DISSERTATIONES MATHEMATICAE UNIVERSITATIS TARTUENSIS
76

DISSERTATIONES MATHEMATICAE UNIVERSITATIS TARTUENSIS
76

NATALJA LEPIK

Estimation of domains under restrictions built upon generalized regression and synthetic estimators

TARTU UNIVERSITY PRESS

Faculty of Mathematics and Computer Science, University of Tartu, Tartu, Estonia

Dissertation has been accepted for the commencement of the degree of Doctor of Philosophy (Ph.D.) in mathematical statistics on August 30, 2011, by the Council of the Faculty of Mathematics and Computer Science, University of Tartu.

Supervisor:	
Associate Professor, Cand. Sc.	Imbi Traat University of Tartu Tartu, Estonia
Opponents:	
Professor, Ph.D.	Risto Lehtonen University of Helsinki Helsinki, Finland
Head of Methodology Dept., Ph.D.	Kaja Sõstra Statistics Estonia Tallinn, Estonia

Commencement will take place on August 30, 2011, at 14.00 in Liivi 2-122.

Publication of this dissertation is financed by the Institute of Mathematical Statistics of the University of Tartu (the Estonian Targeted Financing Project SF0182724s06 and the Estonian Science Foundation grant 8789).

ISSN 1024-4212
ISBN 978-9949-19-754-5 (trükis)
ISBN 978-9949-19-755-2 (PDF)

Tartu Ülikooli Kirjastus
www.tyk.ee
Tellimus nr. 409

To my family.

Contents

List of original publications

Paper in a refereed journal (indexed by the Thomson Reuters Web of Science)

1. Lepik, N. (2007). On the Bias of the Generalized Regression Estimator in Survey Sampling. *Acta Applicandae Mathematicae*, 97(1-3), 41 - 52.

Other publications

2. Jurevitš (Lepik), N. (2004). *Rakendusstatistika. (Applied statistics)*, in Estonian. Tallinn: Ilo
3. Lepik, N. (2008). Estimation under restrictions with biased initial estimators. *Baltic-Nordic Workshop on Survey Sampling Theory and Methodology; Kuressaare.* Tallinn: Statistics Estonia, 2008, 112 - 116.
4. Lepik, N. (2009). Synthetic estimator for domains. *Proceedings of the Baltic-Nordic-Ukrainian Summer School on Survey Statistics.* Kyiv: "TBiMC" Scientific Publishers, 2009, 91 - 97.

The papers 1, 3, 4 are related to the topic of the Thesis.

Acknowledgements

First and foremost I offer my sincerest gratitude to my supervisor, docent Imbi Traat, for the patient guidance, encouragement and advice she has provided throughout my PhD studies. I have been extremely lucky to have a supervisor who cared so much about my work, and who responded to all my questions and queries so promptly.

I would like to express my greatest gratitude to my family for their patience and forbearance while I have spent so many hours studying and working on Thesis.

Finally, I would like to thank all my colleagues in the Institute of Mathematical Statistics of the Tartu University who have assisted me one way or another, especially with challenging alternative views. Special thanks go to my friend and colleague Natalja Eigo from the National Institute for Health Development for conversations that clarified my too theoretical thinking and connected me with the real world.

Part of the work was carried out during the author's visit to the Department of Mathematics and Mathematical Statistics, Umea University, where I would like to thank Professor Gunnar Kulldorff for his support and helpful discussions.

This research was partly supported by the Estonian Science Foundation.

Introduction

Nowadays, demand on accurate statistics of population sub-groups or domains increases. This statistics can be obtained from surveys, or, sometimes, aggregated from registers. It may happen that even if the register contains variables under interest, it does not contain identifies of the domains under our particular interest. As follows, these domain totals can not be produced from that register, they need to be estimated from a survey. The survey has to collect information on the same study variable but together with domain identifiers. As a result, the consistency problem occurs, the domain estimates from the survey do not sum up to the totals available from the registers. Analogical problem arises in the multi-survey situation, where some study variables are common in two or more surveys. Domain estimates from one survey do not sum up to the estimates of larger domains (or population totals) from another survey. Yet, there is one more situation where the consistency problem occurs. Domains themselves and the population total may be estimated by conceptually different estimators in the same survey. As a result, the domain totals do not sum up to the population total, or to the relevant larger domains.

The described inconsistency is annoying from the statistics users viewpoint. Statisticians know that the relationships between population parameters do not necessarily hold for the estimates in a sample. They also know that any auxiliary information incorporated into estimators may increase precision of these estimators. In our situation known relationships between population parameters is a kind of the auxiliary information. If one could use this information in the estimators, one were able to make estimators more accurate and force them to satisfy desired restrictions. Elaboration and study of consistent domain estimators that are more accurate than the initial estimators is general topic of the current thesis.

The problem is not new, consistency of estimators has been considered for some time. For example, if consistency is required between two surveys or between a survey and a register, some authors (Zieschang 1990, Renssen and Nieuwenbroek 1997, Traat and Särndal 2009, Dever and Valliant 2010) have proposed classical calibration approach as a solution. In this approach, the common variables are considered as additional auxiliary variables, and consistency requirement is presented in terms of calibration constraints. Other authors (Kroese and Renssen 1999, Knottnerus and Van Duin 2006) use different calibration approach for this situation, called repeated weighting. They re-calibrate the initially calibrated estimators to satisfy the consistency constraints with outside information.

Yet another approach is proposed by Knottnerus (2003). His general restriction (GR) estimator is constructed upon unbiased initial estimators so that the result satisfies desired linear restrictions. The GR estimator has many good properties like unbiasedness and higher precision compared to that of the initial estimator. Under certain assumptions the GR estimator is optimal in a class of estimators satisfying given restrictions. The ideas of the GR estimator were extended for consistent domain estimation by Sõstra (2007) and further elaborated in Sõstra and Traat (2009).

The Knottnerus approach is able to obtain consistency in both situations: 1) consistency between estimators from different data sources, 2) consistency between estimators in the same data source. In this thesis Knottnerus approach is taken as a basis for consistent estimation. This choice was motivated by the optimality properties and by our focus on the situation 2) in domain estimation.

In addition, a generalization had to be made. All earlier works have concentrated on the unbiased estimation. Similarly, Knottnerus (2003) and Sõstra (2009) assume unbiased initial estimators. In domains case, however, also biased estimators become useful. For example the model-based estimators (Rao, 2003), the synthetic (or projection) estimator (Särndal et al., 1992, pp. 408-412, Yung and Rubin-Bleuer, 2007) are often used for small domains. Though potentially biased, they are appealing due to their small variability. However, it is not known so far, how the biased initial estimators will affect the final GR estimator, after the consistency restrictions are put on. The properties of these final estimators are unknown.

In this thesis we concentrate on the estimation of the domain and the population totals under summation restriction. We allow biased as well unbiased initial estimators for domains. The classes of initial estimators for the GR estimator were chosen to be the generalized regression (GREG) family, and the family of synthetic (SYN) estimators.

The GREG estimator is well studied and widely used by statistical agencies. By using auxiliary information this model-assisted estimator is more precise than the estimator without auxiliary information. The GREG estimator is nearly unbiased (Särndal et al., 1992, p. 237), but at the domain level, and especially in small domains, it may have large variability, (Lehtonen and Pahkinen, 2004, p. 196).

The synthetic estimator is much less studied than GREG, especially in general level. The mutual relationships between the SYN and the GREG are not systematically studied either. The SYN estimator is model-based. Study variable values are predicted by auxiliary variables. Once having predictions, it is very convenient to compute domain or population totals just by summing the predicted values in the respective domains or in the population. The estimates can be produced even for the domains with no sampled units, given the auxiliary variable values for each unit. Due to this convenience, the synthetic estimator is often used in practice (Yung and Rubin-Bleuer, 2007). As an additional positive feature, it has small variability. Its negative feature is potential bias, especially if the underlying model is misspecified.

We assume that both the GREG and the SYN estimators may be simultaneously used in the domains under study, and in the population. It is known that both estimators the GREG and the SYN for domains can be constructed under different model specifications (Lehtonen and Pahkinen 2004, pp. 187-213) – under the P-model (population level model) and under the D-model (domain level model). The estimators under D-models are called the direct estimators, because they use the study variable values only from a given domain. The estimators under P-models uses values of the study variable also from other domains and therefore are called indirect estimators. This thesis concentrates on the properties of the SYN-D and the SYN-P estimators. We are especially interested in the respective biases and in the the mean square errors. Their expressions are needed in our restriction estimators.

The goals of the current thesis are:

- Construction of the new domain estimators (the GR estimators) that satisfy summation restriction. As a generalization, the biased initial estimators are allowed in this construction.

- Working out properties of the proposed GR estimators; the bias, the mean square error (MSE) matrix. Based on these properties, showing superiority of the new GR estimators over initial estimators. Establishing the MSE-matrix ordering of the estimators (in Löwner sense).
- Studying properties of the GREG and the SYN domain estimators, that were chosen as the building blocks for the GR estimator.
- Deriving their linearization-based biases and the MSE expressions, for the population and for the domains case as well. In the domains case considering both the D- and the P-models.
- Deriving cross-MSE's for different estimators.
- Studying a sufficient condition for the equality of the GREG and the SYN estimators, both in the population and in the domain level.
- Developing all properties of the estimators in general level, valid for all sampling designs. Applying them for the simple random sampling without replacement and for the multinomial designs.
- Illustrating and confirming results in a simulation study on real data.

Domain estimation is a multivariate problem, since ordinarily, there are many domains under estimation. This multivariate feature calls for bringing up matrix technique which is used throughout the thesis. The accuracy of the estimators is also measured with multivariate tools – with the MSE matrix. Similarly to the univariate notion, it allows to compare the accuracy of both the unbiased and biased multivariate estimators.

We use the design based approach, i. e. the properties of the estimators (such as expectation, variance/covariance and the MSE) are determined by the sampling design and by the study variable values in the finite population. Sampling design is considered as a multivariate distribution, and a sample as an outcome from that distribution, a realization of the sampling vector. We assume that sample sizes in domains are not too small, i. e. we do not consider small area estimation methods.

Our study method is mathematical. The statements are formulated as propositions, and proved with mathematical methods. Notions and results from probability theory, mathematical statistics and survey sampling theory are used.

This thesis is organized in the following way.

Chapter 1 gives the necessary matrix tools and properties that are used in the further chapters. Some of these tools (e. g. projectors, matrix differentiation and some others) are not so common, especially in survey sampling field. The basics of the design-based inference, with emphasis on the multivariate notions, are presented. Among others, the covariance and the MSE matrices of the estimators are defined, and their properties given.

In Chapter 2 we give the main results of this thesis on the domain estimation under restrictions, where biased initial estimators are allowed. We first introduce the Knottnerus (2003) GR estimator, and give an example where this estimator, if applied to the biased initial estimators, becomes biased itself, and is not better than the initial estimator. In Propositions 2.2 - 2.4 we propose three new GR estimators with the MSE smaller than that of the initial estimators. In Proposition 2.5 we also order MSEs of the three GR estimators and the initial estimator, and find out the best one.

The attention of Chapter 3 is turned to the classes of initial estimators. Two classes are described thoroughly here: the GREG and the SYN estimators, both for the population and for the domain estimation. We describe situations and the condition, when these two estimators are equal. The Taylor expansions of these estimators are derived (up to the second order terms for the GREG). Based on these, we give the approximate biases, covariances and the MSEs of the estimators. The properties leading to the equality of GREG and SYN estimators are also considered. In the domains case we consider two underlying models for the GREG and SYN estimators, the P- and the D-models. We study properties of these estimators and give general expressions for the approximate MSEs between the domain and the population GREG and SYN estimators (they are needed in our GR estimators). Under some conditions these MSEs simplify which is also shown in this chapter.

Chapter 4 presents simulation results. The real data of the healthcare personnel of Estonia is used. The aim is to illustrate the performance of the three GR estimators in the practical situation. The population of about 22000 persons is divided into four domains of different sizes and the samples of size 400 are drawn by the SI and the MN designs. Two study variables are considered, the continuous and the binary variables. The inconsistency problem of the initial domain estimators is illustrated by tables and figures. It is also shown that all three GR estimators satisfy restrictions. The MSE matrices of the GR and the initial estimators are computed from the theoretical formulas, and found also empirically. It is demonstrated that the MSEs of the GR estimators are smaller than that of the initial estimator. It is also demonstrated that the MSE-inequalities of the GR estimators, proved in the theoretical part, hold. The theoretical MSEs were linearization-based approximate MSEs. Nevertheless, empirical MSEs were very close to them. That holds both for the initial estimators and for the GR estimators.

Chapter 5 summarizes main results of the present thesis.

Chapter 1

Preliminaries

In this chapter we give basic definitions and results from matrix algebra that are needed for the following chapters. The books by Harville (1997), Kollo and von Rosen (2005), Puntanen and Styan (2004) are mostly used. We also introduce the design-based inference of sampling theory that is basis for the probabilistic properties of the estimators derived in this thesis (Särndal et al., 1992, and in the language of sampling vectors, Traat, 2000). We add a multivariate perspective to these notions.

1.1 Tools of matrix algebra

1.1.1 Basic notions and properties

Let $\mathbf{A} = (a_{ij}) : n \times m$ be a *real matrix* with elements a_{ij}, where $i = 1, 2, ..., n$ and $j = 1, 2, ..., m$ ($\mathbf{A} \in \mathbb{R}^{n \times m}$). If $n = m$, the matrix is a *square* matrix. The notation $\boldsymbol{a} = (a_i)$ is used for a column vector with elements a_i.

The $n \times m$ matrix with all elements equal to 1 is denoted by $\mathbf{1}_{n \times m}$. A vector of n ones is written as $\mathbf{1}_n$ (or just $\mathbf{1}$). Analogically, $\mathbf{0}_{n \times m}$ denotes a matrix where all elements are zero.

Three different matrix products are used in this thesis. The ordinary product of $\mathbf{A}$ and $\mathbf{B} : m \times l$ is

$$\mathbf{AB} = \mathbf{C} = (c_{ij}) : n \times l,$$

where $c_{ij} = \sum_{k=1}^{m} a_{ik} b_{kl}$. The elementwise *Hadamard* or *Schur* product with a matrix $\mathbf{B} = (b_{ij}) : n \times m$ is

$$\mathbf{A} \circ \mathbf{B} = (a_{ij} b_{ij}) : n \times m. \tag{1.1}$$

The *Kronecker* product with a matrix $\mathbf{B} : k \times l$ is a block matrix, where (i, j)-block is $a_{ij}\mathbf{B}$

$$\mathbf{A} \otimes \mathbf{B} = (a_{ij}\mathbf{B}) : nk \times ml. \tag{1.2}$$

For any matrix $\mathbf{A}$ the *Kroneckerian power* is given by

$$\mathbf{A}^{\otimes k} = \underbrace{\mathbf{A} \otimes ... \otimes \mathbf{A}}_{k \text{ times}}.$$

The *transpose* of $\mathbf{A} = (a_{ij})$ is $\mathbf{A}' = (a_{ji}) : m \times n$. A square matrix $\mathbf{A}$ is *symmetric*, if $\mathbf{A} = \mathbf{A}'$. According to Harville (1997, p.52), it holds for any $\mathbf{A} : n \times m$ and $\mathbf{B}$, $\mathbf{C} : m \times p$,

$$\mathbf{AB} = \mathbf{AC} \text{ if and only if } \mathbf{A}'\mathbf{AB} = \mathbf{A}'\mathbf{AC}. \tag{1.3}$$

For the matrix product the following properties hold:

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}', \tag{1.4}$$
$$(\mathbf{A} \circ \mathbf{B})' = \mathbf{A}' \circ \mathbf{B}', \tag{1.5}$$
$$(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}', \tag{1.6}$$
$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}, \tag{1.7}$$
$$(\mathbf{A} + \mathbf{B}) \otimes (\mathbf{C} + \mathbf{D}) = \mathbf{A} \otimes \mathbf{C} + \mathbf{A} \otimes \mathbf{D} + \mathbf{B} \otimes \mathbf{D} + \mathbf{B} \otimes \mathbf{D}, \tag{1.8}$$

where $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ are of the appropriate dimensions.

For any two matrices $\mathbf{A}, \mathbf{B} : n \times m$,

$$\mathbf{A} = \mathbf{B} \text{ if and only if } \mathbf{A}\boldsymbol{x} = \mathbf{B}\boldsymbol{x}, \tag{1.9}$$

for any vector $\boldsymbol{x} : m \times 1$ (Harville, 1997, Lemma 2.3.2).

The *diagonalization* of $\mathbf{A} : n \times n$ produces a diagonal matrix diag($\mathbf{A}$): $n \times n$, with a_{ii} on the main diagonal. Similarly, diag($\boldsymbol{a}$): $n \times n$ denotes the diagonal matrix having the vector $\boldsymbol{a} = (a_1, a_2, ..., a_n)'$ on the main diagonal.

The $n \times n$ *identity* matrix is denoted by $\mathbb{I}_n$. The identity matrix satisfies

$$\mathbb{I}_n \mathbf{A} = \mathbf{A} \mathbb{I}_m = \mathbf{A}.$$

A square matrix $\mathbf{A}$ is *idempotent* if

$$\mathbf{A}^2 = \mathbf{A}\mathbf{A} = \mathbf{A}.$$

Let $|\mathbf{A}|$ denote a *determinant* of $\mathbf{A} : n \times n$. If $|\mathbf{A}| \neq 0$, then matrix $\mathbf{A}$ is called *non-singular* and then a unique *inverse* of $\mathbf{A}$ exists (Harville, 1997, p.178). The inverse of $\mathbf{A}$ is denoted by $\mathbf{A}^{-1}$ and it is defined by the equality:

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbb{I}_n.$$

Any matrix $\mathbf{B} : m \times n$ that satisfies

$$\mathbf{A} = \mathbf{A}\mathbf{B}\mathbf{A} \tag{1.10}$$

is called a *generalized inverse* of the matrix $\mathbf{A}$ and it is denoted by $\mathbf{A}^{-} = \mathbf{B}$.

For a non-singular matrix $\mathbf{A}$, it is clear that

$$\mathbf{A}\mathbf{A}^{-1}\mathbf{A} = \mathbf{A},$$

and if $\mathbf{A}^{-}$ is the generalized inverse of $\mathbf{A}$, then

$$\mathbf{A}^{-} = (\mathbf{A}^{-1}\mathbf{A})\mathbf{A}^{-}(\mathbf{A}\mathbf{A}^{-1}) = \mathbf{A}^{-1}\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}.$$

A symmetric matrix is *non-negative (positive) definite* if $\boldsymbol{x}'\mathbf{A}\boldsymbol{x} \geq 0$ ($\boldsymbol{x}'\mathbf{A}\boldsymbol{x} > 0$) for any vector $\boldsymbol{x} \neq \mathbf{0}$. It appears that for any matrix $\mathbf{A} : n \times m$ the product $\mathbf{A}\mathbf{A}'$ is non-negative definite (Kollo and von Rosen, 2005, p.12).

A matrix $\mathbf{A} : n \times n$ is a symmetric positive definite matrix if and only if (Harville, 1997, p.219) there exists a non-singular matrix $\mathbf{B}$ such that

$$\mathbf{A} = \mathbf{B}'\mathbf{B}. \tag{1.11}$$

Any positive definite matrix is non-singular (Harville, 1997, p.213).

If for the square matrices $\mathbf{A}$ and $\mathbf{B}$ it holds,

$$\boldsymbol{x}'(\mathbf{A} - \mathbf{B})\boldsymbol{x} > 0 \text{ for any } \boldsymbol{x} \neq \mathbf{0}, \tag{1.12}$$

then $\mathbf{A} \geq \mathbf{B}$ in the sense of *Löwner ordering* (Rao and Rao, 1998, p.508). If $\mathbf{A} \geq \mathbf{B}$, then $a_{ii} \geq b_{ii}$, $\forall i$.

