-1}](\overline{\mathbf{X}}_i^{\top} - \alpha_i\overline{\mathbf{x}}_i^{\top}) \text{ and} \\ \widehat{\mathbf{U}}_m = \frac{1}{n} \sum_{k=1}^m n_k^2 (\widehat{\mathbf{g}}_k(\widehat{\boldsymbol{\beta}}_G)[\widehat{\mathbf{g}}_k(\widehat{\boldsymbol{\beta}}_G)]^{\top}). \end{cases}$$

The matrix $\hat{\mathbf{Q}}_m$ is obtained from $\mathbf{Q}_m$ by replacing the unknown parameters by appropriate estimators. The properties of this estimator will be studied by a simulation study.

Note that the formula of $\text{amse}(\hat{\mu}_{Gi})$ suggests another estimator for the MSE of the standard BLUP estimator of the domain mean discussed by Prasad and Rao (1990). It is given by

$$\text{amse}_R(\hat{\mu}_{Bi}) = (1 - \alpha_i)\sigma_v^2 + (\bar{X}_i - \alpha_i\bar{x}_i)^\top [\mathbf{Q}_b^{-1} \mathbf{U}_b \mathbf{Q}_b^{-1}] (\bar{X}_i - \alpha_i\bar{x}_i) / n.$$

Under the assumption of known variance components, the MSE of the BLUP estimator is given by

$$MSE\{\hat{\mu}_{Bi}(\sigma^{(2)})\} = (1 - \alpha_i)\sigma_v^2 + (\bar{X}_i - \alpha_i\bar{x}_i)^\top \mathbf{Q}_b^{-1} (\bar{X}_i - \alpha_i\bar{x}_i) / n$$

The $\mathbf{Q}_b$ -matrix is given by

$$\sigma_e^2 \mathbf{Q}_b = \frac{1}{n} \sum_{i=1}^m \left(\sum_{j=1}^{n_i} \mathbf{x}_{ij} \mathbf{x}_{ij}^\top - n_i \alpha_i \bar{x}_i \bar{x}_i^\top \right).$$

The $\mathbf{U}_b$ -matrix is given by,

$$\mathbf{U}_b = \frac{1}{n} \sum_{i=1}^m n_i^2 \mathbf{u}_i \mathbf{u}_i^\top,$$

where

$$\mathbf{u}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \alpha_i \bar{x}_i) e_{Bij} / \sigma_e^2,$$

and $e_{Bij} = y_{ij} - \mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_B$; $\hat{\boldsymbol{\beta}}_B$ being the BLUP estimator of $\boldsymbol{\beta}$. Prasad and Rao (1990) developed the second-order approximation to the mean squared error

from $MSE\{\hat{\mu}_{Bi}(\boldsymbol{\sigma}^{(2)})\}$ the formula of which is given above. Another possibility would be to develop a second-order approximation from $amse_R(\hat{\mu}_{Bi})$. The resulting estimator is likely to be more robust to model breakdowns.

5.4. Estimation of the Domain Mean Under the Assumption of Unknown Variance Components.

Our intention in this section is not to develop a comprehensive theory as in section 5.1. Instead, we will mention one possible approach towards extending the theory of section 5.1 to the case where the variance components are unknown. The use of robust estimators for the variance components leads to a fairly complicated statistical procedure. Therefore, some formulas available for standard procedures to assess the accuracy of estimates might not have counterparts in our set-up. In particular, the derivation of a second-order approximation to the MSE similar to that of Prasad and Rao (1990) for the EBLUP estimator is quite complicated although possible and is not attempted here.

The first problem taken up here is to derive a robust estimator $\hat{\boldsymbol{\sigma}}_G$ of the vector $\boldsymbol{\sigma}^\top = (\sigma_v, \sigma_e)$ of variance components. What we propose here is a modification of the method of fitting constants, also known as Henderson's method 3 (see Searle, 1971, chapter 10). Let $\hat{\boldsymbol{\beta}}_\bullet$ be the estimator of $\boldsymbol{\beta}$ given by

$$\hat{\boldsymbol{\beta}}_\bullet = \mathbf{Q}_\bullet^{-1} \sum_{i=1}^m \left(\sum_{j=1}^{n_i} \mathbf{x}_{ij} y_{ij} - n_i \bar{\mathbf{x}}_i \bar{y}_i \right),$$

where $\mathbf{Q}_\bullet$ is given by

$$\mathbf{Q}_\bullet = \sum_{i=1}^m \left[\sum_{j=1}^{n_i} \mathbf{x}_{ij} \mathbf{x}_{ij}^\top - n_i \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^\top \right]$$

and let $\hat{v}_{\bullet i} = \bar{y}_i - \bar{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}_\bullet$. The method of fitting constants yields the following unbiased estimator: $\hat{\boldsymbol{\sigma}}_0^\top = (\hat{\sigma}_{0v}, \hat{\sigma}_{0e})$ where,

$$\hat{\sigma}_{0e}^2 = \frac{1}{n - p - m + 1} \sum_{i=1}^m \sum_{j=1}^{n_i} \hat{e}_{\bullet ij}^2, \quad (5.19)$$

where

$$\hat{e}_{\bullet ij} = y_{ij} - (\mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_\bullet + \hat{v}_{\bullet i}),$$

and

$$\hat{\sigma}_{0v}^2 = \frac{\sum_{i=1}^m \sum_{j=1}^{n_i} e_{0ij}^2 - (n - p) \hat{\sigma}_{0e}^2}{n - \sum_{i=1}^m n_i^2 \bar{\mathbf{x}}_i^\top (\mathbf{X}^\top \mathbf{X})^{-1} \bar{\mathbf{x}}_i}, \quad (5.20)$$

where

$$\begin{cases} e_{0ij} &= y_{ij} - \mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_{1\bullet}, \\ \hat{\boldsymbol{\beta}}_{1\bullet} &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}. \end{cases}$$

It is assumed here that the matrix $\mathbf{X}$ has no column of 1's. That is, in standard models where $\mathbf{x}_{ij}$ is $(1, \mathbf{x}_{ij}^*)^\top$, only $\mathbf{x}_{ij}^*$ will be retained.

These equations must be suitably modified to reduce the sensitivity of the resulting estimators to outlying units. This problem was solved by Fellner (1986) by modifying restricted-maximum-likelihood estimates and under a general mixed linear model. Here, we propose a modification of the fitting constants method under the two-fold nested error regression model. The modification will be carried out along the lines of Fellner (1986) and Huber (1964).

5.4.1. The Proposed Estimators

The estimator $\hat{\sigma}_{G,e}^2$ is implicitly defined by the following equation:

$$\sum_{i=1}^m \sum_{j=1}^{n_i} \psi_H^2 \left(\frac{y_{ij} - (\mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_{\bullet G} + \hat{v}_{\bullet Gi})}{\hat{\sigma}_{G,e}} \right) = (n - p - m + 1) E_\Phi \psi_H^2(Z), \quad (5.21)$$

where Z is a random variable that is normally distributed with mean 0 and variance

1. The estimators $\hat{\boldsymbol{\beta}}_{\bullet G}$ and $\hat{v}_{\bullet Gi}$ are defined by the following system of equations:

$$\begin{cases} \sum_{i=1}^m \sum_{j=1}^{n_i} d_{ij} \mathbf{x}_{ij} \psi_H \left(\frac{y_{ij} - (\mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_{\bullet G} + \hat{v}_{\bullet Gi})}{d_{ij} \hat{\sigma}_{G,e}} \right) = 0, \\ \hat{v}_{\bullet Gi} = (\hat{\sigma}_{G,v}^2 + \hat{\sigma}_{G,e}^2/n_i)^{(1/2)} \psi_H \left(\frac{\bar{y}_i - \bar{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}_{\bullet G}}{(\hat{\sigma}_{G,v}^2 + \hat{\sigma}_{G,e}^2/n_i)^{(1/2)}} \right). \end{cases} \quad (5.22)$$

The problem here is that the second equation of (5.22) depends upon the estimator $\hat{\sigma}_{G,v}^2$ of $\sigma_{G,v}^2$ which is unspecified. We introduce an estimator by modifying the second equation of (5.20) as follows. Let $\hat{\boldsymbol{\beta}}_{1G}$ be a statistic implicitly defined by

$$\sum_{i=1}^m \sum_{j=1}^{n_i} d_{ij} \mathbf{x}_{ij} \psi_H \left(\frac{y_{ij} - \mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_{1G}}{d_{ij} s_G} \right) = 0, \quad (5.23)$$

where

$$s_G^2 = \hat{\sigma}_{G,v}^2 + \hat{\sigma}_{G,e}^2.$$

Then $\hat{\sigma}_{G,v}^2$ is implicitly given by the following equation

$$E_\Phi \psi_H^2(Z) = \frac{h - p}{n - p} \frac{\hat{\sigma}_{G,v}^2}{\hat{\sigma}_{G,v}^2 + \hat{\sigma}_{G,e}^2} + \frac{1}{n - p} \sum_{i=1}^m \sum_{j=1}^{n_i} \psi_H^2 \left(\frac{y_{ij} - \mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_{1G}}{(\hat{\sigma}_{G,v}^2 + \hat{\sigma}_{G,e}^2)^{1/2}} \right), \quad (5.24)$$

where h is given by,

$$h = \sum_{i=1}^m n_i^2 \bar{\mathbf{x}}_i^\top (\mathbf{X}^\top \mathbf{X})^{-1} \bar{\mathbf{x}}_i.$$

To see how equation (5.24) is motivated, note that equation of (5.20) can be rewritten as

$$n^* \hat{\sigma}_{0v}^2 + (n - p) \hat{\sigma}_{0e}^2 = \sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_{1\bullet})^2,$$

or

$$1 = \frac{h - p}{n - p} \frac{\sigma_v^2}{\hat{\sigma}_{0v}^2 + \hat{\sigma}_{0e}^2} + \frac{1}{n - p} \sum_{i=1}^m \sum_{j=1}^{n_i} \left(\frac{y_{ij} - \mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_{1\bullet}}{(\hat{\sigma}_{0v}^2 + \hat{\sigma}_{0e}^2)^{1/2}} \right)^2,$$

where

$$n^* = n - \sum_{i=1}^m n_i^2 \bar{\mathbf{x}}_i^\top (\mathbf{X}^\top \mathbf{X})^{-1} \bar{\mathbf{x}}_i,$$

To get the result, it is enough to replace the square function in brackets by ψ_H^2 .

The two-stage robust estimator of μ_i we propose is given by

$$\hat{\mu}_{Gi}(\hat{\boldsymbol{\sigma}}_G) = \bar{\mathbf{X}}_i^\top \hat{\boldsymbol{\beta}}_{\bullet G} + \hat{\alpha}_i (\bar{y}_i - \bar{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}_{\bullet G}),$$

where $\hat{\alpha}_i = \hat{\sigma}_{G,v}^2 (\hat{\sigma}_{G,v}^2 + \hat{\sigma}_{G,e}^2 / n_i)^{-2}$ and $\hat{\boldsymbol{\sigma}}_G = (\hat{\sigma}_{G,v}, \hat{\sigma}_{G,e})$.

The second important problem to deal with at this stage is to calculate and to estimate the asymptotic mean squared error of the two-stage GM-estimator (2GM) of the μ_i 's. This asymptotic mean squared error is defined as follows:

$$\begin{aligned} AMSE[\hat{\mu}_{Gi}(\hat{\boldsymbol{\sigma}}_G)] &= E_\xi \left[\hat{\mu}_{Gi}(\hat{\boldsymbol{\sigma}}_G) - \mu_i \right]^2, \\ &= E_\xi \left[\hat{\mu}_{Gi}(\boldsymbol{\sigma}_G) - \mu_i \right]^2 + \\ &\quad + E_\xi \left\{ \left[\hat{\mu}_{Gi}(\hat{\boldsymbol{\sigma}}_G) - \hat{\mu}_{Gi}(\boldsymbol{\sigma}_G) \right]^2 + 2 \left[\hat{\mu}_{Gi}(\hat{\boldsymbol{\sigma}}_G) - \hat{\mu}_{Gi}(\boldsymbol{\sigma}_G) \right] \left[\hat{\mu}_{Gi}(\boldsymbol{\sigma}_G) - \mu_i \right] \right\}. \end{aligned}$$

The first term on the right of the above inequality is in fact $AMSE[\hat{\mu}_{Gi}(\sigma_G)]$ which is given in theorem 5.3 . An explicit formula for the second term is intractable. However, a bootstrap procedure such as that of Laird and Louis (1987) can be used to overcome this problem. More research is needed in this area.

5.4.2. An Algorithm

Normally, if the variance components are assumed to be unknown, the whole problem of robust estimation of the domain mean μ_i will be solved only by finding a simultaneous solution to equations (5.21), (5.22), (5.24), and (5.23). At the end of the process, we will have the following statistics: (1) $\hat{\beta}_{\bullet G}$ (2) $\hat{\sigma}_G$ where $\hat{\sigma}_G^\top = (\hat{\sigma}_{G,v}, \hat{\sigma}_{G,e})$. We propose the following algorithm:

- (i) At the first iteration, obtain the variance components by setting $\hat{\sigma}_{G,v}^{2(1)} = \hat{\sigma}_{0v}^2$ and $\hat{\sigma}_{G,e}^{2(1)} = \hat{\sigma}_{0e}^2$ where $\hat{\sigma}_{0v}^2$ and $\hat{\sigma}_{0e}^2$ are the fitting-constant-method estimates of the variance components (see (5.20)). Then set $\hat{\beta}_{\bullet G}^{(1)} = \hat{\beta}_0$ where $\hat{\beta}_0$ is the BLUP estimator of β . It is then possible to calculate the $\hat{v}_{\bullet Gi}^{(1)}$'s and $\hat{\beta}_{1G}^{(1)}$.

Now, suppose that we have already performed t iterations and have to perform iteration $t + 1$.