For any matrix $\mathbf{A} = (\boldsymbol{a}_1|\boldsymbol{a}_2|...|\boldsymbol{a}_m) : n \times m$, where $\boldsymbol{a}_i$ is a column vector of $\mathbf{A}$, its *vectorized form* is defined as

$$\text{vec}(\mathbf{A}) = \begin{pmatrix} \boldsymbol{a}_1 \\ \vdots \\ \boldsymbol{a}_m \end{pmatrix} : nm \times 1. \tag{1.13}$$

Let us have $\mathbf{A} : n \times m, \mathbf{B} : m \times k, \mathbf{C} : k \times l$ and $\mathbf{D} : n \times m$. Then the following properties of the vec-operator hold (Lütkepohl,1996, pp.20-21):

$$\text{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\text{vec}\mathbf{B}; \tag{1.14}$$
$$\text{vec}(\mathbf{A} \circ \mathbf{D}) = \text{diag}(\text{vec}\mathbf{A})\text{vec}\mathbf{D} = \text{diag}(\text{vec}\mathbf{D})\text{vec}\mathbf{A}. \tag{1.15}$$

For vectors $\boldsymbol{a}$ and $\boldsymbol{b}$

$$\text{vec}(\boldsymbol{a}\boldsymbol{b}') = \boldsymbol{b} \otimes \boldsymbol{a}. \tag{1.16}$$

The $pq \times pq$ matrix $\mathbf{K}_{p,q}$, consisting of $q \times p$ blocks is called *commutation matrix*, if in the (i, j)-th block the (j, i)-th element equals to one, while all other elements in that block are zeros (Kollo, van Rosen, 2005, p. 79-82). Main properties of this matrix are

$$\mathbf{K}_{p,q} = \mathbf{K}'_{q,p}; \tag{1.17}$$
$$\mathbf{K}_{p,q}\mathbf{K}_{q,p} = \mathbb{I}_{pq}; \tag{1.18}$$
$$\mathbf{K}_{p,1} = \mathbf{K}_{1,p} = \mathbb{I}_p; \tag{1.19}$$
$$\text{vec}\mathbf{A}' = \mathbf{K}_{p,q}\text{vec}\mathbf{A} \text{ for any } \mathbf{A} : p \times q, \tag{1.20}$$
$$\mathbf{A} \otimes \mathbf{B} = \mathbf{K}_{p,r}(\mathbf{B} \otimes \mathbf{A})\mathbf{K}_{s,q} \text{ for } \mathbf{B} : r \times s. \tag{1.21}$$

1.1.2 Linear spaces and projectors

The *column space* $\mathcal{C}(\mathbf{A})$ of an $n \times m$ matrix $\mathbf{A}$ is the set of all n-dimensional vectors generated by the columns of $\mathbf{A}$:

$$\mathcal{C}(\mathbf{A}) = \{\boldsymbol{y} : \boldsymbol{y} = \mathbf{A}\boldsymbol{x}, \boldsymbol{x} \in \mathbb{R}^m\}. \tag{1.22}$$

The *null space* $\mathcal{N}(\mathbf{A})$ (or *nullity*) of $\mathbf{A}$ is defined by

$$\mathcal{N}(\mathbf{A}) = \{\boldsymbol{x} \in \mathbb{R}^m : \mathbf{A}\boldsymbol{x} = \mathbf{0}\}. \tag{1.23}$$

Many properties of column spaces are given in Kollo and von Rosen(2005, pp. 48-49). Here we bring the necessary.

The relation $\mathcal{C}(\mathbf{A}) \subset \mathcal{C}(\mathbf{B})$ means that every vector in $\mathcal{C}(\mathbf{A})$ belongs to $\mathcal{C}(\mathbf{B})$ but not vice versa.

Proposition 1.1 For any $n \times m$ matrix $\mathbf{A}$ and $n \times p$ matrix $\mathbf{B}$, $\mathcal{C}(\mathbf{A}) \subset \mathcal{C}(\mathbf{B})$ if and only if there exists an $m \times p$ matrix $\mathbf{M}$ such $\mathbf{A} = \mathbf{B}\mathbf{M}$.

Definition 1.1 Let $\mathbf{A}$ be an $n \times m$ matrix. A square matrix $\mathbf{P} : n \times n$ is called a *projector matrix* (or simply *a projector*) *onto column space* of $\mathbf{A}$ if for an arbitrary vector $\boldsymbol{v} \in \mathbb{R}^n$,

$$\mathbf{P}\boldsymbol{v} \in \mathcal{C}(\mathbf{A}), \tag{1.24}$$

and for any vector $\boldsymbol{u} \in \mathcal{C}(\mathbf{A})$,

$$\mathbf{P}\boldsymbol{u} = \boldsymbol{u}. \tag{1.25}$$

A projector $\mathbf{P}$, if a symmetric matrix, is called an *orthogonal projector.*

Harville (1997, p. 166) has shown that the matrix $\mathbf{P}_{\mathbf{A}} : n \times n$ is an orthogonal projector onto $\mathcal{C}(\mathbf{A})$ if and only if

$$\mathbf{P}_{\mathbf{A}} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'. \tag{1.26}$$

Definition 1.2 For any matrix $\mathbf{A} : n \times m$ and for the positive definite symmetric matrix $\mathbf{V} : n \times n$, the symbol $\mathbf{P}_{\mathbf{A};\mathbf{V}}$ is used for the following $n \times n$ matrix:

$$\mathbf{P}_{\mathbf{A};\mathbf{V}} = \mathbf{A}(\mathbf{A}'\mathbf{V}\mathbf{A})^{-}\mathbf{A}'\mathbf{V}. \tag{1.27}$$

If $\mathbf{V} = \mathbb{I}$ in (1.27) then we get the orthogonal projector $\mathbf{P}_{\mathbf{A}}$ in (1.26).

In some literature (e.g. Puntanen and Styan, 2004; Harville, 1997) the matrix $\mathbf{P}_{\mathbf{A};\mathbf{V}}$ is called *the orthogonal projector onto* $\mathcal{C}(\mathbf{A})$ *with respect to* $\mathbf{V}$ (or just *onto* $\mathcal{C}_{\mathbf{V}}(\mathbf{A})$).

Many properties of the matrix $\mathbf{P}_{\mathbf{A};\mathbf{V}}$ are given in Harville (1997, pp. 261-263). The most important of them are given below, they follow from the definition of $\mathbf{P}_{\mathbf{A};\mathbf{V}}$.

Proposition 1.2 Let $\mathbf{A}$ be any $n \times m$ matrix, and $\mathbf{V} : n \times n$ symmetric positive definite matrix. Then,

$$\mathbf{P}_{\mathbf{A};\mathbf{V}}\mathbf{A} = \mathbf{A}, \text{ imlying that } \mathbf{V}\mathbf{P}_{\mathbf{A};\mathbf{V}}\mathbf{A} = \mathbf{V}\mathbf{A}; \tag{1.28}$$
$$\mathbf{V}\mathbf{P}_{\mathbf{A};\mathbf{V}} = (\mathbf{V}\mathbf{P}_{\mathbf{A};\mathbf{V}})', \text{ so } \mathbf{V}\mathbf{P}_{\mathbf{A};\mathbf{V}} \text{ is symmetric;} \tag{1.29}$$
$$\mathbf{V}(\mathbb{I} - \mathbf{P}_{\mathbf{A};\mathbf{V}}) = \big(\mathbf{V}(\mathbb{I} - \mathbf{P}_{\mathbf{A};\mathbf{V}})\big)'; \tag{1.30}$$
$$\mathbf{P}_{\mathbf{A};\mathbf{V}}\mathbf{V}^{-1} = (\mathbf{P}_{\mathbf{A};\mathbf{V}}\mathbf{V}^{-1})'; \tag{1.31}$$
$$\mathbf{P}'_{\mathbf{A};\mathbf{V}}\mathbf{V}\mathbf{A} = \mathbf{V}\mathbf{A}; \tag{1.32}$$
$$\mathbf{P}'_{\mathbf{A};\mathbf{V}}\mathbf{V}\mathbf{P}_{\mathbf{A};\mathbf{V}} = \mathbf{V}\mathbf{P}_{\mathbf{A};\mathbf{V}}; \tag{1.33}$$
$$(\mathbb{I} - \mathbf{P}_{\mathbf{A};\mathbf{V}})'\mathbf{V}(\mathbb{I} - \mathbf{P}_{\mathbf{A};\mathbf{V}}) = \mathbf{V}(\mathbb{I} - \mathbf{P}_{\mathbf{A};\mathbf{V}}); \tag{1.34}$$
$$\mathbf{P}^2_{\mathbf{A};\mathbf{V}} = \mathbf{P}_{\mathbf{A};\mathbf{V}}, \text{that is } \mathbf{P}_{\mathbf{A};\mathbf{V}} \text{ is idempotent;} \tag{1.35}$$
$$\mathbf{P}_{\mathbf{A};\mathbf{V}}\mathbf{B} = \mathbf{B} \text{ for any } \mathbf{B} : n \times l \text{ such that } \mathcal{C}(\mathbf{B}) \subset \mathcal{C}(\mathbf{A}). \tag{1.36}$$

1.1.3 Matrix differentiation and Taylor expansion

Assume, that the matrix $\mathbf{X} : p \times q$ is *mathematically independent and variable* (m.i.v.). It means that all elements of $\mathbf{X}$ are non-constant; no two or more elements are functionally dependent.

Definition 1.3 (Kollo and von Rosen, 2005, p.127) Let the elements of $\mathbf{Y} : r \times s$ be functions of $\mathbf{X} : p \times q$. The matrix $\frac{d\mathbf{Y}}{d\mathbf{X}} : pq \times rs$ is called *matrix derivative* of $\mathbf{Y}$ by $\mathbf{X}$ in a set $\mathcal{A}$, if the partial derivatives $\partial y_{kl}/\partial x_{ij}$ exist, are continuous in $\mathcal{A}$, and their location in the matrix is specified by

$$\frac{d\mathbf{Y}}{d\mathbf{X}} = \frac{\partial}{\partial \text{vec}\mathbf{X}} \text{vec}'\mathbf{Y}, \tag{1.37}$$

where

$$\frac{\partial}{\partial \text{vec}\mathbf{X}} = \left(\frac{\partial}{\partial x_{11}}, \ldots, \frac{\partial}{\partial x_{p1}}, \frac{\partial}{\partial x_{12}}, \ldots, \frac{\partial}{\partial x_{p2}}, \ldots, \frac{\partial}{\partial x_{1q}}, \ldots, \frac{\partial}{\partial x_{pq}} \right)'. \tag{1.38}$$

Some basic properties of the matrix derivative, needed in this work, are listed below (Kollo and von Rosen, 2005, p.149).

Proposition 1.3 Let $\mathbf{X} : p \times q$, m.i.v., $\mathbf{Y} : r \times s$, $\mathbf{Z} : m \times n$ and $\mathbf{A}$, $\mathbf{B}$ be the constant matrices of the proper size. Then

$$\frac{d(c\mathbf{X})}{d\mathbf{X}} = c\mathbb{I}_{pq}, \text{ where } c \text{ is a constant;} \tag{1.39}$$

$$\frac{d(\mathbf{A}'\mathbf{X})}{d\mathbf{X}} = \mathbb{I}_q \otimes \mathbf{A}; \tag{1.40}$$

$$\frac{d(\mathbf{A}'\mathrm{vec}\mathbf{X})}{d\mathbf{X}} = \mathbf{A}; \tag{1.41}$$

$$\frac{d(\mathbf{Y}+\mathbf{Z})}{d\mathbf{X}} = \frac{d\mathbf{Y}}{d\mathbf{X}} + \frac{d\mathbf{Z}}{d\mathbf{X}}, \text{ where } \mathbf{Z} : r \times s; \tag{1.42}$$

$$\frac{d(\mathbf{AYB})}{d\mathbf{X}} = \frac{d\mathbf{Y}}{d\mathbf{X}}(\mathbf{B} \otimes \mathbf{A}'); \tag{1.43}$$

$$\frac{d\mathbf{X}^{-1}}{d\mathbf{X}} = -\mathbf{X}^{-1} \otimes (\mathbf{X}')^{-1}, \text{ if } \mathbf{X} \text{ is non-singular;} \tag{1.44}$$

$$\text{if } \mathbf{Z} = \mathbf{Z}(\mathbf{Y}) \text{ and } \mathbf{Y} = \mathbf{Y}(\mathbf{X}), \text{ then } \frac{d\mathbf{Z}}{d\mathbf{X}} = \frac{d\mathbf{Y}}{d\mathbf{X}}\frac{d\mathbf{Z}}{d\mathbf{Y}}; \tag{1.45}$$

if $\mathbf{W} = \mathbf{W}(\mathbf{Y}(\mathbf{X}), \mathbf{Z}(\mathbf{X}))$, then

$$\frac{d\mathbf{W}}{d\mathbf{X}} = \left.\frac{d\mathbf{Y}}{d\mathbf{X}}\frac{d\mathbf{W}}{d\mathbf{Y}}\right|_{\mathbf{Z}=const} + \left.\frac{d\mathbf{Z}}{d\mathbf{X}}\frac{d\mathbf{W}}{d\mathbf{Z}}\right|_{\mathbf{Y}=const} \tag{1.46}$$

$$\frac{d(\mathbf{X}')}{d\mathbf{X}} = \mathbf{K}_{q,p}, \tag{1.47}$$

$$\frac{d(\mathbf{Y} \otimes \mathbf{Z})}{d\mathbf{X}} = \left\{\frac{d\mathbf{Y}}{d\mathbf{X}} \otimes \mathrm{vec}'\mathbf{Z} + \mathrm{vec}'\mathbf{Y} \otimes \frac{d\mathbf{Z}}{d\mathbf{X}}\right\}(\mathbb{I}_s \otimes \mathbf{K}_{r,n} \otimes \mathbb{I}_m), \tag{1.48}$$

where $\mathbf{K}_{q,p}$ is the commutation matrix.

In this thesis we need the multivariate Taylor series expansion formulated in the next proposition.

Proposition 1.4 (Kollo and von Rosen, 2005, p. 151) If the function $\boldsymbol{f}(\boldsymbol{x})$ from $\mathbb{R}^p$ to $\mathbb{R}^q$ has continuous partial derivatives up to the order $(n+1)$ in a neighborhood $\mathcal{D}$ of a point $\boldsymbol{x}_0$, then the function $\boldsymbol{f}(\boldsymbol{x})$ can be expanded into the Taylor series at the point $\boldsymbol{x}_0$ in the following way:

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x}_0) + \sum_{k=1}^{n} \frac{1}{k!}\left(\mathbb{I}_q \otimes (\boldsymbol{x}-\boldsymbol{x}_0)^{\otimes(k-1)}\right)' \left.\left(\frac{d^k\boldsymbol{f}(\boldsymbol{x})}{d\boldsymbol{x}^k}\right)'\right|_{\boldsymbol{x}=\boldsymbol{x}_0} (\boldsymbol{x}-\boldsymbol{x}_0) + \boldsymbol{r}_n, \tag{1.49}$$

where

$$\boldsymbol{r}_n = \frac{1}{(n+1)!}\left(\mathbb{I}_q \otimes (\boldsymbol{x}-\boldsymbol{x}_0)^{\otimes(k-1)}\right)'\left(\frac{d^{n+1}\boldsymbol{f}(\boldsymbol{x})}{d\boldsymbol{x}^{n+1}}\right)'\Bigg|_{\boldsymbol{x}=\boldsymbol{\xi}}(\boldsymbol{x}-\boldsymbol{x}_0), \quad (1.50)$$

for some $\xi \in \mathcal{D}$.

In the special case of $n = 2$ and $q = 1$, the Taylor expansion up to the second order term follows from (1.49),

$$\begin{aligned} \boldsymbol{f}(\boldsymbol{x}) &= \boldsymbol{f}(\boldsymbol{x}_0) + \left(\frac{d\boldsymbol{f}(\boldsymbol{x})}{d\boldsymbol{x}}\right)'\Bigg|_{\boldsymbol{x}=\boldsymbol{x}_0}(\boldsymbol{x}-\boldsymbol{x}_0) \\ &+ \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_0)'\left(\frac{d^2\boldsymbol{f}(\boldsymbol{x})}{d\boldsymbol{x}^2}\right)'\Bigg|_{\boldsymbol{x}=\boldsymbol{x}_0}(\boldsymbol{x}-\boldsymbol{x}_0) + \boldsymbol{r}_2. \end{aligned} \quad (1.51)$$

1.2 Basics of the design based inference

In this thesis we consider the design-based approach, i.e. properties of the estimators, such as expectation and variance/covariance, are determined by the sampling design and by the study variable values in the finite population.

1.2.1 Sampling design

Consider a finite population $U = (1, 2, ..., N)$ that consist of N units. A probability sample is drawn from U according to some sampling design. Sampling design is a probability distribution of a random sampling vector $\mathbf{I} = (I_1, I_2, ..., I_N)'$:

$$\mathbf{I} \sim p(\boldsymbol{k}) = Pr(\mathbf{I} = \boldsymbol{k}), \quad (1.52)$$

where $\boldsymbol{k} = (k_1, k_2, ..., k_N)'$ is an outcome of $\mathbf{I}$ (Traat et al. 2004, Tillé, 2006). The random variable I_i indicates the number of selections of unit i from U, with $\mathbb{E}(I_i)$ being the expected number of selections.

For without-replacement (WOR) designs $I_i \in \{0, 1\}$ and for the with-replacement (WR) designs $I_i \in \{0, 1, 2, ...\}$. Note that the sample size n can be expressed as

$$n = \mathbf{I}'\mathbf{1},$$

where $\mathbf{1}$ is the N-dimensional vector of ones.

Depending on the sampling design, n can be random or fixed. A sampling design with fixed n is called a fixed size sampling design.

Two sampling designs are used in this thesis, the simple random sampling without replacement (SI) as the representative of the equal probability designs, and the multinomial (MN) design as the representative of the unequal probability designs.

For the SI design, the sampling vector has equal probabilities on all samples of size n, whereas other samples have zero probability,

$$Pr(\mathbf{I} = \boldsymbol{k}) = \binom{N}{n}^{-1}, \text{ if } \mathbf{I}'\mathbf{1} = n.$$

The important characteristics of the SI design, necessary for the design-based inference, are the expectations and the variances/covariances of the elements of $\mathbf{I}$, (Särndal *et al.*, 1991, p.66-72, Cochran, 1977, p.28-29). We use them in matrix form,

$$\mathbb{E}(\mathbf{I}) = (f, f, ..., f)', \tag{1.53}$$
$$\mathbb{C}\mathrm{ov}(\mathbf{I}) = \boldsymbol{\Delta} = f(1-f)(\mathbb{I} - \mathbf{C}), \tag{1.54}$$

where $f = n/N$ is sampling fraction; $\mathbf{C} : N \times N$ is the matrix with zeros on the main diagonal and $(N-1)^{-1}$ elsewhere, and $\mathbb{I}$ is the identity matrix.

For the MN design the distribution of the sampling vector $\mathbf{I}$ is the multinomial, $\mathbf{I} \sim M(n, p_1, p_2, ..., p_N)$ which gives probabilities for all samples of size n,

$$Pr(\mathbf{I} = \boldsymbol{k}) = n! \prod_{i=1}^{N} \frac{p_i^{k_i}}{k_i!}, \text{ if } \mathbf{I}'\mathbf{1} = n.$$

The characteristics of this design (see Traat and Ilves, 2007) in matrix form are:

$$\mathbb{E}(\mathbf{I}) = (np_1, np_2, ..., np_N)' = n\boldsymbol{p}; \tag{1.55}$$
$$\mathbb{C}\mathrm{ov}(\mathbf{I}) = \boldsymbol{\Delta} = \frac{1}{n}\left(\mathrm{diag}(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}'\right), \tag{1.56}$$

where $\boldsymbol{p} = (p_1, p_2, ..., p_N)'$ is the vector of selection probabilities.

The traditional name for the MN design in the literature is – the unequal probability with-replacement design. This name refers to the selection procedure of a sample, selection of units with replacement and with fixed selection probabilities at each selection step. The name multinomial design refers to the probability law of this design – the multinomial distribution.

The equal probability designs have a favorable feature – estimation can be perform without weighting. Thus, the sample mean and the sample proportion estimate unbiasedly the respective population parameters.

The unequal probability designs have another favorable feature, by choosing inclusion probabilities (or for our MN desin selection probabilities) proportional to the study variable, one can make estimators more precise.

1.2.2 Characteristics of estimators

Under design-based approach an estimator $\hat{\theta}$ is a discrete random variable taking values on a finite number of samples $\boldsymbol{k}$.

The design-based expectation of the estimator $\hat{\theta}$ is the weighted average of all possible values $\hat{\theta}(\boldsymbol{k})$ with weight $p(\boldsymbol{k})$ being the probability with which $\boldsymbol{k}$ is chosen,

$$\mathbb{E}(\hat{\theta}) = \sum_{\boldsymbol{k}} \hat{\theta}(\boldsymbol{k}) p(\boldsymbol{k}).$$

The summation goes over all possible samples $\boldsymbol{k}$ that can be obtained under given sampling design $p(\cdot)$.

In this work we need to consider the vector of estimators. We study the domain estimation, where the vector of estimators occurs naturally. For example, $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)'$ may present estimated numbers of unemployed in the three different regions (domains). For this purpose we bring the properties of the estimators in matrix form.

Let $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_d)'$ be random vector of estimators on the true parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_d)'$.

The *expectation* of $\hat{\boldsymbol{\theta}}$ is defined as the expectation of its elements:

$$\mathbb{E}(\hat{\boldsymbol{\theta}}) = (\mathbb{E}(\hat{\theta}_1), \mathbb{E}(\hat{\theta}_2), ..., \mathbb{E}(\hat{\theta}_d))'.$$

The following property holds:

$$\mathbb{E}(\mathbf{A}\hat{\boldsymbol{\theta}} + \boldsymbol{a}) = \mathbf{A}\mathbb{E}(\hat{\boldsymbol{\theta}}) + \boldsymbol{a}, \tag{1.57}$$

where $\mathbf{A} : m \times d$ is a constant matrix and $\boldsymbol{a}$ is a vector of m constants.

Let $\hat{\boldsymbol{\eta}} = (\hat{\eta}_1, \hat{\eta}_2, ..., \hat{\eta}_g)'$ be some other vector of estimators on the true parameter vector $\boldsymbol{\eta} = (\eta_1, \eta_2, ..., \eta_g)'$.

The *covariance* matrix between two random vectors $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\eta}}$ is defined as

$$\mathbb{C}\mathrm{ov}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}) = \mathbb{E}\left[\left(\hat{\boldsymbol{\theta}} - \mathbb{E}(\hat{\boldsymbol{\theta}})\right)(\hat{\boldsymbol{\eta}} - \mathbb{E}(\hat{\boldsymbol{\eta}}))'\right] : d \times g. \tag{1.58}$$

The *variance* of $\hat{\boldsymbol{\theta}}$ is the $d \times d$ matrix

$$\mathbb{C}\mathrm{ov}(\hat{\boldsymbol{\theta}}) = \mathbb{E}\left[\left(\hat{\boldsymbol{\theta}} - \mathbb{E}(\hat{\boldsymbol{\theta}})\right)\left(\hat{\boldsymbol{\theta}} - \mathbb{E}(\hat{\boldsymbol{\theta}})\right)'\right] \tag{1.59}$$

with one-dimensional variances $\mathbb{V}(\hat{\theta}_i)$ on the main diagonal and covariances $\mathbb{C}\mathrm{ov}(\hat{\theta}_i, \hat{\theta}_j)$ outside of it.