- (ii) Given $\hat{\beta}_{\bullet G}^{(t)}$ and $\hat{v}_{\bullet Gi}^{(t)}$, obtain $\hat{\sigma}_{G,e}^{2(t+1)}$ by the following equation (derived from

5.21):

$$\hat{\sigma}_{G,e}^{2(t+1)} = \frac{1}{[n - (m + p) + 1] E_{\Phi} \psi_H^2(Z)} \sum_{i=1}^m \sum_{j=1}^{n_i} \omega_{ij}^{2(t)} \left(y_{ij} - (\mathbf{x}_{ij}^{\top} \hat{\boldsymbol{\beta}}_{\bullet G}^{(t)} + \hat{v}_{\bullet Gi}^{(t)}) \right)^2.$$

where $\omega_{ij}^{2(t)}$ is given by

$$\omega_{ij}^{2(t)} = \begin{cases} \psi_H^2(\xi_{ij}(\hat{\boldsymbol{\beta}}_{\bullet G}^{(t)})) / \xi_{ij}^2(\hat{\boldsymbol{\beta}}_{\bullet G}^{(t)}) & \text{if } \xi_{ij}(\hat{\boldsymbol{\beta}}_{\bullet G}^{(t)}) \neq 0, \\ 1 & \text{otherwise.} \end{cases}$$

and,

$$\begin{aligned} \xi_{ij}(\hat{\boldsymbol{\beta}}_{\bullet G}^{(t)}) &= \left[y_{ij} - (\mathbf{x}_{ij}^{\top} \hat{\boldsymbol{\beta}}_{\bullet G}^{(t)} + \hat{v}_{\bullet Gi}^{(t)}) \right] / \hat{\sigma}_{G,e}^{2(t)}, \\ \hat{v}_{\bullet Gi}^{(t)} &= \left(\hat{\sigma}_{G,v}^{2(t)} + \hat{\sigma}_{G,e}^{2(t)} / n_i \right)^{1/2} \psi_H \left(\frac{\bar{y}_i - \bar{\mathbf{x}}_i^{\top} \hat{\boldsymbol{\beta}}_{\bullet G}^{(t)}}{\left(\hat{\sigma}_{G,v}^{2(t)} + \hat{\sigma}_{G,e}^{2(t)} / n_i \right)^{1/2}} \right). \end{aligned}$$

(iii) Calculate $\hat{\sigma}_{G,v}^{2(t+1)}$ as follows,

$$\hat{\sigma}_{G,v}^{2(t+1)} = \frac{1}{(n - p) E_{\Phi} \psi_H^2(Z)} \left\{ (h - p) \hat{\sigma}_{G,v}^{2(t)} + \sum_{i=1}^m \sum_{j=1}^{n_i} \omega_{1ij}^{2(t)} \left(y_{ij} - \mathbf{x}_{ij}^{\top} \hat{\boldsymbol{\beta}}_{1G}^{(t)} \right)^2 \right\} - \hat{\sigma}_{G,e}^{2(t)}.$$

where $\omega_{1ij}^{2(t)}$ is given by,

$$\omega_{1ij}^{2(t)} = \begin{cases} \psi_H^2(\xi_{1ij}(\hat{\boldsymbol{\beta}}_{1G}^{(t)})) / \xi_{1ij}^2(\hat{\boldsymbol{\beta}}_{1G}^{(t)}) & \text{if } \xi_{1ij}(\hat{\boldsymbol{\beta}}_{1G}^{(t)}) \neq 0, \\ 1 & \text{otherwise,} \end{cases}$$

and

$$\xi_{1ij}(\hat{\boldsymbol{\beta}}_{1G}^{(t)}) = \frac{y_{ij} - \mathbf{x}_{ij}^{\top} \hat{\boldsymbol{\beta}}_{1G}^{(t)}}{\sqrt{\hat{\sigma}_{G,v}^{2(t)} + \hat{\sigma}_{G,e}^{2(t)}}}.$$

If $\hat{\sigma}_{G,v}^{2(t+1)} < 0$ then set $\hat{\sigma}_{G,v}^{2(t+1)} = 0$.

(iv) Update $\hat{\boldsymbol{\beta}}_{1G}^{(t)}$ using the following formula,

$$\hat{\boldsymbol{\beta}}_{1G}^{(t+1)} = \left(\sum_{i=1}^m \sum_{j=1}^{n_i} \omega_{1ij}^{(t)} \mathbf{x}_{ij} \mathbf{x}_{ij}^{\top} \right)^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} \omega_{1ij}^{(t)} \mathbf{x}_{ij} y_{ij}.$$

(v) Update $\hat{\boldsymbol{\beta}}_{\bullet G}^{(t)}$ using the following formula.

$$\hat{\boldsymbol{\beta}}_{\bullet G}^{(t+1)} = \left(\sum_{i=1}^m \sum_{j=1}^{n_i} \omega_{ij}^{(t)} \mathbf{x}_{ij} \mathbf{x}_{ij}^{\top} \right)^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} \omega_{ij}^{(t)} \mathbf{x}_{ij} (y_{ij} - \hat{v}_{\bullet Gi}^{(t)}).$$

(vi) If the convergence criterion is met, then stop. Otherwise, go back to step 2.

If $\Delta \|\hat{\boldsymbol{\beta}}_{\bullet G}\|$ is the change in $\|\hat{\boldsymbol{\beta}}_{\bullet G}\|$ between two consecutive iterations, then the convergence criterion for $\hat{\boldsymbol{\beta}}_{\bullet G}$ can be

$$\frac{\Delta \|\hat{\boldsymbol{\beta}}_{\bullet G}\|}{\|\hat{\boldsymbol{\beta}}_{\bullet G}\|} < \delta_0.$$

The convergence criterion for the variance components is more troublesome as some of them can be extremely small. Fellner (1986) suggested the following criterion:

$$\max\{|\Delta \hat{\sigma}_{G,v}^2|, |\Delta \hat{\sigma}_{G,e}^2|\} < \delta_0 (\hat{\sigma}_{G,v}^2 + \hat{\sigma}_{G,e}^2).$$

In his simulation study, Fellner (1986) used the value $\delta_0 = 0.001$ which seemed to yield satisfactory results.

5.5. Monte Carlo Study

The main purpose of this simulation experiment is to study (i) the efficiency of the GM-estimator $\hat{\mu}_{Gi}$ with respect to the BLUP estimator $\hat{\mu}_{Bi}$ of the i -th domain mean μ_i , under the assumption of known variance components; (ii) the bias of the estimator of the approximate $MSE(\hat{\mu}_{Gi})$ given in (5.18); (iii) the performance of the robust estimator $mse_R(\hat{\mu}_{Bi})$ of $MSE(\hat{\mu}_{Bi})$. The efficiency of $\hat{\mu}_{Gi}$ will be studied relative to some well-known alternatives such as the BLUP estimator, the regression synthetic estimator as well as the approximately unbiased regression estimator.

In order to facilitate the comparison, we have used a nested error regression model similar to that used by Prasad and Rao (1986) and given by

$$y_{ij} = 5.5 + 0.388x_{ij} + v_i + e_{ij}, \text{ for } i = 1, \dots, m \text{ and } j = 1, \dots, n_i.$$

In this experiment, $(n_1, \dots, n_m)$ is a simple random sample selected with replacement from the set $\{2, 3, 4, 5, 6\}$. Moreover, for any i and j , x_{ij} has a shifted exponential distribution $10.0 + \mathcal{E}(1/\lambda_i)$ with mean $10 + 1/\lambda_i$ where each λ_i has a uniform distribution on the interval $(1/10, 2/10)$. For such a model, the regression synthetic estimator of μ_i is given by:

$$\hat{\mu}_{Si} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}_i,$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the ordinary least squares estimators of β_0 and β_1 respectively.

assuming that the general form of the above model is given by:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + v_i + e_{ij}.$$

The approximately unbiased regression estimator is given by

$$\hat{\mu}_{Ri} = \bar{y}_i + \hat{\beta}_1(\bar{X}_i - \bar{x}_i).$$

We have generated 10,000 independent sets $\{e_{ij}, j = 1, \dots, n_i; i = 1, \dots, m\}$ and another 10,000 independent sets $\{v_i, i = 1, \dots, m\}$. Suppose that the distributions of e and v are $\mathcal{L}_e$ and $\mathcal{L}_v$ respectively. These independent sets were generated for the following cases:

(i) $\mathcal{L}_v = \mathcal{N}(0, 64)$ and $\mathcal{L}_e = \mathcal{N}(0, 5292)$, (ii) $\mathcal{L}_v = \mathcal{N}(0, 64)$ and $\mathcal{L}_e = \mathcal{D}(0, 5292)$.

(iii) $\mathcal{L}_v = \mathcal{N}(0, 64)$ and $\mathcal{L}_e = \delta \mathcal{N}(0, 5292\sqrt{x}) + (1 - \delta)\mathcal{N}(0, 5292), \delta \in (0, 1)$.

$\mathcal{D}(\mu, \sigma^2)$ stands for a double exponential distribution with mean μ and scale parameter σ^2 . Similarly, $\mathcal{N}$ symbol stands for normality and δ is simply the contamination rate.

Case (i) has been used in the study merely to assess the loss in efficiency that can be expected from the GM-estimator when the distribution of the errors have short tails. The other cases will show that the GM-estimator can lead to a substantial improvement when the distributions are contaminated or have long tails such as the double exponential.

In order to study the small sample properties of the proposed estimators, we assume that the number of domains is fairly small: $m = 15$. The results reported in table

5.1, are obtained under the assumption that the errors v_i 's and e_{ij} 's are normally distributed. The BLUP estimator is expected to be better. The purpose of this is to assess the loss in efficiency resulting from the use of the GM-estimator.

Table 5.1: Efficiencies of Small Area Mean Estimators and Percent Relative Bias for the Corresponding MSE Estimators
 $\mathcal{L}_v = \mathcal{N}(0, 64)$, $\mathcal{L}_e = \mathcal{N}(0, 5292)$ and $m = 15$.

n_i	$\text{MSE}_{\text{MC}}(\hat{\mu}_{\text{Bi}})$	Efficiency / BLUP			Percent Relative Bias		
		$\hat{\mu}_{\text{Gi}}$	$\hat{\mu}_{\text{Si}}$	$\hat{\mu}_{\text{Ri}}$	$\text{mse}_0(\hat{\mu}_{\text{Bi}})$	$\text{mse}_{\text{R}}(\hat{\mu}_{\text{Bi}})$	$\text{mse}(\hat{\mu}_{\text{Gi}})$
2	149.28	-0.78	-0.88	-93.70	-0.31	-5.74	-4.88
3	146.76	-0.66	-1.83	-91.49	0.44	-5.22	-3.97
4	142.41	-0.61	-1.87	-89.14	0.52	-4.57	-4.03
5	141.02	-0.68	-2.25	-86.67	0.14	-5.09	-4.39
6	138.19	-0.71	-2.64	-83.81	-1.39	-6.38	-5.83

Note that the Monte-Carlo MSE of a domain mean estimator $\hat{\mu}_i$ is given by

$$\text{MSE}_{\text{MC}}(\hat{\mu}_i) = \frac{1}{B} \sum_{b=1}^N (\hat{\mu}_{(b)i} - \mu_{(b)i})^2,$$

where B is the number of replicates used in the simulation study ($B=10,000$), $\mu_{(b)i}$ is the realized value of μ_i and $\hat{\mu}_{(b)i}$ is the corresponding estimator based on the sample b . Typically, we have

$$\mu_{(b)i} = \beta_0 + \beta_1 \bar{X}_i + v_{(b)i},$$

$v_{(b)i}$ being the realized value of the random variable v_i .

The efficiency of any estimator $\hat{\mu}_i$ with respect to the BLUP estimator is measured by

$$\text{eff}(\hat{\mu}_i) = 100 \times \left(\text{MSE}_{\text{MC}}(\hat{\mu}_{\text{Bi}}) / \text{MSE}_{\text{MC}}(\hat{\mu}_i) - 1 \right).$$

It follows from table 5.1 that the use of the GM-estimator will result in a loss of efficiency that ranges from -0.78% to -0.61%. The loss in efficiency of the synthetic is also fairly small although it seems to increase fast as the sample size increases. The regression estimator is very inefficient compared to its competitors.

Concerning the percent relative bias of the MSE estimators, the proposed estimator $mse(\hat{\mu}_{Gi})$ underestimates the actual MSE. The underestimation ranges from -5.83% to -3.97%. The other estimator $mse_R(\hat{\mu}_{Bi})$ to the MSE of the BLUP estimator is slightly more biased with a relative bias around -5%. Furthermore, $mse_0(\hat{\mu}_{Bi})$ is the theoretical MSE of the BLUP estimator. It is estimated and is simply compared here with the Monte-Carlo MSE.

In table 5.2, similar results are reported except that the errors e_{ij} 's are assumed to follow the double exponential distribution with mean 0 and the same variance as in the previous example.

Table 5.2: Efficiencies of Small Area Mean Estimators and Percent Relative Bias for the Corresponding MSE Estimators
 $\mathcal{L}_v = \mathcal{N}(0, 64)$, $\mathcal{L}_e = \mathcal{D}(0, 5292/2)$ and $m = 15$.

n_i	$MSE_{MC}(\hat{\mu}_{Bi})$	Efficiency / BLUP			Percent Relative Bias		
		$\hat{\mu}_{Gi}$	$\hat{\mu}_{Si}$	$\hat{\mu}_{Ri}$	$mse_0(\hat{\mu}_{Bi})$	$mse_R(\hat{\mu}_{Bi})$	$mse(\hat{\mu}_{Gi})$
2	161.64	7.47	-0.77	-92.32	-0.64	-9.95	-2.24
3	150.47	7.72	-1.23	-91.54	-2.16	-6.84	-5.28
4	145.13	7.48	-2.21	-88.92	-0.51	-6.65	-3.22
5	140.66	7.78	-2.18	-86.63	-0.42	-5.56	-3.22
6	147.28	7.08	-2.26	-83.44	-0.94	-9.68	-2.85

It follows from table 5.2 that when the distribution of the e_{ij} 's is double exponential, the GM-estimator takes over the BLUP estimator with a gain in efficiency that

ranges from 7.08% to 7.78%. This alternative to the BLUP estimator is much better than the other two competitors. Also, the proposed MSE estimator to the approximate MSE of $\hat{\mu}_{Gi}$ works fine as far as the relative bias is concerned. The estimator $mse(\hat{\mu}_{Bi})$ leads again to an underestimation of the MSE of the BLUP estimator that can go up to -9.95%. This is due to the small number of domains and this estimator is mainly designed to provide protection against model misspecification.

Table 5.3: Efficiencies of Small Area Mean Estimators and Percent Relative Bias for the Corresponding MSE Estimators
 $\mathcal{L}_v = \mathcal{N}(0, 64)$, $\mathcal{L}_e = (1 - \epsilon)\mathcal{N}(0, 292) + \epsilon\mathcal{N}(0, 100 * 292)$ and $m = 15$.