Next we define the mean square error in a multivariate form. The form is very general and includes in its special cases such well-known notions like covariance matrix of estimators and classical mean square error of the univariate estimator.

Definition 1.4 The *mean square error matrix* (MSE-matrix or shortly MSE) between two random vectors $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\eta}}$ is

$$\mathbb{MSE}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}) = \mathbb{E}\left[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})'\right] : d \times g. \tag{1.60}$$

The elements of $\mathbb{MSE}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}})$ are the cross mean square errors of $\hat{\theta}_i$ and $\hat{\eta}_j$, $\mathbb{E}(\hat{\theta}_i - \theta_i)(\hat{\eta}_j - \eta_j)$.

If $\boldsymbol{\theta} = \boldsymbol{\eta}$ and $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\eta}}$, then a shorter notation is used, $\mathbb{MSE}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}) = \mathbb{MSE}(\hat{\boldsymbol{\theta}})$, where

$$\mathbb{MSE}(\hat{\boldsymbol{\theta}}) = \mathbb{E}\left[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\right]. \tag{1.61}$$

The diagonal elements of the matrix $\mathbb{MSE}(\hat{\boldsymbol{\theta}})$ are the traditional mean square errors of one-dimensional estimator, $\mathbb{E}(\hat{\theta}_i - \theta)^2$.

The *bias* of the estimator $\hat{\boldsymbol{\theta}}$ is defined as vector of biases of elements $\hat{\theta}_i$, $i = 1, 2, ..., d$,

$$\boldsymbol{b}(\hat{\theta}) = \mathbb{E}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta}. \tag{1.62}$$

Proposition 1.5 The operator of $\mathbb{MSE}(\cdot)$ has the following properties.

$$\mathbb{MSE}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}) = \left[\mathbb{MSE}(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\theta}})\right]', \tag{1.63}$$

$$\mathbb{MSE}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}) = \mathbb{C}\mathrm{ov}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}) + \boldsymbol{b}(\hat{\theta})\boldsymbol{b}'(\hat{\eta}), \tag{1.64}$$

$$\mathbb{MSE}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}) = \mathbb{C}\mathrm{ov}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}), \text{ if } \boldsymbol{b}(\hat{\theta}) = \mathbf{0} \text{ or } \boldsymbol{b}(\hat{\eta}) = \mathbf{0}. \tag{1.65}$$

Proof. The property (1.63) follows from the Definition 1.4 and the property (1.4), applied to the right side of the expression.

For the property (1.64) we note that $\mathbb{E}(\hat{\boldsymbol{\theta}} - \mathbb{E}\hat{\boldsymbol{\theta}}) = \mathbf{0}$. Then (1.64) can be obtained in the following way:

$$\begin{aligned}
\mathbb{MSE}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}) &= \mathbb{E}\left[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})'\right] \\
&= \mathbb{E}\left[\left(\hat{\boldsymbol{\theta}} - (\boldsymbol{\theta} + \boldsymbol{b}(\hat{\theta})) + \boldsymbol{b}(\hat{\theta})\right)(\hat{\boldsymbol{\eta}} - (\boldsymbol{\eta} + \boldsymbol{b}(\hat{\eta})) + \boldsymbol{b}(\hat{\eta}))'\right] \\
&= \mathbb{E}\left[\left((\hat{\boldsymbol{\theta}} - \mathbb{E}(\hat{\boldsymbol{\theta}})) + \boldsymbol{b}(\hat{\theta})\right)((\hat{\boldsymbol{\eta}} - \mathbb{E}(\hat{\boldsymbol{\eta}})) + \boldsymbol{b}(\hat{\eta}))'\right] \\
&= \mathbb{E}\left[\left(\hat{\boldsymbol{\theta}} - \mathbb{E}(\hat{\boldsymbol{\theta}})\right)(\hat{\boldsymbol{\eta}} - \mathbb{E}(\hat{\boldsymbol{\eta}}))'\right] + \boldsymbol{b}(\hat{\theta})\boldsymbol{b}'(\hat{\eta}).
\end{aligned}$$

The property (1.65) follows directly from the property (1.64).

□

1.2.3 Estimation of the population and domain totals

The most frequent parameter of interest is the population total t_y,

$$t_y = \boldsymbol{y}'\mathbf{1}, \tag{1.66}$$

where $\boldsymbol{y} = (y_1, y_2, ..., y_N)'$ is the vector of the study variable measured on the population units.

The *linear estimator* of t_y is

$$\hat{t}_y = \boldsymbol{y}'\breve{\mathbf{I}} = \breve{\boldsymbol{y}}'\mathbf{I}. \tag{1.67}$$

In these two forms of the estimator the vector $\breve{\mathbf{I}} = (\breve{I}_1, \breve{I}_2, ..., \breve{I}_N)'$ is the expanded sampling vector, with elements

$$\breve{I}_i = \frac{I_i}{\mathbb{E}(I_i)}, \tag{1.68}$$

the vector $\breve{\boldsymbol{y}}$ is the expanded study variable vector with elements $\breve{y}_i = y_i/\mathbb{E}(I_i)$.

Since $\mathbb{E}(\breve{\mathbf{I}}) = \mathbf{1}$, the estimator (1.67) is design-unbiased, $\mathbb{E}(\hat{t}_y) = t_y$.

Estimate $\hat{t}_y$ is computed by weighting up the sampled values, $\hat{t}_y = \boldsymbol{y}'\breve{\mathbf{I}} = \sum_U \frac{I_i}{\mathbb{E}(I_i)} y_i$, where the summation goes over the all elements $i \in U$. From this prospective we may call the vector $\breve{\mathbf{I}}$ the weight vector.

Under a WOR design (1.67) is the Horvitz-Thompson (HT) estimator and under a WR designs it is the Hansen-Hurwitz estimator. The unified consideration of WOR and WR designs is not the usual one in sampling literature. It has been forcefully developed in Traat (2000), Traat et *al.* (2001, 2004), Meister (2004), Tillé (2006). The WOR designs are prevalent in real surveys. The Multinomial design is often used as an approximation to the complex WOR designs, while deriving properties of the estimators, but sometimes, the multinomial design or other WR designs are also used for drawing samples in real surveys (Traat, Ilves, 2007). We exemplify our results on SI and MN designs, therefore this unified consideration suites very well for this thesis.

The expanded sampling vector $\breve{\mathbf{I}}$ has a crucial role in the estimation. For the SI design its elements are $\breve{I}_i = I_i/f$, where $f = n/N$, for the MN design $\breve{I}_i = I_i/(np_i)$. Later, in this thesis, also the covariance matrix of $\breve{\mathbf{I}}$ is needed. For the SI design

$$\mathbb{C}\mathrm{ov}(\breve{\mathbf{I}}) = \breve{\boldsymbol{\Delta}} = \frac{(1-f)}{f}(\mathbb{I} - \mathbf{C}), \tag{1.69}$$

and for the MN design

$$\mathbb{C}\mathrm{ov}(\breve{\mathbf{I}}) = \breve{\boldsymbol{\Delta}} = \frac{1}{n}\left([\mathrm{diag}(\boldsymbol{p})]^{-1} - \mathbf{1}\mathbf{1}'\right). \tag{1.70}$$

The elements of $\breve{\boldsymbol{\Delta}}$ are $\mathbb{V}(I_i)/(\mathbb{E}I_i)^2$ on the main diagonal and $\mathbb{C}\mathrm{ov}(I_i, I_j)/(\mathbb{E}I_i\mathbb{E}I_j)$ outside.

Estimation of the domain parameters has become an undividable part of the estimation in a whole. As it is defined in Särndal et al. (1992, p. 386) we use the term *domain* for the subpopulation for which separate point estimates and confidence intervals are required. Domains can be for example, socio-economic groups (age by sex), geographical areas (counties, municipalities) or some other sub-populations (one-member, two-member, etc. households).

Estimation of domains can be requested before planning a survey (*planned* domains) or after it (*unplanned* domains). Sample sizes in unplanned domains are random and the respective samples may sometimes consist only of a few units. In this thesis we deal with unplanned domains.

Many methods are developed to construct possibly good estimators for domains with small sample sizes. These methods produce direct and undirect estimators. The domain estimator is called *direct* if it uses the study variable values only from the observed domain. The auxiliary information can be incorporated outside the domain. The linear estimator is the representative of the direct estimators, while the generalized regression and the synthetic estimators, described in Section 3.6 can be both direct or indirect, depending on the choice of a model behind them.

If different estimation methods are used in the domains, then the consistency problem occurs - the estimators do not sum up to the estimators used for the population total, or for the larger domains under interest.

Let U be divided into D non-overlapping domains U_d, $d \in \mathcal{D} = \{1, 2, ..., D\}$ with N_d being the size of the domain U_d. We are interested in the domain totals of study variable $\boldsymbol{y}$:

$$t^d = \sum_{i \in U_d} y_i. \tag{1.71}$$

We assume here that we can identify whether the object $i \in U$ belongs to the domain or not. Traditionally, the domain indicator-vector is defined,

$$\boldsymbol{\delta}_d = (\delta_{1d}, \delta_{2d}, ..., \delta_{Nd})',\ d \in \mathcal{D},$$

where $\delta_{id} = 1$ if $i \in U_d$ and 0 otherwise. It enables to carry over the estimation results of the population total for the domain estimation. Accordingly, a new study variable $\boldsymbol{y}_d$ is defined, $\boldsymbol{y}_d = \boldsymbol{y} \circ \boldsymbol{\delta}_d = \mathrm{diag}(\boldsymbol{\delta})\boldsymbol{y}$ and the domain total (1.71) can be rewritten as

$$t^d = \sum_{i \in U} \delta_{id} y_i = \boldsymbol{y}_d' \mathbf{1}. \tag{1.72}$$

Now (1.72) can be viewed as a population total of the new variable $\boldsymbol{y}_d$.

Many estimators are available for the population totals. Beside the linear estimator, there are estimators using auxiliary information. Auxiliary variable is any variable about which information is available and complete at unit level for all population units (this information may come from registers). For some estimators it is enough to know the population totals of auxiliary variables, and only for the sample units information at the unit level. Properties of the estimators using auxiliary information, the generalized regression and the synthetic estimators, are studied in Section 3.6.

Chapter 2

Estimation of domains under restrictions

In this chapter three new estimators (GR1, GR2, GR3) are defined for domain estimation under restrictions. They are more general than the Knottnerus (2003) GR estimator since they can handle biased initial estimators. Their properties are studied, the expressions for bias and mean square errors are derived. Their ordering with respect to the accuracy is established.

The users of official statistics often require that sample-based estimates satisfy certain restrictions. In the domain's case it is required that the estimates of domain totals sum up to the population total or to its estimate. For example, in time domains, quarterly estimates have to sum up to the yearly total. The relationships holding for the true population parameters do not necessarily hold for the respective estimates. This inconsistency of estimates is annoying for statistics users. On the other hand, known relationships between population parameters is a kind of auxiliary information. Involving this information into estimation process presumably improves estimates. Our goal is to define consistent domain estimators that are more accurate than the initial inconsistent domain estimators.

One solution to the problem of finding estimates under restrictions is the general restriction estimator (GR) proposed by Knottnerus (2003). His estimator is based on the unbiased initial estimators and is unbiased itself.

The advantage of the GR estimator is the variance minimizing property in a class of linear estimators. Sõstra (2007) has developed the GR estimator for estimating domain totals under summation restriction. Optimality property of the domain GR estimator is studied in Sõstra and Traat (2009). In all these works, the unbiased or asymptotically unbiased initial estimators are assumed.

It is well known that there are many useful estimators that are biased. For example, the model-based small area estimators are design-biased. The synthetic estimator can be biased on the domain level. Even the widely used GREG estimator is only asymptotically unbiased. In this thesis we will allow the initial estimators to be biased, and will construct three new restriction estimators, based on the biased initial estimators.

2.1 The GR estimator

Let $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)'$ be the parameter vector under study that satisfies linear restrictions:

$$\mathbf{R}\boldsymbol{\theta} = \boldsymbol{c}, \tag{2.1}$$

where $\mathbf{R}$ is an $r \times k$ matrix of rank r and $\boldsymbol{c}$ is the r-dimensional vector of known constants.

For example, if D domain totals, say t_y^d, where $t_y^d = \sum_{U_d} y_i$, $d = 1, 2, ..., D$, have to sum up to the population total $t_y = \sum_U y_i$, then the components of the restriction equation (2.1) are:

$$\mathbf{R} = (1, 1, ..., 1, -1) : 1 \times (D+1),\ \boldsymbol{\theta} = (t_y^1, t_y^2, ..., t_y^D, t_y)' \text{ and } \boldsymbol{c} = 0. \tag{2.2}$$

Alternatively, the same requirement is achieved by choosing in (2.1)

$$\mathbf{R} = (1, 1, ..., 1) : 1 \times D,\ \boldsymbol{\theta} = (t_y^1, t_y^2, ..., t_y^D)' \text{ and } \boldsymbol{c} = t_y.$$

In the latter case, the population total must be known while developing restriction estimators for domains. In many cases this is not so, and then the estimated population total must be used. Respectively, the components of the restriction equation (2.1) are of type (2.2).

Theorem 2.1 (Knottnerus, 2003, p. 328-329) Let $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, ..., \hat{\theta}_k)'$ be a vector of unbiased estimators of the parameter vector $\boldsymbol{\theta}$ with the variance

$\mathbf{V}$, such that $\mathbf{RVR}'$ can be inverted. Then the general restriction estimator $\hat{\boldsymbol{\theta}}_{GR}$ that satisfies restrictions (2.1) for $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{GR}$, and the variance $\mathbf{V}_{GR}$ of this estimator are:

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{GR} &= \hat{\boldsymbol{\theta}} + \mathbf{K}(\boldsymbol{c} - \mathbf{R}\hat{\boldsymbol{\theta}}), & (2.3) \\ \mathbf{V}_{GR} &= \mathbb{C}\mathrm{ov}(\hat{\boldsymbol{\theta}}_{GR}) = (\mathbb{I} - \mathbf{KR})\mathbf{V}, & (2.4) \end{aligned}$$

where $\mathbb{I}$ is the $k \times k$ identity matrix and

$$\mathbf{K} = \mathbf{VR}'(\mathbf{RVR}')^{-1}. \tag{2.5}$$

Since $\mathbf{RK}$ is the identity matrix, it is easy to check that $\hat{\boldsymbol{\theta}}_{GR}$ satisfies restrictions (2.1):

$$\mathbf{R}\hat{\boldsymbol{\theta}}_{GR} = \mathbf{R}\hat{\boldsymbol{\theta}} + \mathbf{RK}(\boldsymbol{c} - \mathbf{R}\hat{\boldsymbol{\theta}}) = \boldsymbol{c}.$$

Knottnerus (2003, p. 332) shows that $\hat{\boldsymbol{\theta}}_{GR}$ is optimal in a class of estimators that are linear in $\hat{\boldsymbol{\theta}}$ and satisfy restrictions (2.1). In this class, $\hat{\boldsymbol{\theta}}_{GR}$ has minimum variance (in Löwner ordering). For example, other estimators in this class can be received by replacing $\mathbf{V}$ in the expression of $\mathbf{K}$ by any arbitrary $k \times k$ matrix $\mathbf{V}^*$, such that $\mathbf{RV}^*\mathbf{R}$ can be inverted. But the resulting estimators have bigger variance than $\hat{\boldsymbol{\theta}}_{GR}$. In Sõstra (2007, p. 45) it is also shown that $\hat{\boldsymbol{\theta}}_{GR}$ is never less efficient than the initial estimator $\hat{\boldsymbol{\theta}}$, $\mathbf{V}_{GR} \leq \mathbf{V}$ in the sense of Löwner ordering.

Without loss of generality, we further consider linear restrictions in the form

$$\mathbf{R}\boldsymbol{\theta} = \mathbf{0}. \tag{2.6}$$

In general, if $\boldsymbol{c} \neq \mathbf{0}$ in (2.1), it is always possible to choose fixed $\boldsymbol{\theta}_0$ so that

$$\mathbf{0} = \mathbf{R}\boldsymbol{\theta} - \boldsymbol{c} = \mathbf{R}\boldsymbol{\theta} - \mathbf{R}\boldsymbol{\theta}_0 = \mathbf{R}(\boldsymbol{\theta} - \boldsymbol{\theta}_0), \tag{2.7}$$

and consider new parameter $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta} - \boldsymbol{\theta}_0$ instead of $\boldsymbol{\theta}$. For example, in the case of 3 domain totals $t_y^d, d = 1, 2, 3$, the restriction

$$(1,1,1)(t_y^1, t_y^2, t_y^3)' = t_y$$

can be rewritten as

$$(1,1,1)\left[(t_y^1, t_y^2, t_y^3) - \frac{1}{3}(t_y, t_y, t_y)\right]' = 0,$$

where $\boldsymbol{\theta}_0 = \frac{1}{3}(t_y, t_y, t_y)'$. With $\boldsymbol{c} = \mathbf{0}$, the Knottnerus' GR estimator simplifies to the form

$$\hat{\boldsymbol{\theta}}_{GR} = (\mathbb{I} - \mathbf{KR})\hat{\boldsymbol{\theta}}. \tag{2.8}$$

In the following section we allow initial estimator to be biased, and we define three different restriction estimators for this case.

2.2 Restriction estimators handling bias

Assume that estimator $\hat{\boldsymbol{\theta}}$ is biased for $\boldsymbol{\theta}$,

$$\mathbb{E}(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta} + \boldsymbol{b}, \tag{2.9}$$

where $\boldsymbol{b}$ is a vector of biases.

For biased estimators the accuracy of the estimator is ordinarily measured by its mean square error. The GR-estimator (2.8) with biased initial estimator $\hat{\boldsymbol{\theta}}$ is not optimal any more for $\boldsymbol{\theta}$ in the sense of MSE. Although it still satisfies restrictions (2.6), it may have bigger mean square error than that of the initial estimator. We demonstrate this by the following example.

Example 2.1 Let $\boldsymbol{\theta} = (\theta_1, \theta_2)'$ be the vector of unknown parameters, $\mathbf{R} = (1, 1)$ and $\boldsymbol{c} = 0$.
The vector of estimators $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2)'$ has variance $\mathbf{V} = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$ and bias $\boldsymbol{b} = (3, -1)'$.

From (1.61) and (1.64), the mean square error of $\boldsymbol{\theta}$ is

$$\mathbb{MSE}(\hat{\boldsymbol{\theta}}) = \mathbf{V} + \boldsymbol{b}\,\boldsymbol{b}' = \begin{pmatrix} 11 & -2 \\ -2 & 4 \end{pmatrix},$$

which means that $\mathbb{MSE}(\hat{\theta}_1) = 11$ and $\mathbb{MSE}(\hat{\theta}_2) = 4$. We will find now the GR estimator defined in (2.8) and its mean square error.

The matrix $\mathbf{K}$, needed for the $\hat{\boldsymbol{\theta}}_{GR}$, is $\mathbf{K} = \mathbf{VR}'(\mathbf{RVR}')^{-1} \approx (0.43, 0.57)'$, and

$$\hat{\boldsymbol{\theta}}_{GR} = \begin{pmatrix} 0.57\hat{\theta}_1 - 0.43\hat{\theta}_2 \\ -0.57\hat{\theta}_1 + 0.43\hat{\theta}_2 \end{pmatrix}.$$

The variance (2.4),

$$\mathbf{V}_{GR} = \mathbb{C}\mathrm{ov}(\hat{\boldsymbol{\theta}}_{GR}) = \begin{pmatrix} 0.71 & -0.71 \\ -0.71 & 0.71 \end{pmatrix},$$

is smaller than the initial $\mathbf{V}$ (in the sense of Löwner ordering), since

$$\mathbf{V} - \mathbf{V}_{GR} = \begin{pmatrix} 1.29 & 1.71 \\ 1.71 & 2.29 \end{pmatrix}$$

is positive definite (eigenvalues $3.572, 0.008$). However, the obtained $\hat{\boldsymbol{\theta}}_{GR}$ is biased now. The first element of the bias-vector $\boldsymbol{b}(\hat{\boldsymbol{\theta}}_{GR}) = \mathbb{E}(\hat{\boldsymbol{\theta}}_{GR}) - \boldsymbol{\theta}$. $\boldsymbol{b}(\hat{\boldsymbol{\theta}}_{GR})$ is

$$\begin{aligned} \boldsymbol{b}(\hat{\boldsymbol{\theta}}_{GR})^{(1)} &= \mathbb{E}\left(\hat{\boldsymbol{\theta}}_{GR}^{(1)}\right) - \theta_1 = 0.57\,\mathbb{E}(\hat{\theta}_1) - 0.43\,\mathbb{E}(\hat{\theta}_2) - \theta_1 \\ &= 0.57(\theta_1 + 3) - 0.43(\theta_2 - 1) - \theta_1 \\ &= -0.43(\theta_1 + \theta_2) + 2.14 = 2.14. \end{aligned}$$

In the last row $\theta_1 + \theta_2$ vanishes due to the restriction $(1,1)\boldsymbol{\theta} = 0$. Analogously, $\boldsymbol{b}(\hat{\boldsymbol{\theta}}_{GR})^{(2)} = -2.14$.

Finally, the mean square error matrix of the GR estimator is

$$\mathbb{M}\mathrm{SE}(\hat{\boldsymbol{\theta}}_{GR}) = \mathbb{C}\mathrm{ov}(\hat{\boldsymbol{\theta}}_{GR}) + \boldsymbol{b}(\hat{\boldsymbol{\theta}}_{GR})\,\boldsymbol{b}(\hat{\boldsymbol{\theta}}_{GR})' = \begin{pmatrix} 5.31 & -5.31 \\ -5.31 & 5.31 \end{pmatrix},$$

from which we get $\mathbb{M}\mathrm{SE}(\hat{\boldsymbol{\theta}}_{GR}^{(1)}) = \mathbb{M}\mathrm{SE}(\hat{\boldsymbol{\theta}}_{GR}^{(2)}) = 5.31$. Comparing the first components, the mean square error of the GR estimator is smaller than that of $\hat{\boldsymbol{\theta}}$, ($\mathbb{M}\mathrm{SE}(\hat{\theta}_1) = 11$), but for the second component it is bigger ($\mathbb{M}\mathrm{SE}(\hat{\theta}_2) = 4$). Consequently, for the biased initial estimators we can not expect bigger accuracy from the Knottnerus GR estimator.

In the following sections we present three new restriction estimators for biased initial estimators.

2.2.1 Restriction estimator GR1

The first restriction estimator with biased initial estimators is defined in the following proposition, where also its properties are proved.

Proposition 2.2 The estimator

$$\hat{\boldsymbol{\theta}}_{GR1} = (\mathbb{I} - \mathbf{KR})(\hat{\boldsymbol{\theta}} - \boldsymbol{b}), \tag{2.10}$$

with $\mathbf{K} = \mathbf{VR}'(\mathbf{RVR}')^{-1}$ is unbiased for $\boldsymbol{\theta}$. Its variance is

$$\mathbb{C}\mathrm{ov}(\hat{\boldsymbol{\theta}}_{GR1}) = (\mathbb{I} - \mathbf{KR})\mathbf{V}, \tag{2.11}$$

and it is the optimal estimator among all linear estimators in $(\hat{\boldsymbol{\theta}} - \boldsymbol{b})$ that satisfy restriction (2.6).

Proof. The unbiasedness follows from (2.9) and (2.6),

$$\mathbb{E}(\hat{\boldsymbol{\theta}}_{GR1}) = (\mathbb{I} - \mathbf{KR})(\mathbb{E}\hat{\boldsymbol{\theta}} - \boldsymbol{b}) = (\mathbb{I} - \mathbf{KR})\boldsymbol{\theta} = \boldsymbol{\theta}. \tag{2.12}$$

Since $\mathbf{RK} = \mathbf{RVR}'(\mathbf{RV}'\mathbf{R}')^{-1} = \mathbb{I}_r$ (dimensionality is $r \times r$ here), it is obvious that $\hat{\boldsymbol{\theta}}_{GR1}$ satisfies the linear restriction (2.6),

$$\mathbf{R}\hat{\boldsymbol{\theta}}_{GR1} = (\mathbf{R}\mathbb{I} - \mathbf{RKR})(\hat{\boldsymbol{\theta}} - \boldsymbol{b}) = \mathbf{0}.$$

Denoting $\hat{\boldsymbol{\theta}}^* = \hat{\boldsymbol{\theta}} - \boldsymbol{b}$, we know from Knottnerus' results that $\mathbf{R}\hat{\boldsymbol{\theta}}^*_{GR} = (\mathbb{I} - \mathbf{KR})\hat{\boldsymbol{\theta}}^*$ is optimal among linear estimators in $\hat{\boldsymbol{\theta}}^*$ that satisfy restrictions (2.6) for $\mathbf{K} = \mathbb{C}\mathrm{ov}(\hat{\boldsymbol{\theta}}^*)\mathbf{R}' \left[\mathbf{R}\,\mathbb{C}\mathrm{ov}(\hat{\boldsymbol{\theta}}^*)\mathbf{R}'\right]^{-1}$.

But

$$\mathbb{C}\mathrm{ov}(\hat{\boldsymbol{\theta}}^*) = \mathbb{C}\mathrm{ov}(\hat{\boldsymbol{\theta}} - \boldsymbol{b}) = \mathbb{C}\mathrm{ov}(\hat{\boldsymbol{\theta}}) = \mathbf{V},$$

meaning that our estimator in (2.10) is the optimal estimator.

The variance of $\hat{\boldsymbol{\theta}}^*_{GR}$ follows from Knottnerus' result (2.4),

$$\mathbb{C}\mathrm{ov}(\hat{\boldsymbol{\theta}}^*_{GR}) = (\mathbb{I} - \mathbf{KR})\mathbb{C}\mathrm{ov}(\hat{\boldsymbol{\theta}}^*) = (\mathbb{I} - \mathbf{KR})\mathbf{V}.$$

□

Similarly to Knottnerus GR estimator our $\hat{\boldsymbol{\theta}}_{GR1}$ requires quantities that are usually unknown in practise, here the bias $\boldsymbol{b}$ and the variance $\mathbf{V}$. If $\mathbf{V}$ and $\boldsymbol{b}$ are replaced with consistent estimators, $\hat{\boldsymbol{\theta}}_{GR1}$ is consistent itself. In this thesis, however, we concentrate on the GR estimators with known $\boldsymbol{b}$, $\mathbf{V}$ and later $\mathbf{M}$.

2.2.2 Restriction estimator GR2

Below we define an estimator that is free of the knowledge of $\boldsymbol{b}$, satisfies restrictions and is more accurate than the initial estimator $\hat{\boldsymbol{\theta}}$, in $\mathbb{M}$SE terms.

Proposition 2.3 The estimator, satisfying restrictions (2.6), but based on the mean square error $\mathbf{M}$ of the initial estimator $\hat{\boldsymbol{\theta}}$, is

$$\hat{\boldsymbol{\theta}}_{GR2} = (\mathbb{I} - \mathbf{K}^*\mathbf{R})\hat{\boldsymbol{\theta}}, \tag{2.13}$$

where $\mathbf{K}^* = \mathbf{M}\mathbf{R}'(\mathbf{R}\mathbf{M}\mathbf{R}')^{-1}$. The bias of the $\hat{\boldsymbol{\theta}}_{GR2}$ is

$$\boldsymbol{b}(\hat{\boldsymbol{\theta}}_{GR2}) = (\mathbb{I} - \mathbf{K}^*\mathbf{R})\boldsymbol{b}, \tag{2.14}$$

and the mean square error matrix is

$$\mathbb{M}\mathrm{SE}(\hat{\boldsymbol{\theta}}_{GR2}) = (\mathbb{I} - \mathbf{K}^*\mathbf{R})\mathbf{M}. \tag{2.15}$$

Furthermore,

$$\mathbb{M}\mathrm{SE}(\hat{\boldsymbol{\theta}}_{GR2}) \leq \mathbf{M} \tag{2.16}$$

in the sense of Löwner ordering.

Proof. Analogously to the Proposition 2.2, we see that $\mathbf{R}\hat{\boldsymbol{\theta}}_{GR2} = \mathbf{0}$.

For the bias in (2.14) and the mean square error matrix in (2.15) we notice first that due to $\mathbf{R}\boldsymbol{\theta} = \mathbf{0}$,

$$(\mathbb{I} - \mathbf{K}^*\mathbf{R})\boldsymbol{\theta} = \boldsymbol{\theta} - \mathbf{K}^*\mathbf{R}\boldsymbol{\theta} = \boldsymbol{\theta}. \tag{2.17}$$

Then, the bias expression follows from (2.9) and restrictions (2.6),

$$\begin{aligned}
\boldsymbol{b}(\hat{\boldsymbol{\theta}}_{GR2}) &= \mathbb{E}(\hat{\boldsymbol{\theta}}_{GR2}) - \boldsymbol{\theta} = (\mathbb{I} - \mathbf{K}^*\mathbf{R})\mathbb{E}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta} \\
&= (\mathbb{I} - \mathbf{K}^*\mathbf{R})(\boldsymbol{\theta} + \boldsymbol{b}) - \boldsymbol{\theta} = (\mathbb{I} - \mathbf{K}^*\mathbf{R})\boldsymbol{\theta} - \boldsymbol{\theta} + (\mathbb{I} - \mathbf{K}^*\mathbf{R})\boldsymbol{b} \\
&= (\mathbb{I} - \mathbf{K}^*\mathbf{R})\boldsymbol{b}.
\end{aligned}$$

From Definition 1.4 of the mean square error matrix and (2.17) we have

$$\begin{aligned}
\mathbb{M}\mathrm{SE}(\hat{\boldsymbol{\theta}}_{GR2}) &= \mathbb{E}\Big((\mathbb{I} - \mathbf{K}^*\mathbf{R})\hat{\boldsymbol{\theta}} - (\mathbb{I} - \mathbf{K}^*\mathbf{R})\boldsymbol{\theta}\Big)\Big((\mathbb{I} - \mathbf{K}^*\mathbf{R})\hat{\boldsymbol{\theta}} - (\mathbb{I} - \mathbf{K}^*\mathbf{R})\boldsymbol{\theta}\Big)' \\
&= (\mathbb{I} - \mathbf{K}^*\mathbf{R}) \cdot \mathbb{E}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \cdot (\mathbb{I} - \mathbf{K}^*\mathbf{R})' \\
&= (\mathbb{I} - \mathbf{K}^*\mathbf{R})\mathbf{M}(\mathbb{I} - \mathbf{K}^*\mathbf{R})' \\
&= (\mathbb{I} - \mathbf{K}^*\mathbf{R})\mathbf{M} - (\mathbb{I} - \mathbf{K}^*\mathbf{R})\mathbf{M}\mathbf{R}'(\mathbf{K}^*)'. \qquad (2.18)
\end{aligned}$$

We now show that the second term in (2.18) is equal to zero:

$$\begin{aligned}(\mathbb{I} - \mathbf{K}^*\mathbf{R})\mathbf{M}\mathbf{R}'(\mathbf{K}^*)' &= \mathbf{M}\mathbf{R}'(\mathbf{K}^*)' - \mathbf{K}^*\mathbf{R}\mathbf{M}\mathbf{R}'(\mathbf{K}^*)' \\ &= \mathbf{M}\mathbf{R}'(\mathbf{K}^*)' - \left(\mathbf{M}\mathbf{R}'(\mathbf{R}\mathbf{M}\mathbf{R}')^{-1}\right)\mathbf{R}\mathbf{M}\mathbf{R}'(\mathbf{K}^*)' = 0.\end{aligned}$$

Thus, the expression of the MSE in (2.18) is equal to (2.15).

Finally, we show that $\mathbb{MSE}(\hat{\boldsymbol{\theta}}_{GR2}) \leq \mathbf{M}$ in the sense of Löwner ordering. This is equivalent to $\mathbf{M} - \mathbb{MSE}(\hat{\boldsymbol{\theta}}_{GR2}) \geq \mathbf{0}$,

$$\begin{aligned}\mathbf{M} - \mathbb{MSE}(\hat{\boldsymbol{\theta}}_{GR2}) &= \mathbf{M} - (\mathbb{I} - \mathbf{K}^*\mathbf{R})\mathbf{M} \\ &= \mathbf{K}^*\mathbf{R}\mathbf{M} = \mathbf{M}\mathbf{R}'(\mathbf{R}\mathbf{M}\mathbf{R}')^{-1}\mathbf{R}\mathbf{M} \geq \mathbf{0}, \quad (2.19)\end{aligned}$$

because it is of the shape $\mathbf{A}\mathbf{A}'$.

□

2.2.3 Restriction estimator GR3

In the following proposition properties of the estimator $\hat{\boldsymbol{\theta}}_{GR3}$ are proved.

Proposition 2.4 The restriction estimator

$$\hat{\boldsymbol{\theta}}_{GR3} = (\mathbb{I} - \mathbf{K}^*\mathbf{R})(\hat{\boldsymbol{\theta}} - \boldsymbol{b}) \quad (2.20)$$

with $\mathbf{K}^* = \mathbf{M}\mathbf{R}'(\mathbf{R}\mathbf{M}\mathbf{R}')^{-1}$ satisfies restrictions (2.6) and is unbiased for $\hat{\boldsymbol{\theta}}$. It's MSE is the covariance of the estimator and is equal to

$$\mathbb{MSE}(\hat{\boldsymbol{\theta}}_{GR3}) = (\mathbb{I} - \mathbf{K}^*\mathbf{R})\mathbf{V}(\mathbb{I} - \mathbf{K}^*\mathbf{R})'. \quad (2.21)$$

Furthermore,

$$\mathbb{MSE}(\hat{\boldsymbol{\theta}}_{GR3}) \leq \mathbf{M}. \quad (2.22)$$

Proof. The proof of unbiasedness is analogous to the proof of (2.12).

The covariance of $\hat{\boldsymbol{\theta}}_{GR3}$ follows directly from definition of the covariance in matrix form (1.59) and the property (1.57) of the operator $\mathbb{E}(\cdot)$,

$$\begin{aligned}\mathbb{C}\mathrm{ov}(\hat{\boldsymbol{\theta}}_{GR3}) &= \mathbb{E}\left[\left(\hat{\boldsymbol{\theta}}_{GR3} - \mathbb{E}(\hat{\boldsymbol{\theta}}_{GR3})\right)\left(\hat{\boldsymbol{\theta}}_{GR3} - \mathbb{E}(\hat{\boldsymbol{\theta}}_{GR3})\right)'\right] \\ &= \mathbb{E}\left[\left((\mathbb{I} - \mathbf{K}^*\mathbf{R})(\hat{\boldsymbol{\theta}} - \boldsymbol{b}) - \boldsymbol{\theta}\right)\left((\mathbb{I} - \mathbf{K}^*\mathbf{R})(\hat{\boldsymbol{\theta}} - \boldsymbol{b}) - \boldsymbol{\theta}\right)'\right]\end{aligned}$$

Replacing $\boldsymbol{\theta} = (\mathbb{I} - \mathbf{K}^*\mathbf{R})\boldsymbol{\theta}$ we have,

$$\begin{aligned}
\mathbb{C}\text{ov}(\hat{\boldsymbol{\theta}}_{GR3}) &= (\mathbb{I} - \mathbf{K}^*\mathbf{R})\,\mathbb{E}\left[\left(\hat{\boldsymbol{\theta}} - (\boldsymbol{\theta} + \boldsymbol{b})\right)\left(\hat{\boldsymbol{\theta}} - (\boldsymbol{\theta} + \boldsymbol{b})\right)'\right](\mathbb{I} - \mathbf{K}^*\mathbf{R})' \\
&= (\mathbb{I} - \mathbf{K}^*\mathbf{R})\,\mathbb{E}\left[\left(\hat{\boldsymbol{\theta}} - \mathbb{E}(\hat{\boldsymbol{\theta}})\right)\left(\hat{\boldsymbol{\theta}} - \mathbb{E}(\hat{\boldsymbol{\theta}})\right)'\right](\mathbb{I} - \mathbf{K}^*\mathbf{R})' \\
&= (\mathbb{I} - \mathbf{K}^*\mathbf{R})\ \mathbf{V}\ (\mathbb{I} - \mathbf{K}^*\mathbf{R})'.
\end{aligned}$$

Finally, we show that $\mathbf{M} - \mathbb{M}\text{SE}(\hat{\boldsymbol{\theta}}_{GR3}) \geq \mathbf{0}$, which is equivalent to (2.22):

$$\begin{aligned}
\mathbf{M} - (\mathbb{I} - \mathbf{K}^*\mathbf{R})\mathbf{V}(\mathbb{I} - \mathbf{K}^*\mathbf{R})' &= \mathbf{M} - (\mathbb{I} - \mathbf{K}^*\mathbf{R})(\mathbf{M} - \boldsymbol{b}\boldsymbol{b}')(\mathbb{I} - \mathbf{K}^*\mathbf{R})' \\
&= \mathbf{M} - (\mathbb{I} - \mathbf{K}^*\mathbf{R})\mathbf{M}(\mathbb{I} - \mathbf{K}^*\mathbf{R})' \\
&\quad + (\mathbb{I} - \mathbf{K}^*\mathbf{R})\boldsymbol{b}\boldsymbol{b}'(\mathbb{I} - \mathbf{K}^*\mathbf{R}). \qquad (2.23)
\end{aligned}$$

In (2.18) we showed that $(\mathbb{I} - \mathbf{K}^*\mathbf{R})\mathbf{M}(\mathbb{I} - \mathbf{K}^*\mathbf{R})' = (\mathbb{I} - \mathbf{K}^*\mathbf{R})\mathbf{M}$, and in (2.19) that $\mathbf{M} - (\mathbb{I} - \mathbf{K}^*\mathbf{R})\mathbf{M}$ is nonnegative definite. The third term of (2.23) is also nonnegative definite because of the shape $\mathbf{AA'}$. The sum of nonnegative definite matrices is also a nonnegative definite matrix, which proves (2.22) and the proposition as a whole.

□

2.2.4 Comparison of the GR estimators

As we saw from Propositions 2.2-2.4, estimators $\hat{\boldsymbol{\theta}}_{GR1}$, $\hat{\boldsymbol{\theta}}_{GR2}$ and $\hat{\boldsymbol{\theta}}_{GR3}$ have higher accuracy than the initial estimator $\hat{\boldsymbol{\theta}}$. The next result compares the accuracy of all four estimators.

Proposition 2.5 The mean square error matrices of the restriction estimators $\hat{\boldsymbol{\theta}}_{GR1}$, $\hat{\boldsymbol{\theta}}_{GR2}$, $\hat{\boldsymbol{\theta}}_{GR3}$ and the initial estimator $\hat{\boldsymbol{\theta}}$ can be ordered (in the sense of Löwner ordering) as following:

$$\mathbb{M}\text{SE}(\hat{\boldsymbol{\theta}}_{GR1}) \leq \mathbb{M}\text{SE}(\hat{\boldsymbol{\theta}}_{GR3}) \leq \mathbb{M}\text{SE}(\hat{\boldsymbol{\theta}}_{GR2}) \leq \mathbb{M}\text{SE}(\hat{\boldsymbol{\theta}}). \qquad (2.24)$$

Proof. From unbiasedness of $\hat{\boldsymbol{\theta}}_{GR1}$ and $\hat{\boldsymbol{\theta}}_{GR3}$ we note that $\mathbb{M}\text{SE}(\hat{\boldsymbol{\theta}}_{GR1}) = \mathbb{C}\text{ov}(\hat{\boldsymbol{\theta}}_{GR1})$ and $\mathbb{M}\text{SE}(\hat{\boldsymbol{\theta}}_{GR3}) = \mathbb{C}\text{ov}(\hat{\boldsymbol{\theta}}_{GR3})$. From Proposition 2.2 the estimator $\hat{\boldsymbol{\theta}}_{GR1}$ is optimal, i.e. it has the minimum variance (and also the mean square error matrix) among all linear estimators in $(\hat{\boldsymbol{\theta}} - \boldsymbol{b})$. So, it has smaller variance than the estimator $\hat{\boldsymbol{\theta}}_{GR3}$, which is of the same structure. This proves the first inequality.

The second inequality, $\mathbb{MSE}(\hat{\boldsymbol{\theta}}_{GR3}) \leq \mathbb{MSE}(\hat{\boldsymbol{\theta}}_{GR2})$, comes from the expression (2.21) of the Proposition 2.4 and (2.18) from the proof of the Proposition 2.3,

$$\begin{aligned}
\mathbb{MSE}(\hat{\boldsymbol{\theta}}_{GR2}) &= (\mathbb{I} - \mathbf{K}^*\mathbf{R})\mathbf{M}(\mathbb{I} - \mathbf{K}^*\mathbf{R})' \\
&= (\mathbb{I} - \mathbf{K}^*\mathbf{R})(\mathbf{V} + \boldsymbol{b}\boldsymbol{b'})(\mathbb{I} - \mathbf{K}^*\mathbf{R})' \\
&= (\mathbb{I} - \mathbf{K}^*\mathbf{R})\mathbf{V}(\mathbb{I} - \mathbf{K}^*\mathbf{R})' + (\mathbb{I} - \mathbf{K}^*\mathbf{R})(\boldsymbol{b}\boldsymbol{b}')(\mathbb{I} - \mathbf{K}^*\mathbf{R})' \\
&= \mathbb{MSE}(\hat{\boldsymbol{\theta}}_{GR3}) + (\mathbb{I} - \mathbf{K}^*\mathbf{R})\boldsymbol{b}\,[(\mathbb{I} - \mathbf{K}^*\mathbf{R})\boldsymbol{b}]'
\end{aligned}$$

$\geq \mathbb{MSE}(\hat{\boldsymbol{\theta}}_{GR3})$,

because $(\mathbb{I} - \mathbf{K}^*\mathbf{R})\boldsymbol{b}\,[(\mathbb{I} - \mathbf{K}^*\mathbf{R})\boldsymbol{b}]'$ is non-negative definite.

The last inequality comes directly from (2.16) of Proposition 2.3.