ϵ	n_i	$MSE_{MC}(\hat{\mu}_{Bi})$	Efficiency / BLUP			Percent Relative Bias		
			$\hat{\mu}_{Gi}$	$\hat{\mu}_{Si}$	$\hat{\mu}_{Ri}$	$mse_0(\hat{\mu}_{Bi})$	$mse_R(\hat{\mu}_{Bi})$	$mse(\hat{\mu}_{Gi})$
5%	2	118.43	59.06	-2.02	-89.80	0.91	-6.14	5.03
	3	112.98	52.05	-2.90	-89.22	-1.31	-4.93	1.76
	4	108.04	53.21	-3.93	-86.39	0.40	-4.46	4.56
	5	105.42	53.59	-4.72	-83.70	-0.86	-4.78	4.81
	6	106.12	58.41	-5.17	-80.16	0.93	-6.29	6.64
8%	2	153.73	73.57	-0.91	-91.61	0.02	-8.91	4.13
	3	144.25	64.83	-1.60	-91.12	-2.04	-6.32	0.45
	4	139.06	66.40	-2.00	-88.82	-0.44	-6.23	3.38
	5	135.74	66.16	-2.54	-86.33	-1.13	-5.86	3.51
	6	139.82	72.18	-2.66	-83.23	-0.21	-8.70	5.39
10%	2	177.73	76.33	-0.72	-92.21	-0.74	-9.58	3.09
	3	164.87	67.47	-1.18	-91.85	-2.46	-7.27	-0.53
	4	158.85	68.80	-1.45	-89.59	-0.55	-6.94	2.56
	5	155.89	69.02	-1.83	-87.22	-1.55	-6.83	2.46
	6	161.45	74.87	-1.91	-84.40	-0.56	-9.72	4.65

In table 5.1, the errors e_{ij} 's are normally distributed with mean 0 and variance $\sigma_e^2 = 5292$. We now assume that these errors follow a contaminated normal distribution $(1 - \epsilon)\mathcal{N}(0, 292) + \epsilon\mathcal{N}(0, 100 * 292)$ with a contamination rate ϵ taking the values 5%, 8% and 10%. This is a typical case where the GM-estimator is expected to be

very efficient.

It appears clearly from table 5.3 that the GM-estimator as well as the corresponding MSE estimator are much better than all of their competitors. The gain in efficiency can go well beyond 70% for a contamination rate of 10%. But even when the contamination rate is only 5% the gain in efficiency can be as high as 60%. The synthetic estimator yields a small loss in efficiency that can be as high as -5.17% for a contamination rate of 5% and as small as -0.72% for a contamination rate of 10%.

The performances of the MSE estimators are similar to what we observed in table 5.2.

Now we like to assess the accuracy of the MSE estimators for large values of m ($m = 60$), by calculating the percent relative bias of the corresponding estimators.

For the BLUP estimator, we have calculated

$$\begin{aligned} \text{BIAS}[\text{mse}_0(\hat{\mu}_{Bi})] &= 100.0 \times \frac{\text{mse}_0(\hat{\mu}_{Bi}) - \text{MSE}_{MC}(\hat{\mu}_{Bi})}{\text{MSE}_{MC}(\hat{\mu}_{Bi})}, \\ \text{BIAS}[\text{mse}_R(\hat{\mu}_{Bi})] &= 100.0 \times \frac{E_{MC}(\text{mse}_R(\hat{\mu}_{Bi})) - \text{MSE}_{MC}(\hat{\mu}_{Bi})}{\text{MSE}_{MC}(\hat{\mu}_{Bi})}. \end{aligned}$$

For the GM-estimator, we have calculated,

$$\text{BIAS}[\text{mse}(\hat{\mu}_{Gi})] = 100.0 \times \frac{E_{MC}[\text{mse}(\hat{\mu}_{Gi})] - \text{MSE}_{MC}(\hat{\mu}_{Gi})}{\text{MSE}_{MC}(\hat{\mu}_{Gi})},$$

where MSE_{MC} and E_{MC} stand for the Monte Carlo MSE and expectation of the estimator in question, respectively and $\text{mse}_0(\hat{\mu}_{Bi})$ is merely the MSE of the BLUP

estimator which need not be estimated when the variance components are known.

The results are shown in table 5.4 below.

Table 5.4: Percent Relative Bias for MSE Estimators
 $\mathcal{L}_v = \mathcal{N}(0, 64)$ and $m = 60$.

n_i	$\mathcal{L}_e = \mathcal{N}(0, 5292)$			$\mathcal{L}_e = \mathcal{D}(0, 5292/2)$		
	$\text{Bm}_0(\hat{\mu}_{\text{Bi}})$	$\text{Bm}_R(\hat{\mu}_{\text{Bi}})$	$\text{Bm}(\hat{\mu}_{\text{Gi}})$	$\text{Bm}_0(\hat{\mu}_{\text{Bi}})$	$\text{Bm}_R(\hat{\mu}_{\text{Bi}})$	$\text{Bm}(\hat{\mu}_{\text{Gi}})$
2	0.14	-0.44	-0.19	1.13	0.34	0.90
3	0.38	-0.28	-0.09	0.73	0.04	0.51
4	-0.15	-0.72	-0.41	0.29	-0.50	0.11
5	0.11	-0.44	-0.16	0.11	-0.67	0.05
6	0.55	-0.10	0.32	0.44	-0.22	0.39

where Bm stands for $\text{BIAS}[\text{mse}]$. It follows from table 5.4 that when the errors are normally distributed, then all MSE estimators have very small biases as expected. When the e_{ij} 's are distributed according to the double exponential distribution with mean 0 and same variance, the formula of $\text{mse}_0(\hat{\mu}_{\text{Bi}})$ still gives the exact MSE. The proposed estimator $\text{mse}_R(\hat{\mu}_{\text{Bi}})$ leads to an underestimation of about 10% of the true MSE while the estimator proposed for the MSE of $\hat{\mu}_{\text{Gi}}$ leads to an underestimation of about 3%.

CHAPTER 6

Summary and Outstanding Issues for Future Research

This chapter is made up of two sections. The first section provides a summary of the thesis while the second is devoted to issues that need further research.

6.1. Summary

This thesis deals with robust statistical inference in survey sampling. The goal is to contribute to the understanding and the application of the modern robust estimation theory in the context of finite population estimation. The theory of robust estimation was originally developed within the framework of parametric statistical inference in the early 1960's. Since then, it underwent considerable development and the related literature is abundant. Some of the most important techniques are implemented in some leading statistical packages.

However, the literature on the application of robust techniques to survey problems is very limited. Many attempts in the past are limited to the model-based approach which is close to the original framework within which the theory of robust estimation was developed. While design-based inference is certainly the most popular approach in sample surveys, the application of robust estimation techniques in this context was still lacking except some few articles. This thesis is an attempt to fill this gap with the concern of providing a general methodology for deriving robust estimators in a variety of scenarios.

This thesis is divided into two main parts. The first part which comprises the first four chapters addresses the classical problem in sample surveys of finite population mean estimation. The second part, made up of chapter 5, addresses the important problem of small area estimation.

Chapter 1 provides essentially a review of the literature on robust estimation. It appears that most articles are devoted to the outlier identification and treatment in survey sampling. But the outlier treatment is only a special case of the more general problem of robust estimation.

In chapter 2, the framework is that of the common mean superpopulation model. We have assumed that no auxiliary information is available. Under the assumption that the finite population is generated from a symmetrically distributed superpopulation, we have proposed an extension of Huber's work on M-estimators to general complex designs. Since the symmetry assumption on the superpopulation is impor-

tant, we have adopted the type of inference where the superpopulation model and the sampling design are taken into account. We have shown that under some regularity conditions, the M-estimator is consistent for estimating the finite population mean. We have proposed an estimator of its variance which proved very accurate in the simulation study presented at the end of the chapter. This simulation study also indicated that the gain in efficiency obtained from the M-estimator could be quite striking if the symmetry assumption on the superpopulation is approximately satisfied.

Afterwards, we relaxed the symmetry assumption on the superpopulation and introduced two new robust estimators: the CM-estimator and the LM-estimator. The CM-estimator (Consistent M-estimator) is designed to be an M-estimator which is consistent for estimating a finite population mean regardless of the finite population structure. We have proved the consistency of the CM-estimator under mild regularity conditions. An estimator of its variance was proposed as well. The CM-estimator is biased with a bias that in most cases will be of order not larger than $O(n^{-1/2})$ where n is the sample size. Because adjusting the CM-estimator for its bias will lead to a very unstable estimator, it was decided to estimate the mean squared error instead. We then proposed a consistent estimator of the mean squared error. The simulation study at the end of the chapter indicates that when the population is heavily skewed, the CM-estimator can be substantially more efficient than the standard estimators derived from the GREG methodology. Moreover, the proposed

MSE estimator of the CM-estimator works also well.

However, for more complex models and sampling designs, the CM-estimation approach might not lead to consistent estimators. In those situations, we have proposed the LM-estimator which is obtained by adjusting the M-estimator for the bias by a robust L-estimator similar to the classical trimmed mean. Hence the name LM-estimator. Under the common mean superpopulation model, the LM-estimator is not needed as the CM-estimator is always consistent, but performs satisfactorily. We have proved under some regularity conditions that the LM-estimator is consistent for estimating the finite population mean. A consistent estimator of its mean squared error is also proposed. Its performance in the simulation study is good.

In chapter 3, we extended the theory developed in chapter 2 to general linear superpopulation models. We have proved that all the robust estimators proposed in chapter 2 including the M-estimator share the interesting calibration properties of the GREG estimator. We also suggested a general methodology for deriving calibration LM-estimators. Robust estimators can be seen as weighted means where the weights depend upon the variable of interest. This poses a practical problem in large-scale surveys where thousands variables of interest are involved, as sample units cannot share common weights. We addressed this issue by deriving robust weights from key variables and using them for producing estimates for other variables of interest.

The use of complex regression models generates new problems such as that of the

influence of high leverage points. To solve this problem we suggested the use of GM-estimators instead of M-estimators. We have proved consistency of the proposed estimators under this general setting. Estimators for the MSE were developed as well. In order to get more insight, we ended this chapter by looking into the robust estimators under some well-known linear models and slightly complex samples.

Chapter 4 is entirely devoted to the study of jackknife estimators of the variances and the MSE's. Two main concerns motivated the development of the jackknife estimators. (i) The linearization method used to derive the asymptotic variance of the CM-estimator in chapter 3 leads to an approximate formula that requires a large sample size in order to have an acceptable accuracy. (ii) The derivation of the theoretical formula for the variance of the LM-estimator is quite complicated. For some J -functions, it might even be impossible.

We have developed a jackknife estimator for the variance of the CM-estimator. A jackknife estimator for the squared bias was also developed. We proved under some regularity conditions that this jackknife MSE estimator is correct for estimating the asymptotic MSE up to terms of order $O(1/n)$. Jackknife estimators were also developed for the LM-estimator. The simulation study at the end of chapter 4 indicates that even for small samples, the jackknife MSE estimator of the CM-estimator is much more accurate than the estimator obtained by the linearization method. The jackknife MSE estimator of the LM-estimator also work well in the simulation study. It is not better than the estimator obtained by the linearization

method. However, the fact that the jackknife methodology can be used to avoid the tedious algebra involved in deriving MSE estimators without losing too much is a big gain.

In chapter 5, after reviewing the literature on small area estimation and the BLUP estimation methodology, we proposed a robust alternative to the BLUP estimator for estimating small area means. We confined ourselves to the one-fold nested-error regression model under the assumption of known variance components for simplicity. The asymptotic MSE of the proposed GM-estimator was derived as well and an estimator thereof. It was proved that the leading term in the MSE formula is identical to that of the BLUP estimator. This indicates that the gain in efficiency can only be expected when the number of small areas is limited which will occur often in practice.

We have made some suggestions for the development of Empirical GM-estimators (EGM) by modifying the method of fitting constants for estimating the variance components. The properties of the EGM-estimator so obtained have not been studied.

The simulation study at the end of this chapter indicate that when the random effects do not follow the normal distribution, the GM-estimator yields a substantial gain in efficiency. Under normality of the random effects, however, there is a small loss in efficiency.

6.2. Future Research

The application of the theory of robust statistics to survey sampling problems presented in this thesis has addressed many theoretical as well as many practical estimation issues but has raised other questions where more research is needed.

- (i) We have used the large-sample properties of the Winsorized mean presented by Rivest and Hurtubise (1995) for choosing the robustness constant K_n of the CM-estimator. This can only be justified under the assumption that the CM-estimator and the Winsorized mean are asymptotically equivalent. It would be interesting to study the conditions under which $\widehat{Y}_{\text{CM}} - \widehat{Y}_w = O_p(1/n)$ where $\widehat{Y}_w$ is the Winsorized mean with a certain Winsorization cut-off.
 - (ii) The asymptotic properties of the CM and the LM estimators can be improved. The conditions under which the CM-estimator and the GREG estimator do not have the same asymptotic variance should be further clarified. In particular it is desirable to know the conditions under which $\widehat{Y}_{\text{CM}} - \widehat{Y}_{\text{GREG}} = O_p(1/n^\alpha)$ with $\alpha \leq 1/2$. A similar research may be pursued for the LM-estimator as well and the expected gain in efficiency over the GREG estimator can be made more precise.
 - (iii) Some theoretical results that can guide the choice of α_n and that of the parameter λ for the LM-estimator are highly desirable. At least a well-defined procedure for calculating these two quantities would be appropriate.
 - (iv) In multipurpose surveys where many variables of interest are involved, we have
-

suggested to calculate robust weights from a key variable of interest and apply them to produce estimates for the other variables of interest. However the robust weights are obtained from only one variable of interest. These weights would be more convincing if they were obtained from all variables of interest. This would require the development of multivariate M-estimators that can handle a number of variables simultaneously. This is a challenging research problem.

- (v) A traditional theoretical problem incidental to the use of M-estimators is the lack of consistency proof when the M-estimator is calculated simultaneously with the scale parameter. Throughout this thesis we had to assume that

$$\max_{t \geq 1} \hat{\sigma}_{Mt}/\sigma_{Mt} - 1 = o_p(1).$$

Although many simulation results support this assumption, it would be interesting to have a rigorous proof. A better use of the Ortega-Rheinbolt theorem (lemma 3.2) is likely to solve the problem.

- (vi) The CM-estimator and the LM-estimator introduced in chapter 2 are based on a compromise between the identity function and a bounded influence function. The CM-estimator is an M-estimator based on the influence function

$$\psi_{\bullet H}(r) = w_n r + (1 - w_n) \psi_H(r, K_n)$$

K_n being the robustness constant. Under simple random sampling, we often used $w_n = n/N$, the sampling fraction, as in Kokic and Bell (1994). This choice

is more difficult to justify under a complex design. It would be interesting to have a theory for choosing this weight function that takes the design into account. The approach that aims at minimizing the mean squared error must be considered. Also a robustness constant K_n where the sampling design is taken into account is needed. This problem is probably more tractable if studied for fully specified sampling designs.

- (vii) In studying the jackknife variance estimators in chapter 4, we confined ourselves to the simple random sampling design. Extensions to multistage designs in the Krewski-Rao set-up needs to be done.
- (viii) In chapter 5, we introduced the GM-estimator for estimating small area means under the one-fold nested-error regression model. The large-sample properties of the EGM-estimator remains to be studied. It would also be useful to derive the second-order MSE of the resulting estimator. Another possibility might be the bootstrap approach suggested by Laird and Louis (1987).

Extensions of the M-estimation theory to other small area models is another research problem.

APPENDIX A

Technical Proofs of chapter 3

Proof of lemma 3.3.