□

Remark 2.1 As it is shown in Proposition 2.5, $\hat{\boldsymbol{\theta}}_{GR1}$ and $\hat{\boldsymbol{\theta}}_{GR3}$ produce more accurate estimates than $\hat{\boldsymbol{\theta}}_{GR2}$. But they require knowledge of the initial bias $\boldsymbol{b}$. Estimator $\hat{\boldsymbol{\theta}}_{GR2}$ does not involve initial bias explicitly. Though this estimator is not so accurate than $\hat{\boldsymbol{\theta}}_{GR1}$ and $\hat{\boldsymbol{\theta}}_{GR3}$, it may be preferable in the situations where it is easier to estimate the MSE matrix $\mathbf{M}$ than the bias $\boldsymbol{b}$.

2.2.5 Searching optimality among GR2-type estimators

We consider one more restriction estimator similar to GR2. It does not require bias $\boldsymbol{b}$ in its expression, and also, instead of the matrix $\mathbf{M}$, any matrix $\mathbf{B}$ is allowed. The estimator is,

$$\hat{\boldsymbol{\theta}}_{GR4} = (\mathbb{I} - \mathbf{L}\mathbf{R})\hat{\boldsymbol{\theta}}, \tag{2.25}$$

where $\mathbf{L} = \mathbf{B}\mathbf{R}'(\mathbf{R}\mathbf{B}\mathbf{R}')^{-1}$ and $\mathbf{B}$ is an $k \times k$ unknown matrix.

We want to find such $\mathbf{B}$ that produces minimum $\mathbb{MSE}(\hat{\boldsymbol{\theta}}_{GR4})$ among all possible linear estimators of the structure (2.25).

Proposition 2.6 The mean square error of the restriction estimator (2.25) is

$$\mathbb{MSE}(\hat{\boldsymbol{\theta}}_{GR4}) = (\mathbb{I} - \mathbf{LR})\mathbf{M}(\mathbb{I} - \mathbf{LR})', \tag{2.26}$$

where $\mathbf{L} = \mathbf{BR}'(\mathbf{RBR}')^{-1}$.

The proof of the Proposition 2.6 is similar to the proof of (2.17) and (2.18) from the Proposition 2.3.

To find the matrix $\mathbf{B}$ that leads to the minimum of (2.26) we need to know the derivative of this mean square error. Since we are interested in the domain's case, our $\mathbf{R}$ is the k-dimensional row-vector. Derivative is found in the next proposition.

Proposition 2.7 Let $\mathbf{R} : 1 \times k$. Then the first order derivative of the $\mathbb{MSE}(\hat{\boldsymbol{\theta}}_{GR4})$ defined in (2.26) with respect to $\mathbf{B}$ is

$$\begin{aligned} \frac{d\left[\mathbb{MSE}(\hat{\boldsymbol{\theta}}_{GR4})\right]}{d\mathbf{B}} &= \left[\mathbf{R}'(\mathbf{RBR}')^{-1}\mathbf{RM}(\mathbb{I}_k - \mathbf{LR})' \otimes \right. \\ &\quad \left.(\mathbb{I}_k - \mathbf{LR})'\right](\mathbb{I}_{k^2} + \mathbf{K}_{k,k}), \end{aligned} \tag{2.27}$$

where $\mathbf{K}_{k,k}$ is the commutation matrix, $\mathbb{I}_k$ and $\mathbb{I}_{k^2}$ are two identity matrices of the dimensions correspondingly $k \times k$ and $k^2 \times k^2$.

Proof. Let denote $\mathbf{Y} = \mathbb{I}_k - \mathbf{LR}$ and $\mathbf{Z} = \mathbf{YMY}'$. Then, from the derivative formula (1.45) we get

$$\frac{d\left[\mathbb{MSE}(\hat{\boldsymbol{\theta}}_{GR4})\right]}{d\mathbf{B}} = \frac{d\mathbf{Y}}{d\mathbf{B}}\frac{d\mathbf{Z}}{d\mathbf{Y}} = \frac{d(\mathbb{I}_k - \mathbf{LR})}{d\mathbf{B}}\frac{d\left(\mathbf{YMY}'\right)}{d\mathbf{Y}}. \tag{2.28}$$

We will find two derivatives in (2.28) separately, noting that

$$\frac{d(\mathbb{I}_k - \mathbf{LR})}{d\mathbf{B}} = -\frac{d\left(\mathbf{BR}'(\mathbf{RBR}')^{-1}\mathbf{R}\right)}{d\mathbf{B}}. \tag{2.29}$$

Now denoting by $\mathbf{Y}_1 = \mathbf{BR}'$ and $\mathbf{Z}_1 = (\mathbf{RBR}')^{-1}\mathbf{R}$, we can apply formula (1.46) to (2.29),

$$\frac{d(\mathbb{I}_k - \mathbf{LR})}{d\mathbf{B}} = \frac{d\mathbf{Y}_1}{d\mathbf{B}}\frac{d(\mathbf{Y}_1\mathbf{Z}_1)}{d\mathbf{Y}_1}\bigg|_{\mathbf{Z}_1=const} + \frac{d\mathbf{Z}_1}{d\mathbf{B}}\frac{d(\mathbf{Y}_1\mathbf{Z}_1)}{d\mathbf{Z}_1}\bigg|_{\mathbf{Y}_1=const}. \tag{2.30}$$

The first and the second derivative in (2.30) come straightly from formula (1.43),

$$\frac{d\mathbf{Y}_1}{d\mathbf{B}} = \frac{d(\mathbb{I}_k\mathbf{B}\mathbf{R}')}{d\mathbf{B}} = \mathbf{R}' \otimes \mathbb{I}_k, \tag{2.31}$$

$$\left.\frac{d\mathbf{Y}_1\mathbf{Z}_1}{d\mathbf{Y}_1}\right|_{\mathbf{Z}_1=const} = \frac{d(\mathbb{I}_k\mathbf{Y}_1\mathbf{Z}_1)}{d\mathbf{Y}_1} = \mathbf{Z}_1 \otimes \mathbb{I}_k = (\mathbf{R}\mathbf{B}\mathbf{R}')^{-1}\mathbf{R} \otimes \mathbb{I}_k. \tag{2.32}$$

The third derivative from (2.30) is found by applying formula (1.45) twice and then using formula (1.44),

$$\begin{aligned}
\frac{d\mathbf{Z}_1}{d\mathbf{B}} &= \frac{d\left[(\mathbf{R}\mathbf{B}\mathbf{R}')^{-1}\mathbf{R}\right]}{d\mathbf{B}} = \frac{d(\mathbf{R}\mathbf{B}\mathbf{R}')}{d\mathbf{B}}\,\frac{d\left[(\mathbf{R}\mathbf{B}\mathbf{R}')^{-1}\mathbf{R}\right]}{d(\mathbf{R}\mathbf{B}\mathbf{R}')} \\
&= (\mathbf{R}' \otimes \mathbf{R}')\frac{d(\mathbf{R}\mathbf{B}\mathbf{R}')^{-1}}{d(\mathbf{R}\mathbf{B}\mathbf{R}')}\,\frac{d\left[(\mathbf{R}\mathbf{B}\mathbf{R}')^{-1}\mathbf{R}\right]}{d(\mathbf{R}\mathbf{B}\mathbf{R}')^{-1}} \\
&= (\mathbf{R}' \otimes \mathbf{R}')\left[-(\mathbf{R}\mathbf{B}\mathbf{R}')^{-1} \otimes (\mathbf{R}\mathbf{B}'\mathbf{R}')^{-1}\right](\mathbf{R} \otimes 1).
\end{aligned} \tag{2.33}$$

From the property (1.7) of Kronecker and matrix product, the expression (2.33) can be simplified further,

$$\begin{aligned}
\frac{d\mathbf{Z}_1}{d\mathbf{B}} &= \left[-\mathbf{R}'(\mathbf{R}\mathbf{B}\mathbf{R}')^{-1} \otimes \mathbf{R}'(\mathbf{R}\mathbf{B}'\mathbf{R}')^{-1}\right](\mathbf{R} \otimes 1) \\
&= -\mathbf{R}'(\mathbf{R}\mathbf{B}\mathbf{R}')^{-1}\mathbf{R} \otimes \mathbf{R}'(\mathbf{R}\mathbf{B}'\mathbf{R}')^{-1}.
\end{aligned} \tag{2.34}$$

The last derivative of (2.30) comes from (1.43),

$$\left.\frac{d(\mathbf{Y}_1\mathbf{Z}_1)}{d\mathbf{Z}_1}\right|_{\mathbf{Y}_1=const} = \mathbb{I}_k \otimes \mathbf{Y}_1' = \mathbb{I}_k \otimes \mathbf{R}\mathbf{B}'. \tag{2.35}$$

In consideration of properties (1.7) and (1.8) of Kronecker product we have finally for (2.30)

$$\begin{aligned}
\frac{d\mathbf{Y}}{d\mathbf{B}} &= (\mathbf{R}' \otimes \mathbb{I}_k)\left[(\mathbf{R}\mathbf{B}\mathbf{R}')^{-1}\mathbf{R} \otimes \mathbb{I}_k\right] - \\
&\quad \left[\mathbf{R}'(\mathbf{R}\mathbf{B}\mathbf{R}')^{-1}\mathbf{R} \otimes \mathbf{R}'(\mathbf{R}\mathbf{B}'\mathbf{R}')^{-1}\right](\mathbb{I}_k \otimes \mathbf{R}\mathbf{B}') \\
&= \mathbf{R}(\mathbf{R}\mathbf{B}\mathbf{R}')^{-1}\mathbf{R} \otimes \mathbb{I}_k - \mathbf{R}'(\mathbf{R}\mathbf{B}\mathbf{R}')^{-1}\mathbf{R} \otimes \mathbf{R}'(\mathbf{R}\mathbf{B}'\mathbf{R}')^{-1}\mathbf{R}\mathbf{B}' \\
&= \mathbf{R}'(\mathbf{R}\mathbf{B}\mathbf{R}')^{-1}\mathbf{R} \otimes (\mathbb{I}_k - \mathbf{R}'(\mathbf{R}\mathbf{B}'\mathbf{R}')^{-1}\mathbf{R}\mathbf{B}') \\
&= \mathbf{R}'(\mathbf{R}\mathbf{B}\mathbf{R}')^{-1}\mathbf{R} \otimes (\mathbb{I}_k - \mathbf{L}\mathbf{R})'.
\end{aligned} \tag{2.36}$$

The first derivative of (2.28) is found. The second can be found by using formula (1.46). Let denote $\mathbf{Y}_2 = \mathbf{YM}$, $\mathbf{Z}_2 = \mathbf{Y}'$ and $\mathbf{W}_2 = \mathbf{Y}_2\mathbf{Z}_2$. Then,

$$\frac{d(\mathbf{YMY}')}{d\mathbf{Y}} = \frac{d\mathbf{Y}_2}{d\mathbf{Y}} \frac{d(\mathbf{Y}_2\mathbf{Z}_2)}{d\mathbf{Y}_2}\bigg|_{\mathbf{Z}_2=const} + \frac{d\mathbf{Z}_2}{d\mathbf{Y}} \frac{d(\mathbf{Y}_2\mathbf{Z}_2)}{d\mathbf{Z}_2}\bigg|_{\mathbf{Y}_2=const}. \tag{2.37}$$

From (1.43) we have

$$\begin{aligned}
&\frac{d\mathbf{Y}_2}{d\mathbf{Y}} = \frac{d(\mathbb{I}_k\mathbf{YM})}{d\mathbf{Y}} = \mathbf{M} \otimes \mathbb{I}_k, \\
&\frac{d(\mathbf{Y}_2\mathbf{Z}_2)}{d\mathbf{Y}_2}\bigg|_{\mathbf{Z}_2=const} = \mathbf{Z}_2 \otimes \mathbb{I}_k = \mathbf{Y}' \otimes \mathbb{I}_k = (\mathbb{I}_k - \mathbf{LR})' \otimes \mathbb{I}_k, \\
&\frac{d(\mathbf{Y}_2\mathbf{Y}_2)}{d\mathbf{Z}_2}\bigg|_{\mathbf{Y}_2=const} = \mathbb{I}_k \otimes \mathbf{Y}_2' = \mathbb{I}_k \otimes (\mathbf{YM})' = \mathbb{I}_k \otimes \mathbf{M}(\mathbb{I}_k - \mathbf{LR})'.
\end{aligned}$$

For the third derivative of (2.37) the formula (1.47) can be applied,

$$\frac{d\mathbf{Z}_2}{d\mathbf{Y}} = \frac{d\mathbf{Y}'}{d\mathbf{Y}} = \mathbf{K}_{k,k}.$$

The derivative in (2.37) simplifies as following,

$$\begin{aligned}
\frac{d(\mathbf{YMY}')}{d\mathbf{Y}} &= (\mathbf{M} \otimes \mathbb{I}_k)\left[(\mathbb{I}_k - \mathbf{LR})' \otimes \mathbb{I}_k\right] + \mathbf{K}_{k,k}\left[\mathbb{I}_k \otimes \mathbf{M}(\mathbb{I}_k - \mathbf{LR})'\right] \\
&= \mathbf{M}(\mathbb{I}_k - \mathbf{LR})' \otimes \mathbb{I}_k + \mathbf{K}_{k,k}\left[\mathbb{I}_k \otimes \mathbf{M}(\mathbb{I}_k - \mathbf{LR})'\right],
\end{aligned}$$

where property (1.7) was used.

Considering properties (1.18) and (1.21) of the commutation matrix, we have

$$\begin{aligned}
\mathbf{K}_{k,k}\left[\mathbb{I}_k \otimes \mathbf{M}(\mathbb{I}_k - \mathbf{LR})'\right] &= \mathbf{K}_{k,k}\left[\mathbb{I}_k \otimes \mathbf{M}(\mathbb{I}_k - \mathbf{LR})'\right]\mathbf{K}_{k,k}\mathbf{K}_{k,k} \\
&= \left[\mathbf{M}(\mathbb{I}_k - \mathbf{LR})' \otimes \mathbb{I}_k\right]\mathbf{K}_{k,k}.
\end{aligned}$$

Finally, (2.37) can be simplified,

$$\frac{d(\mathbf{YMY}')}{d\mathbf{Y}} = \left[\mathbf{M}(\mathbb{I}_k - \mathbf{LR})' \otimes \mathbb{I}_k\right](\mathbb{I}_{k^2} + \mathbf{K}_{k,k}). \tag{2.38}$$

We have now both components needed for the (2.28), they are in (2.36) and (2.38). Putting them together will lead to the corollary's statement (2.27),

$$
\begin{aligned}
\frac{d\left[\mathbb{M}\mathrm{SE}(\hat{\boldsymbol{\theta}}_{GR4})\right]}{d\mathbf{B}} &= \left[\mathbf{R}'(\mathbf{RBR}')^{-1}\mathbf{R} \otimes (\mathbb{I}_k - \mathbf{LR})'\right] \times \\
&\quad \left[\mathbf{M}(\mathbb{I}_k - \mathbf{LR})' \otimes \mathbb{I}_k\right] (\mathbb{I}_{k^2} + \mathbf{K}_{k,k}) \\
&= \left[\mathbf{R}'(\mathbf{RBR}')^{-1}\mathbf{RM}(\mathbb{I}_k - \mathbf{LR})' \otimes (\mathbb{I}_k - \mathbf{LR})'\right] \times \\
&\quad (\mathbb{I}_{k^2} + \mathbf{K}_{k,k}).
\end{aligned}
$$

$\square$

Remark 2.2 If to take $\mathbf{B} = \mathbf{M}$ in (2.25), then $\hat{\boldsymbol{\theta}}_{GR4} = \hat{\boldsymbol{\theta}}_{GR2}$. The derivative (2.27) in this case is equal to zero, because

$$
\begin{aligned}
&\mathbf{R}'(\mathbf{RMR}')^{-1}\mathbf{RM}\left(\mathbb{I}_k - \mathbf{MR}'(\mathbf{RMR}')^{-1}\mathbf{R}\right)' = \\
&\mathbf{R}'(\mathbf{RMR}')^{-1}\mathbf{RM} - \mathbf{R}'(\mathbf{RMR}')^{-1}\mathbf{RMR}'(\mathbf{RMR}')^{-1}\mathbf{RM} = \mathbf{0}.
\end{aligned}
$$

Thus, $\mathbf{B} = \mathbf{M}$ is the one of the possible minimum points for (2.26).

Chapter 3

Classes of initial estimators

In this chapter the generalized regression (GREG) and the synthetic (SYN) estimators are considered. They are used as the initial estimators in the GR estimator developed in Chapter 2. New results concern the SYN estimator and the mutual relationships between the GREG and SYN estimators. The general bias, variance and mean square error (MSE) expressions for the SYN, and the cross-MSE expression between GREG and SYN are new. They are given both on the population and on the domain levels, whereas for domains, the two different underlying models (the population and the domain model) are assumed. Many interesting relationships are revealed from the comparison of the GREG and SYN estimators. Though, the GREG is known to be asymptotically unbiased, the bias may occur for small sample sizes. Here its approximate bias expression is given from the second-order Taylor expansion.

3.1 Linear estimator

The linear estimator for population total $t_y = \sum_U y_i$ with one study variable $\boldsymbol{y} = (y_1, y_2, ..., y_N)'$ is introduced in Section 1.2.3,

$$\hat{t}_y = \boldsymbol{y}'\breve{\mathbf{I}},$$

where $\breve{\mathbf{I}}$ is the expanded sampling vector defined in (1.68). One-dimensional set-up of the estimation problem is standard in the literature. Here we assume that the totals of many study variables need estimation, which is the case in real surveys, and present the linear estimator and its properties for this multivariate case.

Proposition 3.1 Let $\mathbf{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_k)$ be the $N \times k$ study matrix with ith study variable $\boldsymbol{y}_i = (y_{1i}, y_{2i}, ..., y_{Ni})'$, $i = 1, 2, ..., k$. Then, the linear estimator of the vector of population totals

$$\boldsymbol{t}_{\mathbf{Y}} = \mathbf{Y}'\mathbf{1}_N = (t_{y_1}, t_{y_2}, ..., t_{y_k})'$$

is

$$\hat{\boldsymbol{t}}_{\mathbf{Y}} = \mathbf{Y}'\breve{\mathbf{I}} = (\hat{t}_{y_1}, \hat{t}_{y_2}, ..., \hat{t}_{y_k})'. \tag{3.1}$$

The estimator (3.1) is design-unbiased, it's variance is a $k \times k$ matrix

$$\mathbb{C}\mathrm{ov}(\hat{\boldsymbol{t}}_{\mathbf{Y}}) = \mathbf{Y}'\breve{\boldsymbol{\Delta}}\mathbf{Y}, \tag{3.2}$$

where $\breve{\boldsymbol{\Delta}} = \mathbb{C}\mathrm{ov}(\breve{\mathbf{I}}) = \mathrm{diag}(\mathbb{E}\breve{\mathbf{I}})^{-1}\,\mathbb{C}\mathrm{ov}(\mathbf{I})\,\mathrm{diag}(\mathbb{E}\breve{\mathbf{I}})^{-1}$.

Proof. The unbiasedness of the estimator (3.1) follows from $\mathbb{E}\breve{\mathbf{I}} = \mathbf{1}_N$, where $\mathbf{1}_N$ is the vector of ones.

The variance of the estimator (3.1) follows from (1.59) and (1.57),

$$\begin{aligned}
\mathbb{C}\mathrm{ov}(\hat{\boldsymbol{t}}_{\mathbf{Y}}) &= \mathbb{E}\left[\left(\hat{\boldsymbol{t}}_{\mathbf{Y}} - \mathbb{E}(\hat{\boldsymbol{t}}_{\mathbf{Y}})\right)\left(\hat{\boldsymbol{t}}_{\mathbf{Y}} - \mathbb{E}(\hat{\boldsymbol{t}}_{\mathbf{Y}})'\right]\right. \\
&= \mathbb{E}\left[\left(\mathbf{Y}'\breve{\mathbf{I}} - \mathbf{Y}'\mathbf{1}_N\right)\left(\mathbf{Y}'\breve{\mathbf{I}} - \mathbf{Y}'\mathbf{1}_N\right)'\right] \\
&= \mathbf{Y}'\mathbb{E}\left[\left(\breve{\mathbf{I}} - \mathbf{1}_N\right)\left(\breve{\mathbf{I}} - \mathbf{1}_N\right)'\right]\mathbf{Y} = \mathbf{Y}'\mathbb{C}\mathrm{ov}(\breve{\mathbf{I}})\mathbf{Y}.
\end{aligned}$$

$\square$

Example 3.1 For simple random sampling (SI) design the estimator and its variance take the forms:

$$\hat{\boldsymbol{t}}_{\mathbf{Y}} = f^{-1}\mathbf{Y}'\mathbf{I}, \tag{3.3}$$

$$\mathbb{C}\mathrm{ov}(\hat{\boldsymbol{t}}_{\mathbf{Y}}) = \frac{1-f}{f}\mathbf{Y}'(\mathbb{I} - \mathbf{C})\mathbf{Y}, \tag{3.4}$$

where $f = n/N$ is sampling fraction, and $\mathbf{C} : N \times N$ is the matrix with zeros on the main diagonal and $(N-1)^{-1}$ on other places.

Example 3.2 For the MN design, described in Section 1.2.1, the estimator of the total and its variance are

$$\hat{\boldsymbol{t}}_{\mathbf{Y}} = \mathbf{Y}'\breve{\mathbf{I}} = \frac{1}{n}\mathbf{Y}'(\mathrm{diag}(\boldsymbol{p}))^{-1}\mathbf{I}; \tag{3.5}$$

$$\mathbb{C}\mathrm{ov}(\hat{\boldsymbol{t}}_{\mathbf{Y}}) = \frac{1}{n}\mathbf{Y}'\left[(\mathrm{diag}(\boldsymbol{p}))^{-1} - \mathbf{1}\mathbf{1}'\right]\mathbf{Y}, \tag{3.6}$$

where $\boldsymbol{p} = (p_1, p_2, ..., p_N)'$ is the vector of selection probabilities, and $\mathbf{1}$ is the N-dimensional vector of ones.