For any $\theta \in (0, 1)$ and any vector $(\mathbf{t}, \mathbf{B}) \in \mathbb{R}^p \times \mathbb{R}^p$ such that $\|\mathbf{t}\| = 1$, let $\widehat{\widehat{S}}_\theta(\mathbf{t}, \mathbf{B})$ be given by

$$\widehat{\widehat{S}}_\theta(\mathbf{t}, \mathbf{B}) = \sum_{i=1}^{n_t} \frac{(\mathbf{x}_i^\top \mathbf{t})^2}{N\pi_i\sigma_i^2} \psi'_{\bullet_H}(E_{\theta i}^*).$$

To prove lemma 3.2, it is enough to prove that there exists $\eta_0 > 0$ such that if $\|\mathbf{t}\| = 1$ and $\|\mathbf{B} - \mathbf{B}_{Gt}\| < \delta_0$ then

$$\lim_{t \rightarrow \infty} P\left(\widehat{\widehat{S}}_\theta(\mathbf{t}, \mathbf{B}) \geq \eta_0\right) = 1. \quad (\text{A.1})$$

To see this, note that $\forall \epsilon > 0, \exists \mathbf{t}_\epsilon \in \mathbb{R}^p$ such that $\|\mathbf{t}_\epsilon\| = 1$ satisfying

$$\widehat{\widehat{S}}_\theta(\mathbf{t}_\epsilon, \mathbf{B}) < \inf_{\|\mathbf{t}\|=1} \widehat{\widehat{S}}_\theta(\mathbf{t}, \mathbf{B}) + \epsilon.$$

Consequently, if (A.1) is true, then

$$1 = \lim_{t \rightarrow \infty} P\left(\widehat{\widehat{S}}_\theta(\mathbf{t}_\epsilon, \mathbf{B}) > 2\eta_0\right) \leq \lim_{t \rightarrow \infty} P\left(\inf_{\|\mathbf{t}\|=1} \widehat{\widehat{S}}_\theta(\mathbf{t}, \mathbf{B}) + \epsilon > 2\eta_0\right).$$

This relation is also satisfied for $\epsilon = \eta_0$ and the result follows.

Let $\mathbf{t}$ be an arbitrary p -dimensional vector such that $\|\mathbf{t}\| = 1$ and let $\eta_0 = l_0/2$ where l_0 is given by

$$l_0 = \lambda_0 \lim_{t \rightarrow \infty} \inf_{\|\mathbf{t}\|=1} \frac{1}{N_t} \sum_{i=1}^{N_t} (\mathbf{x}_i^{\top} \mathbf{t} / \sigma_i^2)^2.$$

(λ_0 is defined in assumption A4). It follows from assumptions A5 and A5 that $0 < \eta_0 < \infty$. By Chebychev's inequality we have that,

$$P\left(|\widehat{S}_\theta(\mathbf{t}, \mathbf{B}) - E[\widehat{S}_\theta(\mathbf{t}, \mathbf{B})]| < \eta_0\right) \geq 1 - \frac{V[\widehat{S}_\theta(\mathbf{t}, \mathbf{B})]}{\eta_0^2}.$$

However,

$$\begin{aligned} V[\widehat{S}_\theta(\mathbf{t}, \mathbf{B})] &\leq \frac{1}{N_t^2} \sum_{i=1}^{N_t} \frac{(\mathbf{x}_i^{\top} \mathbf{t} / \sigma_i)^4}{\pi_i} + \frac{1}{N_t^2} \sum_{i \neq j} (\mathbf{x}_i^{\top} \mathbf{t} / \sigma_i)^2 (\mathbf{x}_j^{\top} \mathbf{t} / \sigma_j)^2 \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right). \\ &\leq \left\{ \frac{1}{N_t \min_{1 \leq i \leq N_t} \pi_i} + \max_{1 \leq i \neq j \leq N_t} \left| \frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right| \right\} \frac{1}{N_t} \sum_{i=1}^{N_t} (\mathbf{x}_i^{\top} \mathbf{t} / \sigma_i)^4. \end{aligned}$$

It follows from the assumptions A1, A2 and A6 that $\lim_{t \rightarrow \infty} V[\widehat{S}_\theta(\mathbf{t}, \mathbf{B})] = 0$. Therefore,

$$\lim_{t \rightarrow \infty} P\left(|\widehat{S}_\theta(\mathbf{t}, \mathbf{B}) - E[\widehat{S}_\theta(\mathbf{t}, \mathbf{B})]| < \eta_0\right) = 1.$$

Note that

$$E[\widehat{S}_\theta(\mathbf{t}, \mathbf{B})] = \frac{1}{N} \sum_{i=1}^{N_t} (\mathbf{x}_i^{\top} \mathbf{t})^2 \psi'_{\bullet_H}(E_{\theta i}).$$

Using assumption A4 and the definition of l_0 introduced earlier, it is straightforward to see that $E[\widehat{S}_\theta(\mathbf{t}, \mathbf{B})] \geq l_0$. Therefore

$$\begin{aligned} P\left(|\widehat{S}_\theta(\mathbf{t}, \mathbf{B}) - E[\widehat{S}_\theta(\mathbf{t}, \mathbf{B})]| < \eta_0\right) &\leq P\left(\widehat{S}_\theta(\mathbf{t}, \mathbf{B}) > E[\widehat{S}_\theta(\mathbf{t}, \mathbf{B})] - \eta_0\right), \\ &\leq P\left(\widehat{S}_\theta(\mathbf{t}, \mathbf{B}) > l_0 - \eta_0\right) \leq P\left(\widehat{S}_\theta(\mathbf{t}, \mathbf{B}) > \eta_0\right). \end{aligned}$$

Consequently

$$\lim_{t \rightarrow \infty} P\left(\widehat{S}_\theta(\mathbf{t}, \mathbf{B}) > \eta_0\right) = 1.$$

■

Lemma A.1. *If the assumptions A4, A5, and A7 to A11 are satisfied then*

$$\mathbf{B}_{\text{Cmt}} - \mathbf{B}_{\text{Gt}} = O(n_t^{-1/2}).$$

Proof of lemma A.1

It follows from the mean value theorem that there exists $\theta \in (0, 1)$ such that for any

$\mathbf{B}$ in the neighborhood of $\mathbf{B}_{\text{Gt}}$ we have

$$f(\mathbf{B}) = f(\mathbf{B}_{\text{Gt}}) - \left(\frac{1}{N_t} \sum_{i=1}^{N_t} \sigma_i^{-2} \mathbf{x}_i \mathbf{x}_i^\top \psi'_{\bullet H}(E_{\theta i}^*) \right) (\mathbf{B} - \mathbf{B}_{\text{Gt}}),$$

where $E_{\theta i}^* = (y_i - \mathbf{x}_i^\top \mathbf{B}_\theta) / (d_i \sigma_i)$ and $\mathbf{B}_\theta = \theta \mathbf{B} + (1 - \theta) \mathbf{B}_{\text{Gt}}$. Therefore,

$$\begin{aligned} (\mathbf{B} - \mathbf{B}_{\text{Gt}})^\top f(\mathbf{B}) &= (\mathbf{B} - \mathbf{B}_{\text{Gt}})^\top \left\{ \frac{1}{N_t} \sum_{i=1}^{N_t} \sigma_i^{-1} d_i \mathbf{x}_i \psi_{\bullet H}(E_{\text{Gi},t}^*) \right\} - \\ &\quad - \frac{1}{N_t} \sum_{i=1}^{N_t} [\mathbf{x}_i^\top (\mathbf{B} - \mathbf{B}_{\text{Gt}}) / \sigma_i]^2 \psi'_{\bullet H}(E_{\theta i,t}^*), \\ &= (\mathbf{B} - \mathbf{B}_{\text{Gt}})^\top \left\{ \frac{1}{N_t} \sum_{i=1}^{N_t} \sigma_i^{-1} d_i \mathbf{x}_i [\psi_{\bullet H}(E_{\text{Gi},t}^*) - E_{\text{Gi},t}^*] \right\} - \\ &\quad - \frac{1}{N_t} \sum_{i=1}^{N_t} [\mathbf{x}_i^\top (\mathbf{B} - \mathbf{B}_{\text{Gt}}) / \sigma_i]^2 \psi'_{\bullet H}(E_{\theta i,t}^*), \end{aligned}$$

where $E_{\text{Gi},t}^* = (y_i - \mathbf{x}_i^\top \mathbf{B}_{\text{Gt}}) / (d_i \sigma_i)$. This follows from the fact that $\mathbf{B}_{\text{Gt}}$ is solution

to the equation

$$\sum_{i=1}^{N_t} \sigma_i^{-2} \mathbf{x}_i (y_i - \mathbf{x}_i^\top \mathbf{B}_{\text{Gt}}) = 0.$$

We have that

$$\begin{aligned}
 (\mathbf{B} - \mathbf{B}_{Gt})^\top \mathbf{f}(\mathbf{B}) &\leq \\
 &\leq \|\mathbf{B} - \mathbf{B}_{Gt}\| \left(\frac{1}{N_t} \sum_{i=1}^{N_t} (d_i \|\mathbf{x}_i\| / \sigma_i)^2 \right)^{1/2} \left(\frac{1}{N_t} \sum_{i=1}^{N_t} [\psi_{\bullet H}(E_{Gi,t}^*) - E_{Gi,t}^*]^2 \right)^{1/2} - \\
 &\quad - \|\mathbf{B} - \mathbf{B}_{Gt}\|^2 \inf_{\|\mathbf{t}\|=1} \frac{1}{N_t} \sum_{i=1}^{N_t} (\mathbf{x}_i^\top \mathbf{t} / \sigma_i)^2 \psi'_{\bullet H}(E_{\theta i,t}^*).
 \end{aligned} \tag{A.2}$$

It follows from assumption A11 that

$$\left(\frac{1}{N_t} \sum_{i=1}^{N_t} [\psi_{\bullet H}(E_{Gi,t}^*) - E_{Gi,t}^*]^2 \right)^{1/2} = O(n_t^{-1/2}).$$

Assumption A4, A5, A7 and equation A.2 imply that there exist two constants K_0

and $t(K_0)$ such that if $t > t(K_0)$ then

$$(\mathbf{B} - \mathbf{B}_{Gt})^\top \mathbf{f}(\mathbf{B}) \leq K_0 n_t^{-1/2} \|\mathbf{B} - \mathbf{B}_{Gt}\| - (l_0/2) \|\mathbf{B} - \mathbf{B}_{Gt}\|^2,$$

where

$$l_0 = \lambda_0 \lim_{t \rightarrow \infty} \inf_{\|\mathbf{t}\|=1} \frac{1}{N_t} \sum_{i=1}^{N_t} (\mathbf{x}_i^\top \mathbf{t})^2.$$

Note that if $\|\mathbf{B} - \mathbf{B}_{Gt}\| = (4K_0/l_0)n_t^{-1}$, then $(\mathbf{B} - \mathbf{B}_{Gt})^\top \mathbf{f}(\mathbf{B}) \leq -12(K_0^2/l_0)n_t^{-1} <$

0. It follows from lemma 3.1 (Ortega & Rheinbolt) that $\mathbf{B}_{Cmt} - \mathbf{B}_{Gt} = O(n_t^{-1})$.

■

Proof of theorem 3.3.

The problem is to prove that $\hat{\mathbf{B}}_{Cmt} - \mathbf{B}_{Gt} = O_p(n_t^{-1/2})$. By definition (see Bishop

et al., 1975, page 476), this means that for every $\eta > 0$ there exist two constants

$K(\eta)$ and t_η such that if $t > t_\eta$ then

$$P\left(\|\hat{\mathbf{B}}_{Cmt} - \mathbf{B}_{Gt}\| \leq \frac{K(\eta)}{n_t^{1/2}}\right) \geq 1 - \eta.$$

It follows from lemma 3.1 (Ortega & Rheinbolt) that for every constant K ,

$$P\left((\mathbf{B} - \mathbf{B}_{Gt})^\top \hat{\mathbf{f}}(\mathbf{B}) < 0, \forall \mathbf{B} : \|\mathbf{B} - \mathbf{B}_{Gt}\| = \frac{K}{n_t^{1/2}}\right) \leq P\left(\|\hat{\mathbf{B}}_{\text{CMT}} - \mathbf{B}_{Gt}\| \leq \frac{K}{n_t^{1/2}}\right).$$

as $\hat{\mathbf{B}}_{\text{CMT}}$ is solution to the equation $\hat{\mathbf{f}}(\mathbf{B}) = \mathbf{0}$. Let $\eta_0 > 0$. It is sufficient to prove the existence of $K(\eta_0)$ and t_{η_0} such that if $t > t_{\eta_0}$ then

$$P\left((\mathbf{B} - \mathbf{B}_{Gt})^\top \hat{\mathbf{f}}(\mathbf{B}) < 0, \forall \mathbf{B} : \|\mathbf{B} - \mathbf{B}_{Gt}\|^2 = n_t^{-1} K(\eta_0)\right) \geq 1 - \eta_0.$$

For any p -dimensional vector $\mathbf{B}$ we have that

$$(\mathbf{B} - \mathbf{B}_{Gt})^\top \hat{\mathbf{f}}(\mathbf{B}) = \sum_{i=1}^{n_t} \frac{d_i}{N\pi_i\sigma_i} [\mathbf{x}_i^\top (\mathbf{B} - \mathbf{B}_{Gt})] \psi_{\bullet H}[R_i(\mathbf{B})], \quad (\text{A.3})$$

where $R_i(\mathbf{B}) = (y_i - \mathbf{x}_i^\top \mathbf{B})/(d_i\sigma_i)$. Let us introduce some notations. $E_{Gi,t}^* = R_i(\mathbf{B}_{Gt})$, $E_{\theta i,t}^* = R_i(\mathbf{B}_\theta)$ where $\mathbf{B}_\theta = \theta\mathbf{B} + (1 - \theta)\mathbf{B}_{Gt}$ for some $\theta \in (0, 1)$. Note that

$$R_i(\mathbf{B}) = E_{Gi,t}^* - D_i(\mathbf{B}) \text{ where } D_i(\mathbf{B}) = \mathbf{x}_i^\top (\mathbf{B} - \mathbf{B}_{Gt})/(d_i\sigma_i).$$

It follows from the mean value theorem that

$$\psi_{\bullet H}[R_i(\mathbf{B})] = \psi_{\bullet H}(E_{Gi,t}^*) - D_i(\mathbf{B})\psi'_{\bullet H}(E_{\theta i,t}^*),$$

The equation A.3 together with the above mean value theorem implies that

$$\begin{aligned} (\mathbf{B} - \mathbf{B}_{Gt})^\top \hat{\mathbf{f}}(\mathbf{B}) &= \sum_{i=1}^{n_t} \frac{d_i}{N\pi_i\sigma_i} [\mathbf{x}_i^\top (\mathbf{B} - \mathbf{B}_{Gt})] \psi_{\bullet H}(E_{Gi,t}^*) - \\ &\quad - \sum_{i=1}^{n_t} \frac{1}{N\pi_i} [\mathbf{x}_i^\top (\mathbf{B} - \mathbf{B}_{Gt})/\sigma_i]^2 \psi'_{\bullet H}(E_{\theta i,t}^*), \\ &= \hat{G}_1(\mathbf{B}) - \hat{G}_2(\mathbf{B}). \end{aligned}$$

It follows from Chebychev's inequality that

$$P\left(\left|\widehat{G}_1(\mathbf{B}) - E[\widehat{G}_1(\mathbf{B})]\right| \leq \sqrt{\frac{V[\widehat{G}_1(\mathbf{B})]}{\eta_0/2}}\right) \geq 1 - \eta_0/2,$$

or

$$P\left(\widehat{G}_1(\mathbf{B}) \leq \left|E[\widehat{G}_1(\mathbf{B})]\right| + \sqrt{\frac{V[\widehat{G}_1(\mathbf{B})]}{\eta_0/2}}\right) \geq 1 - \eta_0/2.$$

But,

$$\left|E[\widehat{G}_1(\mathbf{B})]\right| \leq \|\mathbf{B} - \mathbf{B}_{\text{Gt}}\| \cdot |f(\mathbf{B}_{\text{Gt}})|$$

By the mean value theorem, we have $f(\mathbf{B}_{\text{Gt}}) = f(\mathbf{B}_{\text{Cmt}}) + [f'(\mathbf{B}_\theta)](\mathbf{B}_{\text{Gt}} - \mathbf{B}_{\text{Cmt}})$.

where $\mathbf{B}_\theta = \theta\mathbf{B}_{\text{Gt}} + (1 - \theta)\mathbf{B}_{\text{Cmt}}$ for some $\theta \in (0, 1)$. We have that

$$\left|E[\widehat{G}_1(\mathbf{B})]\right| \leq \|\mathbf{B} - \mathbf{B}_{\text{Gt}}\| \|\mathbf{B}_{\text{Gt}} - \mathbf{B}_{\text{Cmt}}\| \frac{1}{N_t} \sum_{i=1}^{N_t} (\|\mathbf{x}_i\|/\sigma_i)^2$$

It follows from assumption A7 and lemma A.1 that

$$\left|E[\widehat{G}_1(\mathbf{B})]\right| = O(n_t^{-1/2} \|\mathbf{B} - \mathbf{B}_{\text{Gt}}\|). \quad (\text{A.4})$$

Furthermore, it is straightforward to see that

$$\begin{aligned} V[\widehat{G}_1(\mathbf{B})] &\leq \left\{ \frac{1}{N_t \min_{1 \leq i \leq N_t} \pi_i} + \right. \\ &\quad \left. + \max_{1 \leq i \neq j \leq N_t} \left| \frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right| \right\} \frac{1}{N_t} \sum_{i=1}^{N_t} \left\{ [d_i \mathbf{x}_i^\top (\mathbf{B} - \mathbf{B}_{\text{Gt}}) / \sigma_i] \psi_{\bullet H}(E_{\text{Gi},t}^*) \right\}^2 \\ &\leq \|\mathbf{B} - \mathbf{B}_{\text{Gt}}\|^2 \left\{ \frac{1}{N_t \min_{1 \leq i \leq N_t} \pi_i} + \max_{1 \leq i \neq j \leq N_t} \left| \frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right| \right\} \times \\ &\quad \times \left(\frac{1}{N_t} \sum_{i=1}^{N_t} (d_i \|\mathbf{x}_i\| / \sigma_i)^4 \right)^{1/2} \left(\frac{1}{N_t} \sum_{i=1}^{N_t} E_{\text{Gi},t}^{*4} \right)^{1/2} \end{aligned}$$

Using assumptions A1, A2, A7 and A9, one gets

$$\sqrt{V[\widehat{G}_1(\mathbf{B})]} = O(n_t^{-1/2} \|\mathbf{B} - \mathbf{B}_{Gt}\|). \quad (A.5)$$

It follows from the equations A.4 and A.5 that there exist two constants $K_1(\eta_0)$ and $t_{1\eta_0}$ such that if $t > t_{1\eta_0}$ then

$$P\left(\widehat{G}_1(\mathbf{B}) \leq \left|E[\widehat{G}_1(\mathbf{B})]\right| + \sqrt{\frac{V[\widehat{G}_1(\mathbf{B})]}{\eta_0/2}}\right) \leq P\left(\widehat{G}_1(\mathbf{B}) \leq K_1(\eta_0)n_t^{-1/2}\|\mathbf{B} - \mathbf{B}_{Gt}\|\right)$$

Concerning $\widehat{G}_2(\mathbf{B})$, we have that

$$\begin{aligned} \widehat{G}_2(\mathbf{B}) &= \sum_{i=1}^{n_t} \frac{1}{N\pi_i\sigma_i^2} [\mathbf{x}_i^\top (\mathbf{B} - \mathbf{B}_{Gt})]^2 \psi'_{\bullet_H}(E_{\theta_i,t}^*), \\ &\geq \|\mathbf{B} - \mathbf{B}_{Gt}\|^2 \inf_{\|\mathbf{t}\|=1} \sum_{i=1}^{n_t} \frac{1}{N\pi_i\sigma_i^2} (\mathbf{x}_i^\top \mathbf{t})^2 \psi'_{\bullet_H}(E_{\theta_i,t}^*). \end{aligned}$$

It follows from lemma 3.2 that there exists $\eta_1 > 0$ such that for every vector $\mathbf{B}$ satisfying $\|\mathbf{B} - \mathbf{B}_{Gt}\| < \delta_0$,

$$\lim_{t \rightarrow \infty} P[\widehat{G}_2(\mathbf{B}) \geq \eta_1 \|\mathbf{B} - \mathbf{B}_{Gt}\|^2] = 1.$$

Therefore, there exists a constant $t_{2\eta_0}$ such that if $t > t_{2\eta_0}$ then

$$P\left(\widehat{G}_2(\mathbf{B}) < \eta_1 \|\mathbf{B} - \mathbf{B}_{Gt}\|^2\right) \leq \eta_0 \text{ for } \mathbf{B}: \|\mathbf{B} - \mathbf{B}_{Gt}\| < \delta_0.$$

If $t > \max(t_{1\eta_0}, t_{2\eta_0})$ and $\|\mathbf{B} - \mathbf{B}_{Gt}\| < \delta_0$ then

$$\begin{aligned} 1 - \eta_0 &\leq P\left(\widehat{G}_1(\mathbf{B}) \leq K_1(\eta_0)n_t^{-1/2}\|\mathbf{B} - \mathbf{B}_{Gt}\|\right) - P\left(\widehat{G}_2(\mathbf{B}) < \eta_1 \|\mathbf{B} - \mathbf{B}_{Gt}\|^2\right), \\ &\leq P\left(\widehat{G}_1(\mathbf{B}) - \widehat{G}_2(\mathbf{B}) \leq K_1(\eta_0)n_t^{-1/2}\|\mathbf{B} - \mathbf{B}_{Gt}\| - \eta_1 \|\mathbf{B} - \mathbf{B}_{Gt}\|^2\right). \end{aligned}$$

Let $K(\eta_0) = 2K_1(\eta_0)/\eta_1$. It is clear that if $t > \max(t_{1\eta_0}, t_{2\eta_0})$ then

$$P\left((\mathbf{B} - \mathbf{B}_{Gt})^\top \hat{\mathbf{f}}(\mathbf{B}) < 0, \forall \mathbf{B} : \|\mathbf{B} - \mathbf{B}_{Gt}\| = n_t^{-1/2} K(\eta_0)\right) \geq$$

$$P\left(\hat{G}_1(\mathbf{B}) - \hat{G}_2(\mathbf{B}) < 0, \forall \mathbf{B} : \|\mathbf{B} - \mathbf{B}_{Gt}\| = n_t^{-1/2} K(\eta_0)\right) \geq$$

$$P\left(\hat{G}_1(\mathbf{B}) - \hat{G}_2(\mathbf{B}) \leq -2(K_1^2(\eta_0)/\eta_1)n_t^{-1}\right) \geq 1 - \eta_0.$$

■

APPENDIX B

Some Proofs and Results Related to Chapter 4

Lemma A.1. *Under the assumption that the J_n -function is Lipschitz continuous and that $\sup_{t \geq 1} \int |r| dF_{Mt}(r) < \infty$, we have that $m_{1t}^* = o_p(1/n_t)$.*

Proof.

We have that,

$$\begin{aligned} -z_{0i} &= \int \left[(\delta_i(r) - F_M(r)) + (F_M(r) - \widehat{F}_{0M}(r)) \right] \left[J_n(F_M(r)) - \right. \\ &\quad \left. - (J_n(F_M(r)) - J_n(\widehat{F}_{0M}(r))) \right] dr, \\ &= -Z_i - \int (\delta_i(r) - F_M(r)) \left(J_n(F_M(r)) - J_n(\widehat{F}_{0M}(r)) \right) dr + \\ &\quad + \int (F_M(r) - \widehat{F}_{0M}(r)) J_n(F_M(r)) dr - \int (F_M(r) - \widehat{F}_{0M}(r)) \times \\ &\quad \times \left(J_n(F_M(r)) - J_n(\widehat{F}_{0M}(r)) \right) dr. \end{aligned}$$

Therefore, $z_{0i} - Z_i = \lambda_{0i} + \lambda_{1t} + \lambda_{2t}$, where

$$\begin{aligned}\lambda_{0i} &= \int (\delta_i(r) - F_M(r)) \left(J_n(F_M(r)) - J_n(\widehat{F}_{0M}(r)) \right) dr, \\ \lambda_{1t} &= - \int (F_M(r) - \widehat{F}_{0M}(r)) J_n(F_M(r)) dr, \\ \lambda_{2t} &= \int (F_M(r) - \widehat{F}_{0M}(r)) \left(J_n(F_M(r)) - J_n(\widehat{F}_{0M}(r)) \right) dr.\end{aligned}$$

Therefore,

$$\begin{aligned}m_{1t}^* &= \frac{(1-f)(1-w_n)^2}{n_t} \frac{1}{n_t-1} \sum_{i=1}^{n_t} \lambda_{0i}^2 + 2 \frac{(1-f)(1-w_n)^2}{n_t-1} (\lambda_{1t} + \lambda_{2t}) \bar{\lambda}_0 + \\ &\quad + \frac{(1-f)(1-w_n)^2}{n_t-1} (\lambda_{1t} + \lambda_{2t})^2.\end{aligned}\tag{A.1}$$

We have to prove that each member of the right-hand side of the above equation is $o_p(1/n_t)$. We start with the first member.

Note that

$$\begin{aligned}\frac{1}{n_t-1} \sum_{i=1}^{n_t} \lambda_{0i}^2 &= \frac{1}{n_t-1} \sum_{i=1}^{n_t} (\lambda_{0i}^* + \lambda_{0i}^{**})^2, \\ \text{where } \lambda_{0i}^* &= \int_0^{+\infty} (\delta_i(r) - F_M(r)) \left(J_n(F_M(r)) - J_n(\widehat{F}_{0M}(r)) \right) dr, \\ \lambda_{0i}^{**} &= \int_{-\infty}^0 (\delta_i(r) - F_M(r)) \left(J_n(F_M(r)) - J_n(\widehat{F}_{0M}(r)) \right) dr.\end{aligned}$$

This implies that,

$$\begin{aligned}\frac{1}{n_t-1} \sum_{i=1}^{n_t} \lambda_{0i}^2 &= \frac{1}{n_t-1} \sum_{i=1}^{n_t} \lambda_{0i}^{*2} + \frac{1}{n_t-1} \sum_{i=1}^{n_t} \lambda_{0i}^{**2} + \frac{2}{n_t-1} \sum_{i=1}^{n_t} \lambda_{0i}^* \lambda_{0i}^{**}, \\ &\leq \frac{1}{n_t-1} \sum_{i=1}^{n_t} \lambda_{0i}^{*2} + \frac{1}{n_t-1} \sum_{i=1}^{n_t} \lambda_{0i}^{**2} + 4 \left(\frac{1}{n_t} \sum_{i=1}^{n_t} \lambda_{0i}^{*2} \right)^{1/2} \left(\frac{1}{n_t} \sum_{i=1}^{n_t} \lambda_{0i}^{**2} \right)^{1/2}.\end{aligned}$$

We prove that $\frac{1}{n_t-1} \sum_{i=1}^{n_t} \lambda_{0i}^{*2} = o_p(1)$. The proof that $\frac{1}{n_t-1} \sum_{i=1}^{n_t} \lambda_{0i}^{**2} = o_p(1)$ is

very similar. We have that,

$$\begin{aligned} \frac{1}{n_t - 1} \sum_{i=1}^{n_t} \lambda_{0i}^{*2} &\leq \frac{1}{n_t - 1} \sum_{i=1}^{n_t} \left(\int_0^\infty (\delta_i(r) - F_M(r))^2 dr \right) \times \\ &\quad \times \left(\int_0^{+\infty} \left[J_n(\widehat{F}_{0M}(r)) - J_n(F_M(r)) \right]^2 dr \right). \end{aligned}$$

Since the J_n -function is Lipschitz continuous, there exists a constant C such that

$\forall t, s \in \mathbb{R}, |J_n(t) - J_n(s)| \leq C|t - s|$. Therefore,

$$\begin{aligned} E \left(\frac{1}{n_t - 1} \sum_{i=1}^{n_t} \lambda_{0i}^{*2} \right) &\leq \\ &\leq C^2 \left(\int_0^{+\infty} V[\widehat{F}_{0M}(r)] dr \right)^{1/2} \left\{ \int_0^{+\infty} E \left(\frac{1}{n_t - 1} \sum_{i=1}^{n_t} (\delta_i(r) - F_M(r))^2 \right) dr \right\}^{1/2}. \\ &\leq C^2 \left(\int_0^{+\infty} V[\widehat{F}_{0M}(r)] dr \right)^{1/2} \left(\int_0^\infty S_\delta^2(r) dr \right)^{1/2}, \end{aligned}$$

where $S_\delta^2(r) = \frac{1}{N_t - 1} \sum_{i=1}^{N_t} (\delta_i(r) - F_M(r))^2$. It is straightforward to see that

$$S_\delta^2(r) \leq \frac{2}{N_t} \sum_{i=1}^{N_t} (1 - F_M(r))^2 \leq 2(1 - F_M(r)).$$

Similarly, we have

$$V(\widehat{F}_{0M}(r)) = \frac{1-f}{n_t} S_\delta^2(r) \leq \frac{2}{n_t} (1 - F_M(r)).$$

Therefore,

$$E \left(\frac{1}{n_t - 1} \sum_{i=1}^{n_t} \lambda_{0i}^{*2} \right) \leq \frac{2C^2}{\sqrt{n_t}} \int_0^\infty (1 - F_M(r)) dr.$$

However, it is a well-known fact that

$$\begin{aligned} \int_{-\infty}^{+\infty} |r| dF_M(r) &= \int_0^{+\infty} (1 - F_M(r)) dr + \int_0^{+\infty} F_M(-r) dr, \\ &\geq \int_0^{+\infty} (1 - F_M(r)) dr. \end{aligned}$$

Thus,

$$E\left(\frac{1}{n_t - 1} \sum_{i=1}^{n_t} \lambda_{0i}^{*2}\right) \leq \frac{2C^2}{\sqrt{n_t}} \int_{-\infty}^{+\infty} |r| dF_M(r) < \infty.$$

It follows from Chebyshev's inequality that $\forall \epsilon > 0$,

$$\begin{aligned} P\left(\left|\frac{1}{n_t - 1} \sum_{i=1}^{n_t} \lambda_{0i}^{*2}\right| > \epsilon\right) &\leq \frac{1}{\epsilon} E\left(\frac{1}{n_t - 1} \sum_{i=1}^{n_t} \lambda_{0i}^{*2}\right), \\ &= O(1/n_t). \end{aligned}$$

This proves that $\frac{1}{n_t - 1} \sum_{i=1}^{n_t} \lambda_{0i}^{*2} = o_p(1)$.