□

Assume now that the population U is divided into D non-overlapping and exhaustive domains U_d, and we are interested in the sum of one study variable $\boldsymbol{y} = (y_1, y_2, ..., y_N)'$ separately in each domain, $t^d = \sum_{i \in U_d} y_i$, $d = 1, 2, ..., D$. Estimation of the entire vector of domain totals is a multivariate problem.

We define the domain indicator-matrix as

$$\boldsymbol{\delta} = (\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, ..., \boldsymbol{\delta}_D) = \begin{pmatrix} \delta_{11} & \delta_{12} & ... & \delta_{1D} \\ \delta_{21} & \delta_{22} & ... & \delta_{2D} \\ ... & ... & ... & ... \\ \delta_{N1} & \delta_{N2} & ... & \delta_{ND} \end{pmatrix}, \tag{3.7}$$

where $\boldsymbol{\delta}_d = (\delta_{1d}, \delta_{2d}, ..., \delta_{Nd})'$ identifies the domain U_d, $d = 1, 2, ..., D$ (i.e. $\delta_{id} = 1$, if $i \in U_d$ and 0 otherwise).

The matrix of study variables by domains is

$$\mathbf{Y} = \mathrm{diag}(\boldsymbol{y})\,\boldsymbol{\delta} = \begin{pmatrix} y_1\delta_{11} & y_1\delta_{12} & ... & y_1\delta_{1D} \\ y_2\delta_{21} & y_2\delta_{22} & ... & y_2\delta_{2D} \\ ... & ... & ... & ... \\ y_N\delta_{N1} & y_N\delta_{N2} & ... & y_N\delta_{ND} \end{pmatrix}. \tag{3.8}$$

With this matrix we can estimate all domain totals simultaneously.

Proposition 3.2 The linear estimator for the vector of domain totals, $\boldsymbol{t}^{\mathcal{D}} = (t^1, t^2, ..., t^D)' = \mathbf{Y}'\mathbf{1}_N$, where $t^d = \sum_{i \in U_d} y_i$, is

$$\hat{\boldsymbol{t}}^{\mathcal{D}} = \boldsymbol{\delta}'\mathrm{diag}(\boldsymbol{y})\breve{\mathbf{I}} = \boldsymbol{\delta}'\mathrm{diag}(\breve{\mathbf{I}})\boldsymbol{y}, \tag{3.9}$$

where $\boldsymbol{\delta}$ is the domain indicator-matrix (3.7) and $\boldsymbol{y} = (y_1, y_2, ..., y_N)'$ is the study variable. The estimator $\hat{\boldsymbol{t}}^{\mathcal{D}}$ is unbiased, and its variance is the $D \times D$ matrix

$$\mathbb{C}\mathrm{ov}\left(\hat{\boldsymbol{t}}^{\mathcal{D}}\right) = \boldsymbol{\delta}'\mathrm{diag}(\boldsymbol{y})\,\check{\Delta}\,\mathrm{diag}(\boldsymbol{y})\boldsymbol{\delta}. \tag{3.10}$$

These properties follow straightforwardly from the expression of $\hat{\boldsymbol{t}}^{\mathcal{D}}$, or from the Proposition 3.1, if to use the domain matrix $\mathbf{Y}$ (3.8) in it.

Due to high variability, the linear estimator can be used for considerably large domains, for those having large sample sizes. For smaller domains modelling and auxiliary information helps to construct better estimators. There are model-dependent and model-assisted methods (Rao, 2003; Lehtonen et al., 2003, 2005). The model-dependent methods are very sensitive to the model misspecification, whereas the model-assisted methods are usually not.

In this thesis we consider the model-assisted GREG estimator and the model-dependent SYN estimator. Both estimators use auxiliary information. The estimators are introduced, first for the population, and then for the domain total. We study their design-based properties. Without loss of generality, and for better understanding, we concentrate on estimation of one study variable and one domain at a time. Formulas for the GREG estimator and its variance in multivariate form can be found in Rajaleid (2004).

3.2 Generalized regression estimator

The generalized regression estimator for the population total uses auxiliary information, which allows to increase the precision of estimates. Auxiliary information consists of auxiliary variables and their totals. It may come from registers or for example, from a previous survey. The variables need not be known at the unit level, i.e separately for each object in U. It is enough to know only totals of the auxiliary variables.

We assume that the finite population is a realization of a superpopulation linear model

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{3.11}$$

where $\mathbf{X}$ is $N \times p$ matrix of p auxiliary variables, $\boldsymbol{\beta} : p \times 1$ is an unknown parameter vector and $\boldsymbol{\varepsilon} : N \times 1$ is an error term, a random vector with uncorrelated components, $\mathbb{C}\mathrm{ov}(\boldsymbol{\varepsilon}) = \mathrm{diag}(\sigma_1^2, \sigma_2^2, ..., \sigma_N^2) = \boldsymbol{\Sigma}$.

Fitting the model by minimizing weighted sum of squared residuals

$$(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta})$$

with respect to $\boldsymbol{\beta}$ gives the generalized least squares estimator $\hat{\boldsymbol{\beta}}$ for the parameter $\boldsymbol{\beta}$ at the population level (Särndal et al., 1992, p. 227),

$$\hat{\boldsymbol{\beta}} = \mathbf{B} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}. \tag{3.12}$$

Computing $\mathbf{B}$ requires knowledge of all y_i from the population U, and therefore can not be done in practice, where only sampled y_i are known. For estimating $\mathbf{B}$ from a sample we rewrite it as a product of two sums,

$$\mathbf{T}_{\mathbf{xx}} = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X} \tag{3.13}$$

and

$$\mathbf{T}_{\mathbf{x}y} = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}. \tag{3.14}$$

The respective design-unbiased estimators can be written in matrix form as:

$$\begin{aligned} \hat{\mathbf{T}}_{\mathbf{xx}} &= \mathbf{X}'\boldsymbol{\Sigma}^{-1}\,\mathrm{diag}(\breve{\mathbf{I}})\,\mathbf{X}, & (3.15)\\ \hat{\mathbf{T}}_{\mathbf{x}y} &= \mathbf{X}'\boldsymbol{\Sigma}^{-1}\,\mathrm{diag}(\breve{\mathbf{I}})\,\boldsymbol{y} = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathrm{diag}(\boldsymbol{y})\breve{\mathbf{I}}. & (3.16)\end{aligned}$$

Inserting (3.15) and (3.16) into (3.12) yields the design-consistent estimator for $\mathbf{B} : p \times 1$,

$$\hat{\mathbf{B}} = \hat{\mathbf{T}}_{\mathbf{xx}}^{-1}\hat{\mathbf{T}}_{\mathbf{x}y}. \tag{3.17}$$

With $\hat{\mathbf{B}}$ one can compute fitted values of the study variable for all elements $i \in U$,

$$\hat{\boldsymbol{y}} = \mathbf{X}\hat{\mathbf{B}}. \tag{3.18}$$

We rewrite the study variable-vector with fitted values,

$$\boldsymbol{y} = \hat{\boldsymbol{y}} + \hat{\boldsymbol{r}}, \tag{3.19}$$

where

$$\hat{\boldsymbol{r}} = \boldsymbol{y} - \mathbf{X}\hat{\mathbf{B}} \tag{3.20}$$

is the vector of residuals. From (3.19) we get the expression for the total,

$$\boldsymbol{y}'\mathbf{1} = \hat{\boldsymbol{y}}'\mathbf{1} + \hat{\boldsymbol{r}}'\mathbf{1}.$$

Here $\hat{\boldsymbol{y}}$ is known for all units, but $\hat{\boldsymbol{r}}$ can be calculated only for sampled units. Estimating $\hat{\boldsymbol{r}}'\mathbf{1}$ by $\hat{\boldsymbol{r}}'\breve{\mathbf{I}}$ leads us to the GREG estimator,

$$\hat{t}_{greg} = \hat{\boldsymbol{y}}'\mathbf{1} + \hat{\boldsymbol{r}}'\breve{\mathbf{I}}. \tag{3.21}$$

The GREG estimator is described and studied widely in the literature, e.g. Cassel et al. (1976), Särndal (1980, 1982), Isaki and Fuller (1982) and Särndal et al. (1992).

The GREG estimator (3.21) depends on the sampling design through $\breve{\mathbf{I}}$ and on the model specification through $\hat{\boldsymbol{y}}$.

According to (3.18), the GREG estimator can be written as

$$\hat{\boldsymbol{t}}_{greg} = (\mathbf{X}\hat{\mathbf{B}})'\mathbf{1} + \hat{\boldsymbol{r}}'\breve{\mathbf{I}}. \tag{3.22}$$

Rearranging terms,

$$\begin{aligned} \hat{\boldsymbol{t}}_{greg} &= (\mathbf{X}\hat{\mathbf{B}})'\mathbf{1} + (\boldsymbol{y} - \mathbf{X}\hat{\mathbf{B}})'\breve{\mathbf{I}} \\ &= \boldsymbol{y}'\breve{\mathbf{I}} - \hat{\mathbf{B}}'(\mathbf{X}'\breve{\mathbf{I}} - \mathbf{X}'\mathbf{1}), \end{aligned}$$

gives

$$\hat{\boldsymbol{t}}_{greg} = \hat{t}_y - \hat{\mathbf{B}}'(\hat{\boldsymbol{t}}_{\mathbf{X}} - \boldsymbol{t}_{\mathbf{X}}), \tag{3.23}$$

which is another version for the GREG formula widely used in literature.

As we see from (3.22), the GREG estimator consists of two sums. The sum

$$(\mathbf{X}\hat{\mathbf{B}})'\mathbf{1} = \sum_U \hat{y}_i$$

is the population total of fitted values $\hat{y}_i = \boldsymbol{x}_i'\hat{\mathbf{B}}$, where $\boldsymbol{x}_i'$ is the row-vector of auxiliary matrix $\mathbf{X}$ corresponding to object $i \in U$. Note that only totals of auxiliary variables are needed for this sum, since

$$(\mathbf{X}\hat{\mathbf{B}})'\mathbf{1} = \hat{\mathbf{B}}'(\mathbf{X}'\mathbf{1}).$$

The sum

$$\hat{\boldsymbol{r}}'\breve{\mathbf{I}} = \sum_s \frac{I_i\hat{r}_i}{\mathbb{E}I_i}$$

is an adjustment term, where $\hat{r}_i$ is an element of the residual vector $\hat{\boldsymbol{r}}$. The adjustment term is computed from the sample.

Auxiliary information helps to reduce variance of the GREG estimator. In Särndal *et al.* (1992, p. 239) is claimed, that the GREG estimator is more precise than the linear estimator (1.67) in the variance sense. The GREG estimator is also asymptotically unbiased, with the bias of order n^{-1} (Särndal *et al.*, 1992, p. 238). It is also known, that the bias ratio (the bias divided by the standard error of the estimator) tends to zero as quickly as $n^{-1/2}$ (Estevao and Särndal, 2004).

The variance and bias expressions of the GREG estimator cannot be obtained exactly because of its complex nature. The linearization technique is used in order to get the approximate variance and bias. We derive the Taylor expansion of the GREG of form (3.23) up to the second order terms since approximate bias becomes visible in these terms. Traditionally, only first order Taylor expansion is used. The expression is given in the next proposition. Its long derivation is put in the Appendix A.

Proposition 3.3 The Taylor expansion of the generalized regression estimator (3.23) up to the second order terms is

$$\begin{aligned} \hat{t}_{greg,sec} &= \hat{t}_y - \mathbf{B}'(\hat{\boldsymbol{t}}_{\mathbf{x}} - \boldsymbol{t}_{\mathbf{x}}) - (\hat{\boldsymbol{t}}_{\mathbf{x}} - \boldsymbol{t}_{\mathbf{x}})'\mathbf{T}_{\mathbf{xx}}^{-1}(\hat{\mathbf{T}}_{\mathbf{x}y} - \mathbf{T}_{\mathbf{x}y}) \\ &+ (\hat{\boldsymbol{t}}_{\mathbf{x}} - \boldsymbol{t}_{\mathbf{x}})'\left(\mathbf{B}' \otimes \mathbf{T}_{\mathbf{xx}}^{-1}\right) \operatorname{vec}(\hat{\mathbf{T}}_{\mathbf{xx}} - \mathbf{T}_{\mathbf{xx}}). \end{aligned} \tag{3.24}$$

□

The approximate variance of the GREG estimator is obtained from the linear part of (3.24),

$$\hat{t}_{greg,\,lin} = \hat{t}_y - \mathbf{B}'(\hat{\boldsymbol{t}}_{\mathbf{x}} - \boldsymbol{t}_{\mathbf{x}}),$$

that can be rewritten as

$$\hat{t}_{greg,\,lin} = (\mathbf{XB})'\mathbf{1} + \boldsymbol{r}'\breve{\mathbf{I}}, \tag{3.25}$$

where $\boldsymbol{r} = \boldsymbol{y} - \mathbf{XB}$. This is an expression with the only random vector $\breve{\mathbf{I}}$. The approximate variance can be easily obtained from (3.25) by using property of the covariance matrix, $\mathbb{C}\text{ov}(\boldsymbol{r}'\breve{\mathbf{I}}) = \boldsymbol{r}'\,\mathbb{C}\text{ov}(\breve{\mathbf{I}})\,\boldsymbol{r}$, and is formulated in the next proposition.

Proposition 3.4 The approximate variance of the GREG estimator (3.23) is

$$A\mathbb{V}(\hat{t}_{greg}) = \boldsymbol{r}'\breve{\Delta}\boldsymbol{r}, \tag{3.26}$$

where $\breve{\Delta} = \mathbb{C}\text{ov}(\breve{\mathbf{I}})$ is the covariance matrix of the expanded sampling vector and $\boldsymbol{r} = \boldsymbol{y} - \mathbf{XB}$ is the residual vector of the population model.

$\square$

The Taylor expansion allows to study also bias of the GREG estimator. Although, it is known that GREG estimator is asymptotically unbiased, for small samples the bias may exist. In special cases, the bias of GREG estimator is studied by Lepik (2007). In the next proposition we derive the general bias expression of the GREG estimator (3.23). Note that the linear part of the Taylor expansion gives zero bias.

Proposition 3.5 The approximate bias of $\hat{t}_{greg}$, obtained from the Taylor expansion (3.24), is:

$$\begin{aligned} Ab(\hat{t}_{greg}) &= -\mathrm{vec}'[\mathbb{C}\mathrm{ov}(\hat{\boldsymbol{t}}_{\mathbf{x}}, \hat{\mathbf{T}}_{\mathbf{x}y})]\mathrm{vec}(\mathbf{T}_{\mathbf{xx}}^{-1}) \\ &\quad + \mathrm{vec}'\left[\mathbb{C}\mathrm{ov}\left(\hat{\boldsymbol{t}}_{\mathbf{x}}, \mathrm{vec}(\hat{\mathbf{T}}_{\mathbf{xx}})\right)\right]\mathrm{vec}(\mathbf{B}' \otimes \mathbf{T}_{\mathbf{xx}}^{-1}). \end{aligned} \tag{3.27}$$

Proof. From the definition of the bias and the Taylor expansion of GREG (3.24) we find,

$$\begin{aligned} Ab(\hat{t}_{greg}) &= \mathbb{E}(\hat{t}_{greg,\,sec}) - t_y \\ &= -\mathbb{E}\left[(\hat{\boldsymbol{t}}_{\mathbf{x}} - \boldsymbol{t}_{\mathbf{x}})'\mathbf{T}_{\mathbf{xx}}^{-1}(\hat{\mathbf{T}}_{\mathbf{x}y} - \mathbf{T}_{\mathbf{x}y})\right] \\ &\quad + \mathbb{E}\left[(\hat{\boldsymbol{t}}_{\mathbf{x}} - \boldsymbol{t}_{\mathbf{x}})'(\mathbf{B}' \otimes \mathbf{T}_{\mathbf{xx}}^{-1})\mathrm{vec}(\hat{\mathbf{T}}_{\mathbf{xx}} - \mathbf{T}_{\mathbf{xx}})\right]. \end{aligned} \tag{3.28}$$

Both summands in (3.28) are scalars, so the operation of vectorization can be applied for them. After that the property (1.14) of the vec-operator enables to bring out the middle non-random matrix:

$$\begin{aligned} Ab(\hat{t}_{greg}) &= -\mathbb{E}\left[(\hat{\mathbf{T}}_{\mathbf{x}y} - \mathbf{T}_{\mathbf{x}y})' \otimes (\hat{\boldsymbol{t}}_{\mathbf{x}} - \boldsymbol{t}_{\mathbf{x}})'\right]\mathrm{vec}(\mathbf{T}_{\mathbf{xx}}^{-1}) \\ &\quad + \mathbb{E}\left[\mathrm{vec}'(\hat{\mathbf{T}}_{\mathbf{xx}} - \mathbf{T}_{\mathbf{xx}}) \otimes (\hat{\boldsymbol{t}}_{\mathbf{x}} - \boldsymbol{t}_{\mathbf{x}})'\right]\mathrm{vec}(\mathbf{B}' \otimes \mathbf{T}_{\mathbf{xx}}^{-1}). \end{aligned}$$

The modified form of the property (1.16), $\boldsymbol{b}' \otimes \boldsymbol{a}' = \mathrm{vec}'(\boldsymbol{a}\boldsymbol{b}')$, gives for vectors

$$\begin{aligned} Ab(\hat{t}_{greg}) &= -\mathbb{E}\left[\mathrm{vec}'\{(\hat{\boldsymbol{t}}_{\mathbf{x}} - \boldsymbol{t}_{\mathbf{x}})(\hat{\mathbf{T}}_{\mathbf{x}y} - \mathbf{T}_{\mathbf{x}y})'\}\right]\mathrm{vec}(\mathbf{T}_{\mathbf{xx}}^{-1}) \\ &\quad + \mathbb{E}\left[\mathrm{vec}'\{(\hat{\boldsymbol{t}}_{\mathbf{x}} - \boldsymbol{t}_{\mathbf{x}})\mathrm{vec}'(\hat{\mathbf{T}}_{\mathbf{xx}} - \mathbf{T}_{\mathbf{xx}})\}\right]\mathrm{vec}(\mathbf{B}' \otimes \mathbf{T}_{\mathbf{xx}}^{-1}). \end{aligned}$$

We may exchange the vec- and expectation operations. By definition (1.58) it will lead to the covariances between the vectors of estimators. This proves the bias expression (3.27).

□

The magnitude of the bias (3.27) depends on many ingredients, such as the study and auxiliary variables, the model relating them, and the sampling design. Some special cases of (3.27) have been studied by many authors. For example, Särndal *et al.* (1992, pp. 245-258), Deng and Chhikara (1990), Cochran (1977, pp. 160-162) studied the bias in the case of SI and the ratio model.

3.3 Synthetic estimator

Synthetic estimator (SYN) is defined as a sum of the fitted values under superpopulation linear model (3.11) (Särndal et al., 1992, p. 399). It is closely related to the GREG estimator, in fact it is the first sum in the expression (3.22) of the GREG estimator,

$$\hat{t}_{syn} = (\mathbf{X}\hat{\mathbf{B}})'\mathbf{1} = \hat{\mathbf{B}}'\boldsymbol{t}_{\mathbf{x}}, \tag{3.29}$$

where $\hat{\mathbf{B}} = \hat{\mathbf{T}}_{\mathbf{xx}}^{-1}\hat{\mathbf{T}}_{\mathbf{x}y}$. The estimator (3.29) is non-linear. To study its properties we expand (3.29) into Taylor series. The derivation is given in the Appendix B. Only linear terms are developed. The second order terms would have resulted in too long expressions. Since they are not needed in the present work, they were omitted.

Proposition 3.6 The linear term of the Taylor expansion of the synthetic estimator (3.29) is

$$\hat{t}_{syn,\,lin} = (\mathbf{B}' - \mathbf{B}'\hat{\mathbf{T}}_{\mathbf{xx}}\mathbf{T}_{\mathbf{xx}}^{-1} + \hat{\mathbf{T}}_{\mathbf{x}y}'\mathbf{T}_{\mathbf{xx}}^{-1})\boldsymbol{t}_{\mathbf{x}}. \tag{3.30}$$

□

The next proposition gives an alternative expression for $\hat{t}_{syn,\,lin}$ which displays close relationship with $\hat{t}_{greg,\,lin}$ in (3.25).