Using a similar approach we can prove that $\frac{1}{n_t - 1} \sum_{i=1}^{n_t} \lambda_{0i}^{**2} = o_p(1)$. This proves that

$$\frac{(1-f)(1-w_n)^2}{n_t} \frac{1}{n_t - 1} \sum_{i=1}^{n_t} \lambda_{0i}^2 = o_p(1/n_t).$$

Now we prove that the second member on the right-hand side of equation (A.1) is $o_p(1/n_t)$.

It is straightforward to see that

$$\begin{aligned} \bar{\lambda}_0 &= \int \left(\hat{F}_{0M}(r) - F_M(r) \right) \left[J_n(F_M(r)) - J_n(\hat{F}_{0M}(r)) \right] dr, \\ &= -\lambda_{2t}. \end{aligned}$$

Therefore,

$$\begin{aligned} E(\lambda_{1t} + \lambda_{2t}) \bar{\lambda}_0 &= -\left(E\lambda_{1t}\lambda_{2t} + E\lambda_{2t}^2 \right), \\ &\leq \left(E\lambda_{1t}^2 \right)^{1/2} \left(E\lambda_{2t}^2 \right)^{1/2} + E\lambda_{2t}^2. \end{aligned}$$

By noting that $\lambda_{1t}^2 = (\lambda_{1t}^* + \lambda_{1t}^{**})^2$ where

$$\begin{aligned} \lambda_{1t}^* &= \int_0^{+\infty} [\hat{F}_{0M}(r) - F_M(r)] J_n[F_M(r)] dr \text{ and} \\ \lambda_{1t}^{**} &= \int_{-\infty}^0 [\hat{F}_{0M}(r) - F_M(r)] J_n[F_M(r)] dr. \end{aligned}$$

it can be proved that $E\lambda_{1t}^2 = O(1/\sqrt{n_t})$. This is obtained by the same used for proving that the first term of (B1) is $o_p(1/n_t)$. Similarly, we have that $E\lambda_{2t}^2 = O(1/\sqrt{n_t})$.

The last term to be dealt with only involves λ_{1t} and λ_{2t} and is necessarily $o_p(1/n_t)$.

This proves in fact that $m_{1t}^* = o_p(1/n_t)$.

■

Lemma A.2. *Under the assumption that the J_n -function is Lipschitz continuous.*

$\sup_{t \geq 1} \int |r| dF_{\text{Mt}}(r) < \infty$, and that $\lim_{t \rightarrow \infty} \frac{1}{N_t - 1} \sum_{i=1}^{N_t} U_i^2 = \sigma_U^2 > 0$, we have that $m_{2t}^ = o_p(1/n_t)$.*

Proof.

We have that

$$\begin{aligned} m_{2t} &= 2 \frac{(1-f)(1-w_n)}{n_t} \frac{1}{n_t - 1} \sum_{i=1}^{n_t} U_i (z_{0i} - Z_i), \\ &\leq \frac{4}{n_t} \left(\frac{1}{n_t} \sum_{i=1}^{n_t} U_i^2 \right)^{1/2} \left(\frac{1}{n_t} \sum_{i=1}^{n_t} (z_{0i} - Z_i)^2 \right)^{1/2}. \end{aligned}$$

We have just proved in lemma A.1 that

$$\frac{1}{n_t} \sum_{i=1}^{n_t} (z_{0i} - Z_i)^2 = o_p(1).$$

Furthermore, we have

$$\frac{1}{n_t} \sum_{i=1}^{n_t} U_i^2 = O_p \left(\frac{1}{N_t - 1} \sum_{i=1}^{N_t} U_i^2 \right).$$

■

Proof of lemma 4.6.

The Taylor series expansion of $\hat{g}_t$ in the neighborhood of B_{CM} is given by

$$\begin{aligned} \hat{g}_t(\hat{B}_{\text{CM}}) &= \hat{g}_t(B_{\text{CM}}) + (\hat{B}_{\text{CM}} - B_{\text{CM}})g'_t(B_{\text{CM}}) + (\hat{B}_{\text{CM}} - B_{\text{CM}})[\hat{g}'_t(B_{\text{CM}}) - \\ &\quad - g'_t(B_{\text{CM}})] + (1/2)(\hat{B}_{\text{CM}} - B_{\text{CM}})^2\hat{g}''_t(B_{\text{CM}}^{(\theta)}). \end{aligned}$$

It follows from this that

$$\begin{aligned} \hat{B}_{\text{CM}} - \hat{Y}_{\text{G}} &= (B_{\text{CM}} - \bar{Y}) - \left\{ [g'_t(B_{\text{CM}})]^{-1} \hat{g}_t(B_{\text{CM}}) + (\hat{Y}_{\text{G}} - \bar{Y}) \right\} - \\ &\quad - (g'_t(B_{\text{CM}}))^{-1} \left[(\hat{B}_{\text{CM}} - B_{\text{CM}})(\hat{g}'_t(B_{\text{CM}}) - g'_t(B_{\text{CM}})) + \right. \\ &\quad \left. + (1/2)(\hat{B}_{\text{CM}} - B_{\text{CM}})^2\hat{g}''_t(B_{\text{CM}}^{(\theta)}) \right]. \end{aligned} \quad (\text{A.2})$$

Therefore,

$$(\hat{B}_{\text{CM}} - \hat{Y}_{\text{G}})^2 = (B_{\text{CM}} - \bar{Y})^2 + [\hat{Q}_{\text{CM}} - (\hat{Y}_{\text{G}} - \bar{Y})]^2 + 2(B_{\text{CM}} - \bar{Y})[\hat{Q}_{\text{CM}} - (\hat{Y}_{\text{G}} - \bar{Y})] + \Delta_t. \quad (\text{A.3})$$

where,

$$\hat{Q}_{\text{CM}} = \frac{1}{n_t} \sum_{i=1}^{n_t} Q_{\text{CM}i}, \quad \Delta_t = u_t^2 + 2[\hat{Q}_{\text{CM}} - (\hat{Y}_{\text{G}} - B_{\text{CM}})]u_t \text{ and}$$

$$u_t = -\frac{1}{g'_t(B_{\text{CM}})} \left[(\hat{B}_{\text{CM}} - B_{\text{CM}})(\hat{g}'_t(B_{\text{CM}}) - g'_t(B_{\text{CM}})) + (1/2)(\hat{B}_{\text{CM}} - B_{\text{CM}})^2\hat{g}''_t(B_{\text{CM}}^{(\theta)}) \right].$$

It is quite straightforward to see that $b_{2\text{JK}} = O_p(1/n_t)$. Therefore to complete the proof, we simply have to prove that $\Delta(b_{2\text{JK}} < 0) = o_p(1)$, which will be done now.

Let $B_{\text{A}}(\hat{Y}_{\text{CM}}) = B_{\text{CM}} - \bar{Y}$ and $B_{0\text{A}}(\hat{Y}_{\text{CM}})$ be given by

$$B_{0\text{A}}(\hat{Y}_{\text{CM}}) = \begin{cases} B_{\text{A}}(\hat{Y}_{\text{CM}}) & \text{if } B_{\text{A}}(\hat{Y}_{\text{CM}}) \neq 0, \\ 1/n_t & \text{if } B_{\text{A}}(\hat{Y}_{\text{CM}}) = 0. \end{cases}$$

The reason for replacing $B_A(\widehat{Y}_{CM})$ with $B_{0A}(\widehat{Y}_{CM})$ is that $B_{0A}^2(\widehat{Y}_{CM})$ will always be strictly positive. Let $\epsilon > 0$. We have that

$$\begin{aligned} P\left(\Delta(b_{2JK} < 0) > \epsilon\right) &= P\left(\Delta(b_{2JK} < 0) = 1\right) = P(b_{2JK} < 0), \\ &= P\left(b_{2JK} - B_{0A}^2(\widehat{Y}_{CM}) < -B_{0A}^2(\widehat{Y}_{CM})\right), \\ &\leq P\left(|b_{2JK} - B_{0A}^2(\widehat{Y}_{CM})| > B_{0A}^2(\widehat{Y}_{CM})\right), \end{aligned}$$

It follows from Chebychev's inequality that

$$P\left(|b_{2JK} - B_{0A}^2(\widehat{Y}_{CM})| > B_{0A}^2(\widehat{Y}_{CM})\right) \leq \left(E|b_{2JK} - B_{0A}^2(\widehat{Y}_{CM})|\right) / B_{0A}^2(\widehat{Y}_{CM}).$$

It follows from (A.3) that

$$\begin{aligned} b_{2JK} - B_{0A}^2(\widehat{Y}_{CM}) &= \\ &= \left(B_A^2(\widehat{Y}_{CM}) - B_{0A}^2(\widehat{Y}_{CM})\right) + \left\{\left[\widehat{Q}_{CM} - (\widehat{Y}_G - \bar{Y})\right]^2 - \widehat{V}_{JK}(\widehat{B}_{CM} - \widehat{Y}_G)\right\} + \\ &\quad + 2(B_{CM} - \bar{Y})[\widehat{Q}_{CM} - (\widehat{Y}_G - \bar{Y})] + \Delta_t. \end{aligned}$$

Since, $E[\widehat{Q}_{CM} - (\widehat{Y}_G - \bar{Y})]^2 = V_A(\widehat{B}_{CM} - \widehat{Y}_G)$, we have that

$$\begin{aligned} E|b_{2JK} - B_{0A}^2(\widehat{Y}_{CM})| &\leq V_A(\widehat{B}_{CM} - \widehat{Y}_G) - E[\widehat{V}_{JK}(\widehat{B}_{CM} - \widehat{Y}_G)] + \\ &\quad + 2|B_A(\widehat{Y}_{CM})|\sqrt{V_A(\widehat{B}_{CM} - \widehat{Y}_G)} + E(\Delta_{0t}), \end{aligned}$$

where $\Delta_{0t} = \left(B_A^2(\widehat{Y}_{CM}) - B_{0A}^2(\widehat{Y}_{CM})\right) + \Delta_t$. It follows from assumption (iii) of theorem 4.2 that

$$E|b_{2JK} - B_{0A}^2(\widehat{Y}_{CM})| \leq 2|B_A(\widehat{Y}_{CM})|\sqrt{V_A(\widehat{B}_{CM} - \widehat{Y}_G)} + E(\Delta_{0t}) + o(1/n_t).$$

Since $B_{0A}^2(\widehat{Y}_{CM}) = O(1/n_t)$ we have that

$$\begin{aligned} \left(E|b_{2JK} - B_{0A}^2(\widehat{Y}_{CM})|\right) / B_{0A}^2(\widehat{Y}_{CM}) &\leq 2\sqrt{V_A(\widehat{B}_{CM} - \widehat{Y}_G)} / |B_{0A}(\widehat{Y}_{CM})| + \\ &\quad + E(\Delta_{0t}) / B_{0A}^2(\widehat{Y}_{CM}) + o(1). \end{aligned} \tag{A.4}$$

Note that,

$$\begin{aligned} V_A(\hat{B}_{CM} - \hat{Y}_G) &= \frac{1-f}{n_t} \frac{1}{N_t-1} \sum_{i=1}^{N_t} (Q_{CMi} - (y_i - \bar{Y}))^2, \\ &\leq \frac{2}{n_t} \left(\frac{1}{N_t} \sum_{i=1}^{N_t} (Q_{CMi} - E_{CMi})^2 \right) + B_A^2(\hat{Y}_{CM}). \end{aligned}$$

Let $\tilde{E}_{CMi} = E_{CMi}/\sigma_{CM}$ and λ_{1t} given by

$$\lambda_{1t} = \frac{1}{N_t} \sum_{i=1}^{N_t} \psi'_{\bullet H}(\tilde{E}_{CMi}, K_n).$$

It is quite straightforward to see that

$$Q_{CMi} - E_{CMi} = (1 - w_{n_t})\lambda_{1t}^{-1} \left[\sigma_{CM}\psi_H(\tilde{E}_{CMi}, K_n) - E_{CMi} \right] - (1 - \lambda_{1t}^{-1})E_{CMi}.$$

Therefore,

$$\begin{aligned} (Q_{CMi} - E_{CMi})^2 &= (1 - w_{n_t})^2 \lambda_{1t}^2 \left(\sigma_{CM}\psi_H(\tilde{E}_{CMi}) - E_{CMi} \right)^2 + (1 - \lambda_{1t}^{-1})^2 E_{CMi}^2 - \\ &\quad - 2(\lambda_{1t}^{-1} - 1)(1 - w_{n_t})\lambda_{1t} \left[E_{CMi}^2 - \sigma_{CM}\psi_H(\tilde{E}_{CMi}, K_n) \times E_{CMi} \right], \\ &\leq \left(\sigma_{CM}\psi_H(\tilde{E}_{CMi}) - E_{CMi} \right)^2 + (1 - \lambda_{1t}^{-1})^2 E_{CMi}^2. \end{aligned}$$

The last inequality follows from the fact that $0 \leq \lambda_{1t} \leq 1$. Now we have that

$$\left(\sigma_{CM}\psi_H(\tilde{E}_{CMi}) - E_{CMi} \right)^2 \leq \frac{1}{K_n^2 \sigma_{CM}^2} E_{CMi}^4.$$

This follows from the assumption that $|\psi_H(r, K_n) - r| \leq r^2/K_n \forall r \in \mathbb{R}$. Therefore,

$$V_A(\hat{B}_{CM} - \hat{Y}_G) \leq \frac{2}{n_t} \left(\frac{1}{K_n^2 \sigma_{CM}^2} \times \frac{1}{N_t} \sum_{i=1}^{N_t} E_{CMi}^4 + (1 - \lambda_{1t}^{-1})^2 \times \frac{1}{N_t} \sum_{i=1}^{N_t} E_{CMi}^2 \right),$$

and

$$\begin{aligned} \sqrt{V_A(\hat{B}_{CM} - \hat{Y}_G)} &\leq \frac{2}{\sqrt{n_t}} \left\{ \frac{1}{K_n \sigma_{CM}} \left(\frac{1}{N_t} \sum_{i=1}^{N_t} E_{CMi}^4 \right)^{1/2} + |1 - \lambda_{1t}^{-1}| \times \right. \\ &\quad \left. \times \left(\frac{1}{N_t} \sum_{i=1}^{N_t} E_{CMi}^2 \right)^{1/2} \right\}. \end{aligned}$$

It follows from assumption (i) of theorem 4.2 that

$$\begin{aligned} \sqrt{V_A(\widehat{B}_{CM} - \widehat{Y}_G)} / |B_{0A}(\widehat{Y}_{CM})| &= O(n_t^{-1/2}) \left(O(1/K_n) + O(|1 - \lambda_{1t}^{-1}|) \right) O(K_n), \\ &= O(n_t^{-1/2}) + O(n_t^{-1/2} K_n |1 - \lambda_{1t}^{-1}|). \end{aligned}$$

We have $|1 - \lambda_{1t}| = O(1/K_n)$. In fact,

$$\lambda_{1t} = w_{n_t} + (1 - w_{n_t}) \frac{1}{N_t} \sum_{i=1}^{N_t} \psi'_H(\tilde{E}_{CMi})$$

and

$$\begin{aligned} 1 - \lambda_{1t} &= (1 - w_{n_t}) \frac{1}{N_t} \sum_{i=1}^{N_t} \left\{ 1 - \left(\psi'_H(\tilde{E}_{CMi}) \right) \right\} \\ &\leq \frac{2}{K_n} \frac{1}{N_t} \sum_{i=1}^{N_t} |\tilde{E}_{CMi}|. \end{aligned}$$

This follows from the assumption that $\forall r \in \mathbb{R}$, $|1 - \psi'_H(r, K_n)| < 2|r|/K_n$. It follows from this that, $O(n_t^{-1/2} K_n |1 - \lambda_{1t}^{-1}|) = O(n_t^{-1/2})$ since $|1 - \lambda_{1t}^{-1}| = O(1/K_n)$.

It follows from equation (A.4) that it is enough to prove $E(\Delta_{0t}) = o(n_t^{-1})$ to complete the proof.

Because the functions $\widehat{g}'_t(\cdot)$ and $g'_t(\cdot)$ are bounded by 1, it is easy to check that $\Delta_{0t} = o_p(1/n_t)$. It follows from lemma 4.1 that if for some $\epsilon > 0$, we have $\sup_{t \geq 1} E|n_t \Delta_{0t}|^{1+\epsilon} < \infty$ then $E(\Delta_{0t}) = o(1/n_t)$. Therefore, it is sufficient to prove that $E|n_t \Delta_{0t}|^{3/2} < \infty$ ($\epsilon = 1/2$).

Note that $E|n_t \Delta_{0t}|^{3/2} \leq \sqrt{E|n_t \Delta_{0t}|^3}$. Therefore, it is enough to prove that $E|n_t \Delta_{0t}|^3 < \infty$. We recall that for any real numbers a , b and r we have the C_r -inequality $(|a| + |b|)^r \leq 2^r(|a|^r + |b|^r)$. Therefore, using the C_r -inequality,

$$\begin{aligned} E|n_t \Delta_{0t}|^3 &\leq 2^3 \left\{ [n_t |B_A^2(\widehat{Y}_{CM}) - B_{0A}^2(\widehat{Y}_{CM})|]^3 + E(n_t |\Delta_t|)^3 \right\}, \\ &= O(n_t^{-3}) + 2^3 E(n_t |\Delta_t|)^3. \end{aligned}$$

From this, we only have to prove that $E(n_t|\Delta_t|)^3 < \infty$. We have $\Delta_t = u_t^2 + 2u_t[\widehat{Q}_{\text{CM}} - (\widehat{Y}_G - \bar{Y}) - (\bar{Y} - B_{\text{CM}})]$. Therefore, applying the C_r -inequality many times, we get,

$$\begin{aligned} (n_t|\Delta_t|)^3 &\leq 2^3(\sqrt{n_t}u_t)^6 + 2^6(\sqrt{n_t}u_t)^3 \times (\sqrt{n_t}|\widehat{Q}_{\text{CM}}|)^3 + \\ &\quad + 2^9(\sqrt{n_t}u_t)^3(\sqrt{n_t}|\widehat{Y}_G - \bar{Y}|)^3 + 2^9(\sqrt{n_t}u_t)^3(\sqrt{n_t}|\bar{Y} - B_{\text{CM}}|)^3. \end{aligned}$$

Using Schwartz's inequality we get,

$$\begin{aligned} E(n_t|\Delta_t|)^3 &\leq 2^3E(\sqrt{n_t}u_t)^6 + 2^6E^{1/2}(\sqrt{n_t}u_t)^6 \times E^{1/2}(\sqrt{n_t}|\widehat{Q}_{\text{CM}}|)^6 + \\ &\quad + 2^9E^{1/2}(\sqrt{n_t}u_t)^6 \cdot E^{1/2}(\sqrt{n_t}|\widehat{Y}_G - \bar{Y}|)^6 + \\ &\quad + 2^9(\sqrt{n_t}|\bar{Y} - B_{\text{CM}}|)^3E^{1/2}(\sqrt{n_t}u_t)^6. \end{aligned}$$

The proof would be completed if we prove that $E(\sqrt{n_t}u_t)^6 < \infty$ uniformly in t .

Since the functions $\widehat{g}'_t(\cdot)$ and $g'_t(\cdot)$ are bounded by 1 and $\widehat{g}''_t(\cdot)$ is bounded by a constant, we have that

$$(\sqrt{n_t}u_t)^6 \leq 2^6(g'_t(B_{\text{CM}}))^{-1} \left[2^6(\sqrt{n_t}(\widehat{B}_{\text{CM}} - B_{\text{CM}}))^6 + C_0 n_t^{-3} (\sqrt{n_t}(\widehat{B}_{\text{CM}} - B_{\text{CM}}))^{12} \right].$$

The result follows from condition (ii) of theorem 4.2.

■

APPENDIX C

Some Proofs and Results Related to Chapter 5

Lemma C.1

Under the conditions of theorem 5.3 $E_\xi(A_i) = o(1/m)$ where A_i is given by

$$A_i = (\bar{X}_i - \alpha_i \bar{\mathbf{x}}_i)^\top \mathbf{Q}^{-1} \left[\hat{\mathbf{g}}(\boldsymbol{\beta}) \mathbf{d}_m^\top(\theta) + \mathbf{d}_m(\theta) \hat{\mathbf{g}}^\top(\boldsymbol{\beta}) + \mathbf{d}_m(\theta) \mathbf{d}_m^\top(\theta) \right] \mathbf{Q}^{-1} (\bar{X}_i - \alpha_i \bar{\mathbf{x}}_i).$$

Proof.

It follows from theorem C of Serfling (1980, page 15) that it is enough to prove that

$$(i) \quad A_i = o_p(1/m),$$

$$(ii) \quad \sup_{m \geq 1} E_\xi |mA_i|^{1+\epsilon} = O(1) \text{ for some } \epsilon \text{ (} 0 < \epsilon < 1 \text{)}.$$

Following the proof of theorem 5.3, we have that,

$$\hat{\mathbf{g}}(\boldsymbol{\beta}) \mathbf{d}_m^\top(\theta) = -\hat{\mathbf{g}}(\boldsymbol{\beta}) (\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})^\top (\mathbf{Q}_m - \mathbf{Q}) + \hat{\mathbf{g}}(\boldsymbol{\beta}) \mathbf{r}_m^\top(\theta).$$

It follows from assumption (ii) of theorem 5.3 and Slutsky's theorem that the asymptotic distribution of $\sqrt{m}(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})$ is a multivariate normal distribution $\mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = \mathbf{Q}^{-1} \mathbf{U} \mathbf{Q}^{-1}$. Therefore, assumptions (i) and (ii) of theorem 5.3 imply that,

$$\begin{aligned} \hat{\mathbf{g}}(\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})^\top (\mathbf{Q}_m - \mathbf{Q}) &= O_p(m^{-1/2}) O_p(m^{-1/2}) o_p(1) = o_p(1/m), \\ \hat{\mathbf{g}}(\boldsymbol{\beta}) \mathbf{r}_m^\top(\theta) &= (1/2) \hat{\mathbf{g}}(\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})^\top \hat{\mathbf{g}}_2(\boldsymbol{\beta}_\theta)(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}), \\ &= O_p(m^{-3/2}) = o_p(1/m). \end{aligned}$$

Thus, $\hat{\mathbf{g}}(\boldsymbol{\beta}) \mathbf{d}_m^\top(\theta) = o_p(1/m)$.

Furthermore, we have that

$$\begin{aligned} \mathbf{d}_m(\theta) \mathbf{d}_m^\top(\theta) &= [(\mathbf{Q}_m - \mathbf{Q})(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}) + \mathbf{r}_m(\theta)] [(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})^\top (\mathbf{Q}_m - \mathbf{Q}) + \mathbf{r}_m^\top(\theta)], \\ &= (\mathbf{Q}_m - \mathbf{Q})(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})^\top (\mathbf{Q}_m - \mathbf{Q}) + \mathbf{r}_m(\theta) \mathbf{r}_m^\top(\theta) + \\ &\quad + \mathbf{r}_m(\theta)(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})^\top (\mathbf{Q}_m - \mathbf{Q}) + (\mathbf{Q}_m - \mathbf{Q})(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}) \mathbf{r}_m^\top(\theta). \\ &= o_p(1/m) + o_p(m^{-3/2}) + o_p(1/m) + o_p(1/m^2). \end{aligned}$$

Thus, $A_i = o_p(1/m)$.

The next step is to show that there exists $\epsilon_0 \in (0, 1)$ such that

$$\sup_{m \geq 1} E_\xi |mA_i|^{1+\epsilon_0} = O(1).$$

For any ϵ_0 fixed, we have that,

$$\begin{aligned} |mA_i|^{1+\epsilon_0} &\leq \left\{ (\|\bar{\mathbf{X}}_i\| + \|\bar{\mathbf{x}}_i\|)^2 \|\mathbf{Q}^{-1}\|^2 \left\| \sqrt{m} \mathbf{d}_m(\theta) \times \sqrt{m} \hat{\mathbf{g}}^\top(\boldsymbol{\beta}) + \right. \right. \\ &\quad \left. \left. \sqrt{m} \hat{\mathbf{g}}(\boldsymbol{\beta}) \times \sqrt{m} \mathbf{d}_m^\top(\theta) + \sqrt{m} \mathbf{d}_m(\theta) \times \sqrt{m} \mathbf{d}_m(\theta) \right\|^{1+\epsilon_0} \right\}. \end{aligned}$$

Recall that for any $p \times p$ -matrix $\mathbf{M}$,

$$\|\mathbf{M}\|^2 = \sum_{i=1}^p \sum_{j=1}^p M_{ij}^2.$$

In what follows, we will need the following C_r inequality:

$$E|X + Y|^r \leq C_r (E|X|^r + E|Y|^r) \text{ where } C_r = 1 \text{ if } r < 1 \text{ and } C_r = 2^{r-1} \text{ if } r \geq 1.$$

and X and Y are two random variables.

Using the above C_r inequality twice, we obtain that

$$\begin{aligned} E_\xi |mA_i|^{1+\epsilon_0} &\leq \left[4\|\mathbf{Q}^{-1}\|(\|\bar{\mathbf{X}}_i\| + \|\bar{\mathbf{x}}_i\|) \right]^{2(1+\epsilon_0)} \left\{ E\|\sqrt{m}\mathbf{d}_m(\theta)\sqrt{m}\hat{\mathbf{g}}^\top(\boldsymbol{\beta})\|^{1+\epsilon_0} + \right. \\ &\quad \left. + E\|\sqrt{m}\mathbf{d}_m(\theta)\sqrt{m}\mathbf{d}_m(\theta)\|^{1+\epsilon_0} \right\} \end{aligned}$$

Note that $\mathbf{Q}$ is a fixed positive-definite matrix. Therefore its norm is finite. It follows from assumption (iv) that $\max_{1 \leq i \leq m} \|\bar{\mathbf{X}}_i\| = O(1)$ and $\max_{1 \leq i \leq m} \|\bar{\mathbf{x}}_i\| = O(1)$. It remains to show that,

$$(1) \quad E_\xi \|\sqrt{m}\mathbf{d}_m(\theta)\sqrt{m}\hat{\mathbf{g}}^\top(\boldsymbol{\beta})\|^{1+\epsilon_0} < \infty,$$

$$(2) \quad E_\xi \|\sqrt{m}\mathbf{d}_m(\theta)\sqrt{m}\mathbf{d}_m^\top(\theta)\|^{1+\epsilon_0} < \infty.$$

Concerning (1), it follows from Hölder's inequality that

$$E_\xi \|\sqrt{m}\mathbf{d}_m(\theta)\sqrt{m}\hat{\mathbf{g}}^\top(\boldsymbol{\beta})\|^{1+\epsilon_0} \leq E_\xi^{1/2} \|\sqrt{m}\mathbf{d}_m(\theta)\|^{2(1+\epsilon_0)} \cdot E_\xi^{1/2} \|\sqrt{m}\hat{\mathbf{g}}(\boldsymbol{\beta})\|^{2(1+\epsilon_0)}.$$

It follows from assumption (iii) with $\epsilon_0 = \delta_0/2$ that

$$E_\xi \|\sqrt{m}\mathbf{d}_m(\theta)\sqrt{m}\hat{\mathbf{g}}^\top(\boldsymbol{\beta})\|^{1+\epsilon_0} = O(1).$$

since

$$\begin{aligned} E_\xi^{1/2} \|\sqrt{m} \mathbf{d}_m(\theta)\|^{2(1+\epsilon_0)} &\leq \\ &\leq E_\xi^{1/2} \left\{ (\|\mathbf{Q}_m\| + \|\mathbf{Q}\|) \|\sqrt{m}(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})\| + \|\sqrt{m} \mathbf{r}_m(\theta)\| \right\}^{2(1+\epsilon_0)}. \end{aligned}$$

and that

$$\|\mathbf{Q}_m\|^2 \leq \frac{1}{n\sigma_e^2} \sum_{i=1}^m \sum_{j=1}^{n_i} \|\mathbf{x}_{ij}\|^2 + \left(\max_{1 \leq i \leq m} \|\bar{\mathbf{x}}_i\| \right)^2 = O(1).$$

Relation (2) can easily be proved by noting that

$$\begin{aligned} E_\xi \|\sqrt{m} \mathbf{d}_m(\theta) \sqrt{m} \mathbf{d}_m^\top(\theta)\|^{1+\epsilon_0} &\leq E_\xi^{1/2} \|\sqrt{m} \mathbf{d}_m(\theta)\|^{2(1+\epsilon)} \times E_\xi^{1/2} \|\sqrt{m} \mathbf{d}_m(\theta)\|^{2(1+\epsilon_0)}. \\ &= E_\xi \|\sqrt{m} \mathbf{d}_m(\theta)\|^{2(1+\epsilon_0)}. \end{aligned}$$