Proposition 3.7 The alternative form of the Taylor expansion (3.30) is

$$\hat{t}_{syn,\,lin} = (\mathbf{XB})'\mathbf{1} + \boldsymbol{r}'\mathrm{diag}(\mathbf{S})\breve{\mathbf{I}}, \tag{3.31}$$

where $\boldsymbol{r} = \boldsymbol{y} - \mathbf{XB}$ and

$$\mathbf{S} = \boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{T}_{\mathbf{XX}}^{-1}\boldsymbol{t}_{\mathbf{X}}. \tag{3.32}$$

Proof. We start from (3.30) in the form

$$\hat{t}_{syn,\,lin} = \left[\mathbf{B}' + (\hat{\mathbf{T}}_{\mathbf{X}y} - \hat{\mathbf{T}}_{\mathbf{XX}}\mathbf{B})'\mathbf{T}_{\mathbf{XX}}^{-1}\right]\boldsymbol{t}_{\mathbf{X}}. \tag{3.33}$$

Using formula (3.15) for $\hat{\mathbf{T}}_{\mathbf{XX}}$ and the property (1.15), we may write

$$\hat{\mathbf{T}}_{\mathbf{XX}}\mathbf{B} = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathrm{diag}(\mathbf{XB})\breve{\mathbf{I}}.$$

Inserting this and $\hat{\mathbf{T}}_{\mathbf{X}y}$ from (3.16) into (3.33) the estimator $\hat{t}_{syn,\,lin}$ can be rewritten in the following way,

$$\begin{aligned}
\hat{t}_{syn,\,lin} &= \left[\mathbf{B}' + \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathrm{diag}(\boldsymbol{y})\breve{\mathbf{I}} - \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathrm{diag}(\mathbf{XB})\breve{\mathbf{I}}\right)'\mathbf{T}_{\mathbf{XX}}^{-1}\right]\boldsymbol{t}_{\mathbf{X}} \\
&= \left[\mathbf{B}' + \breve{\mathbf{I}}'\mathrm{diag}(\boldsymbol{y} - \mathbf{XB})\boldsymbol{\Sigma}^{-1}\mathbf{X}\cdot\mathbf{T}_{\mathbf{XX}}^{-1}\right]\boldsymbol{t}_{\mathbf{X}} \\
&= \left[\mathbf{B}' + \breve{\mathbf{I}}'\mathrm{diag}(\boldsymbol{r})\boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{T}_{\mathbf{XX}}^{-1}\right]\boldsymbol{t}_{\mathbf{X}}.
\end{aligned}$$

Denoting $\mathbf{S} = \boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{T}_{\mathbf{XX}}^{-1}\boldsymbol{t}_{\mathbf{X}}$, we have

$$\hat{t}_{syn,\,lin} = \mathbf{B}'\boldsymbol{t}_{\mathbf{X}} + \breve{\mathbf{I}}'\mathrm{diag}(\boldsymbol{r})\mathbf{S}.$$

According to the property (1.15) we have $\mathrm{diag}(\boldsymbol{r})\mathbf{S} = \mathrm{diag}(\mathbf{S})\boldsymbol{r}$. Now

$$\hat{t}_{syn,\,lin} = \mathbf{B}'\boldsymbol{t}_{\mathbf{X}} + \breve{\mathbf{I}}'\mathrm{diag}(\mathbf{S})\boldsymbol{r}.$$

Since $t_{syn,\,lin}$ is a scalar, then it can be transposed, which completes the proof.

$\square$

Proposition 3.8 The second term of $\hat{t}_{syn,\,lin}$ in the alternative expression (3.31), $\boldsymbol{r}'\mathrm{diag}(\mathbf{S})\breve{\mathbf{I}}$, is the unbiased estimator of zero:

$$\mathbb{E}\left(\boldsymbol{r}'\mathrm{diag}(\mathbf{S})\breve{\mathbf{I}}\right) = 0. \tag{3.34}$$

Proof. We have,

$$\mathbb{E}\left(\boldsymbol{r}'\mathrm{diag}(\mathbf{S})\breve{\mathbf{I}}\right) = \boldsymbol{r}'\mathrm{diag}(\mathbf{S})\mathbb{E}\left(\breve{\mathbf{I}}\right) = \boldsymbol{r}'\mathrm{diag}(\mathbf{S})\mathbf{1}.$$

Here, both $\mathbf{S}$ and $\mathbf{1}$ are vectors, thus we may exchange the operation of diagonalization, $\mathrm{diag}(\mathbf{S})\mathbf{1} = \mathrm{diag}(\mathbf{1})\mathbf{S} = \mathbf{S}$. Then from definitions of $\mathbf{S}$ and $\boldsymbol{r}$ we have for the expectation

$$\begin{aligned}\mathbb{E}\left(\boldsymbol{r}'\mathrm{diag}(\mathbf{S})\breve{\mathbf{I}}\right) &= \boldsymbol{r}'\mathbf{S} = \mathbf{S}'\boldsymbol{r} = \left(\boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{T}_{\mathbf{XX}}^{-1}\mathbf{X}'\mathbf{1}\right)'(\boldsymbol{y} - \mathbf{XB})\\ &= \mathbf{1}'\mathbf{X}\mathbf{T}_{\mathbf{XX}}^{-1}\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}\right) - \mathbf{1}'\mathbf{X}\mathbf{T}_{\mathbf{XX}}^{-1}\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)\mathbf{B}\\ &= \mathbf{1}'\mathbf{X}\mathbf{T}_{\mathbf{XX}}^{-1}\mathbf{T}_{\mathbf{X}y} - \mathbf{1}'\mathbf{X}\mathbf{T}_{\mathbf{XX}}^{-1}\mathbf{T}_{\mathbf{XX}}\mathbf{B} = 0,\end{aligned}$$

due to $\mathbf{B} = \mathbf{T}_{\mathbf{XX}}^{-1}\mathbf{T}_{\mathbf{X}y}$.

□

The above proposition explains small variance of the synthetic estimator; the approximate variance comes from an estimator of zero.

In the following proposition we give the formulas for the approximate bias and the variance of the synthetic estimator. They are obtained from the Taylor expansion (3.30) and the alternative formula (3.31).

Proposition 3.9 The approximate bias and variance of $\hat{t}_{syn}$, obtained from the Taylor expansion, are

$$A\boldsymbol{b}(\hat{t}_{syn}) = -\boldsymbol{r}'\mathbf{1} \tag{3.35}$$

$$A\mathbb{V}(\hat{t}_{syn}) = \boldsymbol{r}'\cdot\mathrm{diag}(\mathbf{S})\,\breve{\Delta}\,\mathrm{diag}(\mathbf{S})\cdot\boldsymbol{r}, \tag{3.36}$$

where $\mathbf{S} = \boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{T}_{\mathbf{XX}}^{-1}\boldsymbol{t}_{\mathbf{X}}$, $\boldsymbol{r} = \boldsymbol{y} - \mathbf{XB}$ and $\breve{\boldsymbol{\Delta}} = \mathbb{C}\mathrm{ov}(\breve{\mathbf{I}})$.

Proof. Taking into account the unbiasedness of $\hat{\mathbf{T}}_{\mathbf{XX}}$ and $\hat{\mathbf{T}}_{\mathbf{X}y}$ for the respective totals, the approximate bias comes directly from (3.30) and from the definition of bias,

$$\begin{aligned}A\boldsymbol{b}(\hat{t}_{syn}) &= \mathbb{E}(\hat{t}_{syn,\,lin}) - t_y = \mathbb{E}\left((\mathbf{B}' - \mathbf{B}'\hat{\mathbf{T}}_{\mathbf{XX}}\mathbf{T}_{\mathbf{XX}}^{-1} + \hat{\mathbf{T}}_{\mathbf{X}y}'\mathbf{T}_{\mathbf{XX}}^{-1})\boldsymbol{t}_{\mathbf{X}}\right) - t_y\\ &= \mathbf{B}'\boldsymbol{t}_{\mathbf{X}} - \mathbf{B}'\boldsymbol{t}_{\mathbf{X}} + \mathbf{T}_{\mathbf{X}y}'\mathbf{T}_{\mathbf{XX}}^{-1}\boldsymbol{t}_{\mathbf{X}} - t_y\\ &= \mathbf{B}'\boldsymbol{t}_{\mathbf{X}} - t_y = \mathbf{B}'\mathbf{X}'\mathbf{1} - \boldsymbol{y}'\mathbf{1} = -(\boldsymbol{y} - \mathbf{XB})'\mathbf{1} = -\boldsymbol{r}'\mathbf{1}.\end{aligned}$$

The approximate variance of $\hat{t}_{syn}$ comes from the second term of the formula (3.31):

$$\begin{aligned} A\mathbb{V}(\hat{t}_{syn,}) &= \mathbb{C}\mathrm{ov}(\boldsymbol{r}' \operatorname{diag}(\mathbf{S})\, \breve{\mathbf{I}}) \\ &= \boldsymbol{r}' \operatorname{diag}(\mathbf{S})\, \mathbb{C}\mathrm{ov}(\breve{\mathbf{I}}) \operatorname{diag}(\mathbf{S})\, \boldsymbol{r}, \end{aligned}$$

which is (3.36).

$\square$

Bias of $\hat{t}_{syn}$ can be also approximated by direct comparison of $\hat{t}_{greg}$ and $\hat{t}_{syn}$:

$$\hat{t}_{greg} = \hat{t}_{syn} + \hat{\boldsymbol{r}}'\breve{\mathbf{I}}, \tag{3.37}$$

where $\hat{\boldsymbol{r}} = \boldsymbol{y} - \mathbf{X}\hat{\mathbf{B}}$. From the knowledge, $\mathbb{E}(\hat{t}_{greg}) \approx t_y$, it follows that

$$t_y \approx \mathbb{E}(\hat{t}_{syn}) + \mathbb{E}(\hat{\boldsymbol{r}}'\breve{\mathbf{I}}).$$

From here the approximate bias is

$$\mathbb{E}(\hat{t}_{syn}) - t_y \approx -\mathbb{E}(\hat{\boldsymbol{r}}'\breve{\mathbf{I}}).$$

Further approximation of $\hat{\boldsymbol{r}}$ by $\boldsymbol{r}$, gives the result (3.35).

Remark 3.1 It follows from (3.35) that approximate bias of $\hat{t}_{syn}$ does not depend on the sampling design, but only on the assisting model (through the sum of residuals). However, the approximate variance (3.36) depends on the sampling design (through $\breve{\Delta}$), and on the assisting model as well.

We end this section by an accuracy measure of the synthetic estimator, by the mean square error.

Proposition 3.10 The approximate mean square error of the synthetic estimator $\hat{t}_{syn}$ is

$$A\mathbb{M}\mathrm{SE}(\hat{t}_{syn}) = \boldsymbol{r}' \left(\operatorname{diag}(\mathbf{S})\breve{\boldsymbol{\Delta}} \operatorname{diag}(\mathbf{S}) + \mathbf{1}\mathbf{1}'\right) \boldsymbol{r}. \tag{3.38}$$

Proof. The result (3.38) follows directly from the property (1.64) of the $\mathbb{M}\mathrm{SE}(\cdot)$ operator and the Proposition 3.9,

$$\begin{aligned} A\mathbb{M}\mathrm{SE}(\hat{t}_{syn}) &= A\mathbb{V}(\hat{t}_{syn}) + A\boldsymbol{b}(\hat{t}_{syn}) \left(A\boldsymbol{b}(\hat{t}_{syn})\right)' \\ &= \boldsymbol{r}'\operatorname{diag}(\mathbf{S})\breve{\boldsymbol{\Delta}} \operatorname{diag}(\mathbf{S})\boldsymbol{r} + \boldsymbol{r}'\mathbf{1}\mathbf{1}'\boldsymbol{r} \\ &= \boldsymbol{r}' \left(\operatorname{diag}(\mathbf{S})\breve{\boldsymbol{\Delta}} \operatorname{diag}(\mathbf{S}) + \mathbf{1}\mathbf{1}'\right) \boldsymbol{r}. \end{aligned} \tag{3.39}$$

$\square$

3.4 A property relating GREG and SYN estimators

The expression of the GREG estimator (3.22) tells us that $\hat{t}_{greg} = \hat{t}_{syn}$, if the residual part is zero, $\hat{\boldsymbol{r}}'\breve{\mathbf{I}} = 0$. Below we deal with the respective condition.

According to Särndal et al. (1992, p.231), a sufficient condition for $\hat{\boldsymbol{r}}'\breve{\mathbf{I}} = 0$ is that there exists a constant (not depending on i) column vector $\boldsymbol{\lambda}$ such that for all $i \in U$,

$$\sigma_i^2 = \boldsymbol{x}_i'\boldsymbol{\lambda}, \tag{3.40}$$

where σ_i^2 is the error variance of the model (3.11), a diagonal element of the matrix $\boldsymbol{\Sigma}$. We present this sufficient condition in matrix form.

Proposition 3.11 A sufficient condition for $\hat{\boldsymbol{r}}'\breve{\mathbf{I}} = 0$ is that there exists a vector $\boldsymbol{\lambda}$ such that

$$\boldsymbol{\Sigma} = \mathrm{diag}(\mathbf{X}\boldsymbol{\lambda}). \tag{3.41}$$

Proof. Assume that condition (3.41) holds. We want to show

$$\hat{\boldsymbol{r}}'\breve{\mathbf{I}} = (\boldsymbol{y}' - \hat{\mathbf{B}}'\mathbf{X}')\breve{\mathbf{I}} = 0.$$

First, let us see that $\mathbf{X}'\breve{\mathbf{I}} = \hat{\mathbf{T}}_{\mathbf{xx}}\boldsymbol{\lambda}$. Developing,

$$\mathbf{X}'\breve{\mathbf{I}} = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\breve{\mathbf{I}} = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathrm{diag}(\mathbf{X}\boldsymbol{\lambda})\breve{\mathbf{I}},$$

and noticing that both $\mathbf{X}\boldsymbol{\lambda}$ and $\breve{\mathbf{I}}$ are $N \times 1$ vectors, so that the property (1.15) can be applied for them, we get

$$\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathrm{diag}(\mathbf{X}\boldsymbol{\lambda})\breve{\mathbf{I}} = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathrm{diag}(\breve{\mathbf{I}})\mathbf{X}\boldsymbol{\lambda} = \hat{\mathbf{T}}_{\mathbf{xx}}\boldsymbol{\lambda}.$$

Analogously, we can show that

$$\boldsymbol{y}'\breve{\mathbf{I}} = \boldsymbol{y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\breve{\mathbf{I}} = \boldsymbol{y}'\boldsymbol{\Sigma}^{-1}\mathrm{diag}(\mathbf{X}\boldsymbol{\lambda})\breve{\mathbf{I}} \overset{(1.15)}{=} \hat{\mathbf{T}}_{\mathbf{x}y}'\boldsymbol{\lambda}.$$

So we have:

$$\begin{aligned} \hat{\boldsymbol{r}}'\breve{\mathbf{I}} &= \hat{\mathbf{T}}_{\mathbf{x}y}'\boldsymbol{\lambda} - \hat{\mathbf{B}}'\hat{\mathbf{T}}_{\mathbf{xx}}\boldsymbol{\lambda} \\ &\overset{(3.17)}{=} \hat{\mathbf{T}}_{\mathbf{x}y}'\boldsymbol{\lambda} - \hat{\mathbf{T}}_{\mathbf{x}y}'\boldsymbol{\lambda} = 0. \end{aligned}$$

□

Remark 3.2 The relationship (3.41) holds in many practical situations (Särndal et al., 1992, p. 232, Särndal, 2007). Some situations are listed below, and the validity of 3.41 is shown for them.

(1) Covariance matrix $\mathbf{\Sigma}$ has a simple structure, $\mathbf{\Sigma} = \sigma^2\mathbb{I}$, where σ^2 is a constant, and regression model has an intercept, that is $x_{i1} = 1$ for all $i \in U$. Then the relationship (3.41) holds with $\boldsymbol{\lambda}' = (\sigma^2, 0, ..., 0)$. Really,

$$\mathbf{X}\boldsymbol{\lambda} = \begin{pmatrix} 1 & x_{12} & ... & x_{1p} \\ 1 & x_{22} & ... & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{N2} & ... & x_{NP} \end{pmatrix} \begin{pmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \sigma^2 \\ \sigma^2 \\ \vdots \\ \sigma^2 \end{pmatrix} = \sigma^2\mathbf{1}.$$

(2) σ_i^2 is proportional to one out of the p-auxiliary variables, that is for some auxiliary variable j, $j = 1, 2, ..., p$,

$$\sigma_i^2 \propto x_{ij},\ i \in U.$$

Then $x_{ij} = c\sigma_i^2$ for all i and for some constant c. A suitable $\boldsymbol{\lambda}$ for the relationship (3.41) is $\boldsymbol{\lambda}' = (0, ..., 0, 1/c, 0, ...0)$, where $1/c$ is the j-th element of $\boldsymbol{\lambda}$. For the product $\mathbf{X}\boldsymbol{\lambda}$ we have

$$\mathbf{X}\boldsymbol{\lambda} = \begin{pmatrix} x_{11} & ... & c\sigma_1^2 & ... & x_{1p} \\ x_{21} & ... & c\sigma_2^2 & ... & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{N1} & ... & c\sigma_N^2 & ... & x_{Np} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 1/c \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 \\ \sigma_2^2 \\ \vdots \\ \sigma_N^2 \end{pmatrix}.$$

(3) σ_i^2 is proportional to a linear combination of auxiliary variables,

$$\sigma_i^2 \propto \sum_{j=1}^{p} a_j x_{ij},$$

for all $i \in U$ and some constants $a_1, a_2, ..., a_p$. Then $\sum_{j=1}^{p} a_j x_{ij} = c\sigma_i^2$ for some constant c. It is easy to see that (3.41) holds with $\boldsymbol{\lambda}' = (a_1/c, a_2/c, ..., a_p/c)$:

$$\mathbf{X}\boldsymbol{\lambda} = \begin{pmatrix} x_{11} & x_{12} & ... & x_{1p} \\ x_{21} & x_{22} & ... & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{N1} & x_{N2} & ... & x_{Np} \end{pmatrix} \begin{pmatrix} \frac{a_1}{c} \\ \frac{a_2}{c} \\ \vdots \\ \frac{a_p}{c} \end{pmatrix} = \frac{1}{c} \begin{pmatrix} \sum_{j=1}^{p} x_{1j}a_j \\ \sum_{j=1}^{p} x_{2j}a_j \\ \vdots \\ \sum_{j=1}^{p} x_{Nj}a_j \end{pmatrix} = \begin{pmatrix} \sigma_1^2 \\ \sigma_2^2 \\ \vdots \\ \sigma_N^2 \end{pmatrix}.$$

(4) Auxiliary matrix $\mathbf{X}$ consists of classification variables that are used to code membership in one of the p mutually exclusive and exhaustive groups. Thus, the ith row of $\mathbf{X}$ is

$$\boldsymbol{x}_i' = \boldsymbol{\gamma}_i' = (\gamma_{i1}, \gamma_{i2}, ..., \gamma_{ip}),$$

where for $j = 1, 2, ..., p$, $\gamma_{ij} = 1$ if i belongs to the group j, and $\gamma_{ij} = 0$ if not. Here we assume that objects from the same group j have the same variance σ_j^2, implying that the $N \times N$ matrix $\boldsymbol{\Sigma}$ has $\sigma_1^2, \sigma_2^2, ..., \sigma_p^2$ with certain repetitions on its diagonal. Then a suitable vector $\boldsymbol{\lambda}$ is $\boldsymbol{\lambda}' = (\sigma_1^2, \sigma_2^2, ..., \sigma_p^2)$. Exemplifying, let $N = 5$ and the first, the second, the fourth object belong to the first group with variance σ_1^2; the third and the fifth object belong to the second group with the variance σ_2^2. Then

$$\mathbf{X}\boldsymbol{\lambda} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sigma_1^2 \\ \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 \\ \sigma_1^2 \\ \sigma_2^2 \\ \sigma_1^2 \\ \sigma_2^2 \end{pmatrix}.$$

(5) Combination of a continuous auxiliary variable x_i and classification variable described in (4), i. e. let the ith row of $\mathbf{X}$ be

$$\boldsymbol{x}_i' = (\boldsymbol{\gamma_i}', x_i\boldsymbol{\gamma}_i'),\ i \in U.$$

Let us assume like in the previous case that objects from the same group j have the same variance σ_j^2, $j = 1, 2, ..., p$. A suitable vector $\boldsymbol{\lambda}$ with dimension $2p$ is $\boldsymbol{\lambda}' = (\sigma_1^2, ..., \sigma_p^2, 0, ..., 0)$. Then $\mathbf{X}\boldsymbol{\lambda}$, analogically to (4), is the diagonal of $\boldsymbol{\Sigma}$.

(6) The extensions of the above are available.

$\square$

As an opposite example, we describe below a situation where (3.41) does not hold, and, consequently, the GREG and the synthetic estimators differ from each other. Assume, that $\boldsymbol{\Sigma} = \boldsymbol{\sigma}^2\mathbb{I}$, and auxiliary matrix $\mathbf{X}$ consists of only one variable, different from $\mathbf{1}$, or from any other constant vector. Corresponding regression model has no intercept. Then, it is not possible to find such λ (which is now a scalar) that the condition

$$\mathbf{X}\lambda = \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix} \lambda = \begin{pmatrix} \sigma^2 \\ \sigma^2 \\ \cdots \\ \sigma^2 \end{pmatrix}$$

holds. This implies that the residual term in $\hat{t}_{greg}$ is different from zero. Really, noting that with one auxiliary variable,

$$\hat{\mathrm{T}}_{\mathbf{xx}} = (x_1, x_2, \cdots, x_N)\frac{1}{\sigma^2}\,\mathrm{diag}(\breve{\mathbf{I}}) \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_N \end{pmatrix} = \frac{1}{\sigma^2}\sum_U x_i^2 \breve{I}_i,$$

$$\hat{\mathrm{T}}_{\mathbf{x}y} = \frac{1}{\sigma^2}\sum_U x_i y_i \breve{I}_i,$$

and respectively from (3.17)

$$\hat{\mathrm{B}} = \frac{\sum_U x_i y_i \breve{I}_i}{\sum_U x_i^2 \breve{I}_i}.$$

So, we have for the residual term

$$\hat{\boldsymbol{r}}'\breve{\mathbf{I}} = \left(\boldsymbol{y} - \mathbf{X}\hat{\mathrm{B}}\right)'\breve{\mathbf{I}} = \sum_U y_i \breve{I}_i - \hat{\mathrm{B}}\sum_U x_i \breve{I}_i = \hat{t}_y - \hat{\mathrm{B}}\hat{t}_x.$$

It follows that for $\mathbf{X} \neq$ *constant*, $\hat{\boldsymbol{r}}'\breve{\mathbf{I}} \neq 0$ in each sample, and correspondingly $\hat{t}_{greg} \neq \hat{t}_{syn}$.