■

References

- [1] Anscombe, F.J. (1960). "Rejection of Outliers", *Technometrics*. **2**, 123-147.
- [2] Arvesen, J.N. (1969). "Jackknifing U -statistics", *Annals of Mathematical Statistics*, **40**, 2076-2100.
- [3] Barnett, V. and Lewis, T. (1984). *Outliers in statistical data*. New-York: Wiley
- [4] Battese, G.E., Harter, R.M. and Fuller, W.A. (1988). "An Error-Component Model for Prediction of County Crop Areas using Survey and Satellite Data. *Journal of the American Statistical Association*, **83**, 28-36
- [5] Beaton, A.E. and Tukey, J.W. (1974). "The Fitting of Power Series. Meaning Polynomials, Illustrated on Band-Spectroscopic Data". *Technometrics*. **16**. 147-185.
- [6] Bhat, B.R. (1981). *Modern Probability Theory*, Wiley Eastern Ltd.
- [7] Bickel, P.J. (1976). "Another Look at Robustness: A Review of Reviews and Some New Developments", *Scandinavian Journal of Statistics*, **3**, 145-168.
- [8] Bickel, P.J., and Lehmann, E.L. (1975). Descriptive statistics for nonparametric models. I. Introduction. II. Location. *Annals of Statistics* **3**, 1038-1044. **3**. 1045-1069. [1.1b]

- [9] Billingsley, P. (1986). *Probability and Measure* (2nd. ed.) New York: Wiley.
 - [10] Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1975). *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, Cambridge, Mass.
 - [11] Boos, D.D., and Serfling, R.J. (1980). A note on differentials and the CLT and LIL for statistical functions, with application to M-estimates. *The Annals of Statistics* 8, 618-624.
 - [12] Campbell, N.A. (1980). Robust procedures in multivariate analysis I: Robust covariance estimation. *Applied Statistics*, 29, 231-237.
 - [13] Carrol, R.J., and Welsh, A.H. (1988). "A note on asymmetry and robustness in linear regression". *Technometrics* 19, 15-18.
 - [14] Cassel, C.M., Särndal, C.E., and Wretman, J.H. (1976). Some results on generalized difference estimation and generalized regression estimation for finite populations. *Biometrika* 63, 615-620.
 - [15] Cassel, C.M., Särndal, C.E., and Wretman, J.H. (1976). *Foundation of Inference in Survey Sampling*. New York: Wiley.
 - [16] Chambers, R.L.(1982), *Robust Finite Population Estimation*. Ph.D. thesis. Johns Hopkins University, Baltimore.
 - [17] Chambers, R.L.(1986), "Outlier Robust Finite Population Estimation", *Journal of the American Statistical Association*, 81, 1063-1069.
 - [18] Chambers, R.L., and Kokic, P.N. (1993), "Outlier Robust Sample Survey Inference", *Proceedings of the 49th Session, International Statistical Institute*.
-

-
- [19] Christensen, R., Pearson, L.M., and Johnson, W. (1992). "Case-Deletion Diagnostics for Mixed Models", *Technometrics*, 43, 38-45.
- [20] Cochran, W.G. (1977). *Sampling Techniques* (3rd. Ed.). New York: Wiley.
- [21] Cook, R.D., and Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman & Hall, London.
- [22] Dalén, J. (1987), Practical Estimators of a Population Total which Reduce the Impact of Large Observations. R & D Report, Statistics Sweden.
- [23] Deville, J.C. and Särndal, C.E. (1992). "Calibration Estimators and Generalized Raking Techniques in Survey Sampling". *Journal of the American Statistical Association*, 87, 376-382.
- [24] Durrett, R. (1991). *Probability: Theory and Examples*, Wadsworth & Brooks/Cole Advanced Books & Software. Pacific Grove, California.
- [25] Efron, B. (1979). "Bootstrap Methods: Another Look at the Jackknife". *Annals of Statistics* 7, 1-26.
- [26] Ernst, L.R. (1980), "Comparison of estimators of the mean which adjust for large observations", *Sankhyā: The Indian Journal of Statistics (Series C)*. 42. 1-16.
- [27] Fay, R.E. and Herriot, R.A. (1979). "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data", *Journal of the American Statistical Association*, 74, 269-277.
- [28] Fellner, W.H. (1986), "Robust Estimation of Variance Components", *Techno-*
-

- metrics*, 28, 51-60.
- [29] Fuller, W.A. (1991). "Simple Estimators of the Mean of Skewed Populations". *Statistica Sinica*, 1, 137-158.
- [30] Fuller, W.A. (1993). "Estimators for Long-Tailed Distribution", *Proceedings of the 49th Session, International Statistical Institute*.
- [31] Ghangurde, P.D. (1989a), Outlier Robust Estimation in Finite Population Sampling. Technical Report, Social Survey Methods Division, Statistics Canada.
- [32] Ghosh, M. and Rao, J.N.K. (1994). "Small Area Estimation: An Appraisal" *Statistical Science*, 9, 55-93.
- [33] Glasser, G.J. (1962). "On the Complete Coverage of Large Units in a Statistical Study", *International Statistical Review*, 30, 28-32.
- [34] Gnedenko, B.V. (1962), *The Theory of Probability*, 4th ed., Chelsea, New York.
- [35] Godambe, V.P. (1982), "Estimation in Survey Sampling", *Journal of the American Statistical Association*, 77, 393-406.
- [36] Gross, W.F., Bode, G., Taylor, J.M. and Lloyd-Smith, C.W. (1986). "Some Finite Population Estimators which Reduce the Contribution of Outliers". *Proceedings of the Pacific Statistical Congress*, Auckland, New Zealand, 20-24 May 1985.
- [37] Gwet, J.P., and Rivest, L.P. (1992), "Outlier Resistant Alternatives to the Ratio Estimator". *Journal of the American Statistical Association*, 87, 1174-1182.
-

-
- [38] Hampel, F.R. (1974). "The influence curve and its role in robust estimation". *Journal of the American Statistical Association*, 69, 383-393.
- [39] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.E.(1986), *Robust Statistics: The Approach Based on Influence Functions*. New York: John Wiley.
- [40] Hansen, M.H., Hurwitz, W.N., and Madow, W.G. (1953), *Sample Survey Methods and Theory*, New York: John Wiley & Sons, Vol. I.
- [41] Henderson, C.R. (1963). "Selection Index and Expected Genetic Advance". *Statistical Genetics and Plant Breeding*, National Research Council Publication 982, National Academy of Sciences, pp. 141-163.
- [42] Hidioglou, M.A., and Srinath, K.P.(1981), "Some Estimators of a Population Total From Simple Random Samples Containing Large Units". *Journal of the American Statistical Association*, 76, 690-695.
- [43] Hidioglou, M.A. and Särndal, C.E. (1985). "An Empirical Study of Some Regression Estimators for Small Domains", *Survey Methodology*, 11, 65-77.
- [44] Hill, R.W. (1977). Robust regression when there are outliers in the carriers. Ph.D. thesis. Harvard university, Cambridge, Mass.[6.3a, 6.4b]
- [45] Holt, D., and Smith, T.M.F.(1979), "Post Stratification", *Journal of the Royall Statistical Society, Series A*, 142, 33-46.
- [46] Huber, P.J.(1964), "Robust Estimation of a Location Parameter", *The Annals of Mathematical Statistics*. 35, 73-101.
-

- [47] Huber, P.J.(1981), *Robust Statistics*, John Wiley & Sons, New York.
 - [48] Huggins, R. M. (1993). "A robust approach to the analysis of repeated measures". *Biometrics* **49**, 715-720.
 - [49] Hulliger B.(1995), "Outlier Robust Horvitz-Thompson Estimators", *Survey Methodology*, 21, 79-87.
 - [50] Isaki, C.T., and Fuller, W.A. (1982). "Survey design under the regression superpopulation model", *Journal of the American Statistical Association* **77**, 89-96.
 - [51] Jones. H.L. (1974), "Jackknife Estimation of Functions of Stratum Means". *Biometrika*, **61**, 343-348.
 - [52] Kish, L. (1965). *Survey Sampling*. New York: Wiley.
 - [53] Kokic, P.N. and Bell, P.A. (1994), "Optimal Winsorizing Cutoffs for a Stratified Finite Population Estimator", *Journal of Official Statistics*, 10, 419-435.
 - [54] Krewski, D. and Rao, J.N.K. (1981). "Inference from Stratified Samples: Properties of the Linearization, Jackknife and Balanced Repeated Replication Methods". *The Annals of Statistics*, 9, 1010-1019.
 - [55] Laird, N.M. and Louis, T.A. (1987). "Empirical Bayes Confidence Intervals Based on Bootstrap Sample", *Journal of the American Statistical Association*, **82**, 739-750.
 - [56] Lahiri, S. N. (1992), "Bootstrapping M -estimators of a Multiple Linear Regression Parameter", *The Annals of Statistics*, **3**, 1548-1570.
-

-
- [57] Lavallée, P. and Hidirolou, M.A. (1988), "On the Stratification of Skewed Populations", *Survey Methodology*, 14, pp. 33-43.
- [58] Lee, H. (1995), "Outliers in Business Surveys", in Cox, Binder, Chinnappa, Christianson, Colledge and Kott (eds.), *Business Surveys Methods*, New York: John Wiley & Sons.
- [59] McCarthy, P.J. (1966). "Replicaton: An Approach to the Analysis of Data from Complex Surveys". *Vital and Health Statistics*, Series 2, No. 14, National Center for Health Statistics, Public Health Service, Washington, D.C.
- [60] Miller, R.G. (1964). "An Unbalanced Jackknife", *Annals of Statistics*, 2, pp. 880-891.
- [61] Ortega, J. and Rheinboldt, W. (1970). *Iterative Solutions of Non-linear Equations in Several Variables*. Page 163. Academic, New York.
- [62] Prasad, N.G.N. and Rao, J.N.K. (1990). "The Estimation of the Mean Squared Error of Small Area Estimators", *Journal of the American Statistical Association*, 85, 163-171.
- [63] Quenouille, M.H. (1949). "Approximate Tests of Correlation in Time Series". *Journal of the Royal Statistical Society, B* 11, 68-84.
- [64] Quenouille, M.H. (1956). "Note on Bias in Estimation". *Biometrika* 43, 353-360.
- [65] Rao, C.R. and Zhao, L.C. (1992). "Linear representation of M-estimates". *The Canadian Journal of Statistics* 20, 359-368.
-

- [66] Rao, J.N.K.(1966), "Alternative Estimators in PPS Sampling for Multiple Characteristics", *Sankhyā: The Indian Journal of Statistics(Series A)*, 28, 47-60
 - [67] Rao, J.N.K. (1978), "Sampling Designs Involving Unequal Probabilities of Selection and Robust Estimation of a Finite Population Total". In *Contributions to Survey Sampling and Applied Statistics*.
 - [68] Rao, J.N.K. and Choudhry, G.H. (1995). "Small Area Estimation: Overview and Empirical Study". in Cox, Binder, Chinnappa, Christianson, Colledge and Kott (eds.), *Business Surveys Methods*. New York: John Wiley & Sons.
 - [69] Reeds, J.A. (1978). "Jackknifing Maximum Likelihood Estimates", *The Annals of Statistics*, 6, 727-739.
 - [70] Rey, W.J.J.(1983), *Introduction to Robust and Quasi-Robust Statistical Methods*. New York: Springer-Verlag.
 - [71] Richardson, A.M. and Welsh, A.H. (1995). "Robust Restricted Maximum Likelihood in Mixed Linear Models". *Biometrics*, 51, 1429-1439.
 - [72] Rivest, L.P.(1989), "De l'unicité des estimateurs robustes en régression lorsque le paramètre d'échelle et le paramètre de régression sont estimés simultanément", *Revue Canadienne de Statistique*, 17, 147-156.
 - [73] Rivest, L.P.(1993). "Winsorization of Survey Data", *Proceedings of the 49th Session, International Statistical Institute*.
 - [74] Rivest, L.P.(1994), "Statistical Properties of Winsorized Means for Skewed
-

- Distributions", *Biometrika*, 81, 373-383.
- [75] Rivest, L.P. and Hurtubise, D. (1995). "On Searls' Winsorized Mean for Skewed Populations", *Survey Methodology*, 21, 107-116.
- [76] Robinson, G.K. (1991). "That BLUP is a Good Thing: the Estimation of Random Effects (with discussuin). *Statistical Science*, 6, 15-51.
- [77] Robinson, P.M. and Särndal, C.E. (1983). "Asymptotic Properties of the Generalized Regression Estimator in Probability Sampling", *Sankhyā: The Indian Journal of Statistics(Series B)*, 45, 240-248
- [78] Royall, R.M. (1970). "On Finite Population Sampling Theory under Certain Linear Regression Models". *Biometrika*, 57, 377-387.
- [79] Särndal, C.E. (1984). "Design-Consistent Versus Model-Dependent Estimators for Small Domains", *Journal of the American Statistical Association*, 79, 624-631.
- [80] Särndal, C.E. and Hidiroglou, M.A. (1989). "Small Domain Estimation: A Conditional Analysis", *Journal of the American Statistical Association*, 84, 266-275.
- [81] Särndal, C.E, Swensson, B. and Wretman, J. (1989). "The Weighted Residual Technique for Estimating the Variance of the General Regression Estimator of the Finite Population Total", *Biometrika*, 76, 527-537.
- [82] Särndal, C.E, Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.
-

- [83] Schaible, W.L., Brock, D.B. and Schnack, G.A. (1977). "An Empirical Comparison of the Simple Inflation, Synthetic and Composite Estimators for Small Area Statistics", *Proceedings of the Social Statistics Section*, pp. 1017-1021. American Statistical Association, Washington, D.C.
 - [84] Scott, A. and Wu, C.F. (1981). "On the Asymptotic Distribution of Ratio and Regression Estimators", *Journal of the American Statistical Association*. 76. 98-102.
 - [85] Searl, S.R. (1971). *Linear Models*, Johns Wiley & Sons.
 - [86] Searls, D.T. (1966), "An Estimator which reduces large true observations". *Journal of the American Statistical Association*, 61, 1200-1204.
 - [87] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. New-York: John Wiley.
 - [88] Shao, J.(1994), "L-Statistics in Complex Survey Problems", *The Annals of Statistics*, 22, 946-967.
 - [89] Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*, Springer.
 - [90] Shao, J. and Wu, C.F.J. (1992). "Asymptotic Properties of the Balanced Repeated Replication Method For Sample Quantiles", *The Annals of Statistics*. 20, 1571-1593.
 - [91] Smith, T.M.F. (1987) "Influential Observations in Survey Sampling", *Journal of Applied Statistics*, 14, 143-152
 - [92] Stukel. D. (1991) *Small Area Estimation Under One and Two-Fold Nested*
-

Error Regression Model, Ph.D. Thesis, Carleton University.

- [93] Tukey, J.W. (1958). "Bias and Confidence in Not Quite Large Samples". *Annals of Mathematical Statistics* **29**, 614.
 - [94] Wright, R.L. (1983). "Finite population sampling with multivariate auxiliary information". *Journal of the American Statistical Association* **78**, 879-884.
-