3.5 About equality of GREG and SYN estimators

We have seen several situations where $\hat{\boldsymbol{r}}'\breve{\mathbf{I}} = 0$, in which case $\hat{t}_{greg} = \hat{t}_{syn}$. An additional situation is the perfect linear relationship in the population,

$$\boldsymbol{y} = \mathbf{X}\mathbf{B}.$$

Then the vector of residuals in the population,

$$\boldsymbol{r} = \boldsymbol{y} - \mathbf{XB} \tag{3.42}$$

is equal to zero. For the sample estimator of $\mathbf{B}$ we have in this case

$$\begin{aligned}\hat{\mathbf{B}} &= \hat{\mathbf{T}}_{\mathbf{xx}}^{-1}\hat{\mathbf{T}}_{\mathbf{x}y} = \hat{\mathbf{T}}_{\mathbf{xx}}^{-1}\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\,\mathrm{diag}(\breve{\mathbf{I}})\,\boldsymbol{y}\right) = \hat{\mathbf{T}}_{\mathbf{xx}}^{-1}\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\,\mathrm{diag}(\breve{\mathbf{I}})\,\mathbf{X}\right)\mathbf{B} \\ &= \hat{\mathbf{T}}_{\mathbf{xx}}^{-1}\hat{\mathbf{T}}_{\mathbf{xx}}\mathbf{B} = \mathbf{B}.\end{aligned}$$

Consequently, the estimated vector of residuals, $\hat{\boldsymbol{r}} = \boldsymbol{y} - \mathbf{X}\hat{\mathbf{B}}$, is equal to the vector of residuals in the population, $\boldsymbol{r}$, and therefore zero. As a result, $\hat{t}_{greg} = (\mathbf{X}\hat{\mathbf{B}})'\mathbf{1} = (\mathbf{XB})'\mathbf{1} = \boldsymbol{y}'\mathbf{1} = t_y$, which is an error-free estimate.

An interesting non-trivial condition was given in Proposition 3.11: $\hat{t}_{greg} = \hat{t}_{syn}$, if for some constant vector $\boldsymbol{\lambda}$, $\boldsymbol{\Sigma} = \mathrm{diag}(\mathbf{X}\boldsymbol{\lambda})$. From the equality of estimators follows the equality of their Taylor expansions. Consequently, the condition $\boldsymbol{\Sigma} = \mathrm{diag}(\mathbf{X}\boldsymbol{\lambda})$ for some $\boldsymbol{\lambda}$, is also sufficient for $\hat{t}_{greg,\,lin} = \hat{t}_{syn,\,lin}$. On the other hand, comparing expressions (3.25) and (3.31), we see directly that a sufficient condition for their equality is $\mathbf{S} = \mathbf{1}$. The next proposition shows that these conditions are equivalent.

Proposition 3.12 The condition $\mathbf{S} = \mathbf{1}$ holds if and only if there exists a p-vector $\boldsymbol{\lambda}$ such that

$$\boldsymbol{\Sigma} = \mathrm{diag}(\mathbf{X}\boldsymbol{\lambda}).$$

Proof. Suppose $\boldsymbol{\Sigma} = \mathrm{diag}(\mathbf{X}\boldsymbol{\lambda})$. Then, according to the property (1.15)

$$\boldsymbol{\Sigma}\mathbf{1} = \mathrm{diag}(\mathbf{X}\boldsymbol{\lambda})\mathbf{1} = \mathrm{diag}(\mathbf{1})\mathbf{X}\boldsymbol{\lambda} = \mathbf{X}\boldsymbol{\lambda}. \tag{3.43}$$

It is easy to show now that $\mathbf{S} = \mathbf{1}$. For this purpose we rewrite $\mathbf{S} = \boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{T}_{\mathbf{xx}}^{-1}\boldsymbol{t}_{\mathbf{x}}$ in the following way,

$$\mathbf{S} = \boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{T}_{\mathbf{xx}}^{-1}\left(\mathbf{X}'(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma})\mathbf{1}\right).$$

Then, due to (3.43) we have for $\mathbf{S}$

$$\begin{aligned}\mathbf{S} &= \boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{T}_{\mathbf{xx}}^{-1}\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\lambda}\right) \\ &= \boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{T}_{\mathbf{xx}}^{-1}\mathbf{T}_{\mathbf{xx}}\boldsymbol{\lambda} \\ &= \boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\lambda} \\ &= \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\mathbf{1} = \mathbf{1}.\end{aligned}$$

Let us now assume that $\mathbf{S} = \mathbf{1}$ and study conditions for $\mathbf{\Sigma}$. For this purpose we rewrite $\mathbf{S}$ as

$$\begin{aligned} \mathbf{S} &= \mathbf{\Sigma}^{-1}\mathbf{X} \cdot (\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1} \cdot \mathbf{X}'\mathbf{1} \\ &= \mathbf{P}'_{\mathbf{X};\mathbf{\Sigma}^{-1}}\mathbf{1}, \end{aligned}$$

where a projector-matrix $\mathbf{P}_{\mathbf{X};\mathbf{\Sigma}^{-1}}$ is defined in (1.27). From the property of the projector (1.29) and the symmetric matrix $\mathbf{\Sigma}^{-1}$, we note that

$$\mathbf{P}'_{\mathbf{X};\mathbf{\Sigma}^{-1}} = \mathbf{P}'_{\mathbf{X};\mathbf{\Sigma}^{-1}}\mathbf{\Sigma}^{-1}\mathbf{\Sigma} = (\mathbf{\Sigma}^{-1}\mathbf{P}_{\mathbf{X};\mathbf{\Sigma}^{-1}})'\mathbf{\Sigma} = \mathbf{\Sigma}^{-1}\mathbf{P}_{\mathbf{X};\mathbf{\Sigma}^{-1}}\mathbf{\Sigma}.$$

Thus, the assumption $\mathbf{S} = \mathbf{1}$ is equivalent to the condition

$$\mathbf{\Sigma}^{-1}\mathbf{P}_{\mathbf{X};\mathbf{\Sigma}^{-1}}\mathbf{\Sigma}\mathbf{1} = \mathbf{1}. \tag{3.44}$$

Premultiplying both parts of (3.44) by $\mathbf{\Sigma}$ gives

$$\mathbf{P}_{\mathbf{X};\mathbf{\Sigma}^{-1}}\mathbf{\Sigma}\mathbf{1} = \mathbf{\Sigma}\mathbf{1}. \tag{3.45}$$

Now the property (1.36) of the projector $\mathbf{P}_{\mathbf{X};\mathbf{\Sigma}^{-1}}$ can be applied to (3.45); it gives the condition for column spaces, i.e. $\mathcal{C}(\mathbf{\Sigma}\mathbf{1}) \subset \mathcal{C}(\mathbf{X})$. The last condition means (Proposition 1.1) that there exists a matrix $\mathbf{M} : p \times 1$ such that

$$\mathbf{\Sigma}\mathbf{1} = \mathbf{X}\mathbf{M}. \tag{3.46}$$

Since $\mathbf{\Sigma}\mathbf{1} = \boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2, ..., \sigma_N^2)'$, the expression (3.46) is equivalent to

$$\boldsymbol{\sigma}^2 = \mathbf{X}\mathbf{M},$$

or

$$\mathrm{diag}(\boldsymbol{\sigma}^2) = \mathrm{diag}(\mathbf{X}\mathbf{M}).$$

Denoting $\mathbf{M}$ by $\boldsymbol{\lambda}$ completes the proof.

□

Remark 3.3 We note that the sufficient condition, $\mathbf{S} = \mathbf{1}$, is not necessary for the equality of Taylor expansions, $\hat{t}_{greg,\,lin} = \hat{t}_{syn,\,lin}$. The necessary condition is

$$\boldsymbol{r}'\breve{\mathbf{I}} = \boldsymbol{r}'\mathrm{diag}(\mathbf{S})\breve{\mathbf{I}}.$$

This condition does not imply that $\mathbf{S} = \mathbf{1}$.

As it shown in (3.35), the approximate bias of the SYN is $-\boldsymbol{r}'\mathbf{1}$. Generally, $\boldsymbol{r}'\mathbf{1} \neq 0$. But in the cases where $\hat{t}_{greg} = \hat{t}_{syn}$, the approximate bias should be zero. The next proposition shows this.

Proposition 3.13 If the auxiliary matrix $\mathbf{X}$ is such that $\boldsymbol{\Sigma} = \mathrm{diag}(\mathbf{X}\boldsymbol{\lambda})$ for some constant vector $\boldsymbol{\lambda}$, then

$$\boldsymbol{r}'\mathbf{1} = 0. \tag{3.47}$$

Proof. We rewrite the sum of residuals $\boldsymbol{r}'\mathbf{1}$ in the following way:

$$\boldsymbol{r}'\mathbf{1} = \mathbf{1}'\boldsymbol{r} = \mathbf{1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{r}.$$

Using the assumption $\boldsymbol{\Sigma} = \mathrm{diag}(\mathbf{X}\boldsymbol{\lambda})$ and property (1.15) we have for the last expression,

$$\boldsymbol{r}'\mathbf{1} = \mathbf{1}'\mathrm{diag}(\mathbf{X}\boldsymbol{\lambda})\boldsymbol{\Sigma}^{-1}\boldsymbol{r} = [\mathrm{diag}(\mathbf{X}\boldsymbol{\lambda})\mathbf{1}]'\boldsymbol{\Sigma}^{-1}\boldsymbol{r} = [\mathrm{diag}(\mathbf{1})\mathbf{X}\boldsymbol{\lambda}]'\boldsymbol{\Sigma}^{-1}\boldsymbol{r}.$$

Due to $\mathrm{diag}(\mathbf{1}) = \mathbb{I}$, the last expression simplifies further,

$$\boldsymbol{r}'\mathbf{1} = \boldsymbol{\lambda}'\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{r}. \tag{3.48}$$

Now from the definition of residuals, $\boldsymbol{r} = \boldsymbol{y} - \mathbf{XB}$, and formulas (3.13) - (3.14) for $\mathbf{T}_{\mathbf{XX}}$ and $\mathbf{T}_{\mathbf{X}y}$, we get from (3.48)

$$\boldsymbol{r}'\mathbf{1} = \boldsymbol{\lambda}'(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y} - \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{XB}) = \boldsymbol{\lambda}'(\mathbf{T}_{\mathbf{X}y} - \mathbf{T}_{\mathbf{XX}}\mathbf{B}) = 0,$$

because $\mathbf{B} = \mathbf{T}_{\mathbf{XX}}^{-1}\mathbf{T}_{\mathbf{X}y}$.

□

Summarizing results found above we conclude.

1. From the condition $\boldsymbol{\Sigma} = \mathrm{diag}(\mathbf{X}\boldsymbol{\lambda})$ follows the equality $\hat{t}_{greg} = \hat{t}_{syn}$. But the opposite conclusion does not generally hold.

2. From the equality of estimators follows the equality of their Taylor expansion. Thus if $\boldsymbol{\Sigma} = \mathrm{diag}(\mathbf{X}\boldsymbol{\lambda})$, then also $\hat{t}_{greg,lin} = \hat{t}_{syn,lin}$ (but not necessarily in the opposite way).

3. The condition $\boldsymbol{\Sigma} = \mathrm{diag}(\mathbf{X}\boldsymbol{\lambda})$ holds if and only if $\mathbf{S} = \mathbf{1}$. Thus, if $\mathbf{S} = \mathbf{1}$, then $\hat{t}_{greg} = \hat{t}_{syn}$ and $\hat{t}_{greg,lin} = \hat{t}_{syn,lin}$, moreover, the sum of population residuals is zero, $\boldsymbol{r}'\mathbf{1} = 0$.

4. Under the condition $\boldsymbol{\Sigma} = \mathrm{diag}(\mathbf{X}\boldsymbol{\lambda})$, $A\mathbb{MSE}(\hat{t}_{syn}) = A\mathbb{V}(\hat{t}_{syn}) = A\mathbb{V}(\hat{t}_{greg})$.

3.6 GREG and SYN estimators for domains

In Sections 3.2 and 3.3 the GREG and the SYN estimators for the population total were studied. In this section we describe different possibilities to use the GREG and the SYN estimators for the domain total $t^d = \sum_{U_d} y_i$. We recall from (1.72) that domain total can be expressed in terms of vectors as

$$t^d = \boldsymbol{y}_d' \mathbf{1},$$

where $\boldsymbol{y}_d = \operatorname{diag}(\boldsymbol{\delta}_d)\boldsymbol{y}$ and $\boldsymbol{\delta}_d$ is the domain indicator-vector.

The GREG and the SYN estimators for t^d are defined as (Särndal et al., 1992, p. 399):

$$\begin{aligned} \hat{t}^d_{greg} &= \sum_{U_d} \hat{y}_i + \sum_{s_d} \hat{r}_i \breve{I}_i, \\ \hat{t}^d_{syn} &= \sum_{U_d} \hat{y}_i, \end{aligned}$$

which we present in matrix form as

$$\hat{t}^d_{greg} = [\operatorname{diag}(\boldsymbol{\delta}_d)\,\hat{\boldsymbol{y}}]'\mathbf{1} + [\operatorname{diag}(\boldsymbol{\delta}_d)\,\hat{\boldsymbol{r}}]'\,\breve{\mathbf{I}}, \tag{3.49}$$

$$\hat{t}^d_{syn} = [\operatorname{diag}(\boldsymbol{\delta}_d)\,\hat{\boldsymbol{y}}]'\mathbf{1}, \tag{3.50}$$

where $\hat{\boldsymbol{y}} = (\hat{y}_i) : N \times 1$ is the vector of fitted values build on some linear model, and $\hat{\boldsymbol{r}} = (\hat{r}_i) = \boldsymbol{y} - \hat{\boldsymbol{y}} : N \times 1$ is the vector of sample fit residuals.

Formulas (3.49) and (3.50) differ from the corresponding estimators (3.21) and (3.29) of the population total by the involved vector $\boldsymbol{\delta}_d$, which forces summation to go over domain U_d and over sample in this domain $s_d = s \cap U_d$.

3.6.1 Estimators under D- and P-models

The fitted values $\hat{\boldsymbol{y}}$ in the domain estimators (3.49) and (3.50) depend on the model. In domains' case the assisting model can be specified in several ways. Fixed-effects and mixed linear model specifications for $\hat{t}^d_{greg}$ and $\hat{t}^d_{syn}$ are described and thoroughly studied in Lehtonen and Pahkinen (2004, pp. 187-213).

Here we consider two cases: the fixed-effects D-model and the fixed-effects P-model (Lehtonen and Pahkinen, 2004, pp. 200). In the fixed-effects D-model the vector β_d is specified separately for each domain, so that

$$y_i = \boldsymbol{x}_i' \boldsymbol{\beta}_d + \varepsilon_i \tag{3.51}$$

for $i \in U_d$, $d = 1, 2, ..., D$. If the model could be fitted in the whole subpopulation U_d, the generalized least squares estimator of $\boldsymbol{\beta}_d$ would be

$$\mathbf{B}_d = \left(\mathbf{T}^d_{\mathbf{xx}}\right)^{-1} \mathbf{T}^d_{\mathbf{x}y}, \tag{3.52}$$

where

$$\mathbf{T}^d_{\mathbf{xx}} = \mathbf{X}_d' \boldsymbol{\Sigma}^{-1} \mathbf{X}_d, \tag{3.53}$$

$$\mathbf{T}^d_{\mathbf{x}y} = \mathbf{X}_d' \boldsymbol{\Sigma}^{-1} \boldsymbol{y}_d, \tag{3.54}$$

and $\mathbf{X}_d = \text{diag}(\boldsymbol{\delta}_d)\mathbf{X}$ is the auxiliary matrix known in U_d and $\boldsymbol{\Sigma} = \mathbb{C}\mathbf{ov}(\boldsymbol{\varepsilon})$, $\boldsymbol{\varepsilon} = (\varepsilon_i) : N \times 1$.

Design-consistent estimator for $\mathbf{B}_d$ based on the observed data in domain d is

$$\hat{\mathbf{B}}_d = \left(\hat{\mathbf{T}}^d_{\mathbf{xx}}\right)^{-1} \hat{\mathbf{T}}^d_{\mathbf{x}y}, \tag{3.55}$$

with

$$\hat{\mathbf{T}}^d_{\mathbf{xx}} = \mathbf{X}_d' \boldsymbol{\Sigma}^{-1} \text{diag}(\breve{\mathbf{I}}) \mathbf{X}_d, \tag{3.56}$$

$$\hat{\mathbf{T}}^d_{\mathbf{x}y} = \mathbf{X}_d' \boldsymbol{\Sigma}^{-1} \text{diag}(\breve{\mathbf{I}})\, \boldsymbol{y}_d. \tag{3.57}$$

The vector of fitted values for the domain U_d in the first part of (3.49) is

$$\text{diag}(\boldsymbol{\delta}_d)\hat{\boldsymbol{y}} = \text{diag}(\boldsymbol{\delta}_d)\mathbf{X}\hat{\mathbf{B}}_d = \mathbf{X}_d\hat{\mathbf{B}}_d,$$

and the vector of sample fit residuals in the second part of (3.49) is

$$\text{diag}(\boldsymbol{\delta}_d)\hat{\boldsymbol{r}}_d = \text{diag}(\boldsymbol{\delta}_d)\left(\boldsymbol{y} - \mathbf{X}\hat{\mathbf{B}}_d\right) = \text{diag}(\boldsymbol{\delta}_d)\boldsymbol{y} - \text{diag}(\boldsymbol{\delta}_d)\mathbf{X}\hat{\mathbf{B}}_d = \boldsymbol{y}_d - \mathbf{X}_d\hat{\mathbf{B}}_d.$$

Thus, the GREG and the SYN estimators (3.49)-(3.50) build on the assisting D-model (3.51) are

$$\hat{t}^d_{greg\text{-}D} = \left(\mathbf{X}_d\hat{\mathbf{B}}_d\right)' \mathbf{1} + \hat{\boldsymbol{r}}_{d\text{-}D}' \breve{\mathbf{I}}, \tag{3.58}$$

$$\hat{t}^d_{syn\text{-}D} = \left(\mathbf{X}_d\hat{\mathbf{B}}_d\right)' \mathbf{1}, \tag{3.59}$$

where

$$\hat{\boldsymbol{r}}_{d\text{-}D} = \boldsymbol{y}_d - \mathbf{X}_d \hat{\mathbf{B}}_d. \tag{3.60}$$

We use subindices *greg-D* and *syn-D* respectively, in order to emphasize the assisting D-model, standing behind the estimators.

The estimators $\hat{t}^d_{greg\text{-}D}$ and $\hat{t}^d_{syn\text{-}D}$ are called direct estimators, since they use study variable values only from a given domain d (in $\hat{\mathbf{B}}_d$ and $\hat{\boldsymbol{r}}_{d\text{-}D}$). But these estimators require modeling in each domain separately, which may be time consuming. Practicians prefer to use the same model for all domains, especially if there are many domains. Then so called fixed-effects P-models can be used for constructing the GREG and the synthetic estimators for a domain total.

P-model uses an assisting model, defined at the population level,

$$y_i = \boldsymbol{x}_i' \boldsymbol{\beta} + \varepsilon_i, \tag{3.61}$$

for $i \in U$. This model was described in (3.11), and the corresponding sample-based estimator for $\boldsymbol{\beta}$ is given in (3.17). Here, this estimator is given with subindex P to stress its connection to the P-model,

$$\hat{\mathbf{B}}_P = \hat{\mathbf{T}}^{-1}_{\mathbf{xx}} \hat{\mathbf{T}}_{\mathbf{x}y}. \tag{3.62}$$

The estimators of t^d follow from the formulas (3.49) and (3.50). Thus, the GREG and the SYN estimators build on the population model (3.61) are

$$\hat{t}^d_{greg\text{-}P} = (\mathbf{X}_d \hat{\mathbf{B}}_P)' \mathbf{1} + \hat{\boldsymbol{r}}'_{d\text{-}P} \breve{\mathbf{I}} \tag{3.63}$$

$$\hat{t}^d_{syn\text{-}P} = (\mathbf{X}_d \hat{\mathbf{B}}_P)' \mathbf{1}, \tag{3.64}$$

where

$$\hat{\boldsymbol{r}}_{d\text{-}P} = \boldsymbol{y}_d - \mathbf{X}_d \hat{\mathbf{B}}_P. \tag{3.65}$$

Note that the estimators under P-model use values of $\boldsymbol{y}$ also from other domains that d (in $\hat{\mathbf{B}}_P$ and $\hat{\boldsymbol{r}}_{d\text{-}P}$). It is said that they borrow strength from other domains. For this reason, the estimators $\hat{t}^d_{greg\text{-}P}$ and $\hat{t}^d_{syn\text{-}P}$ are called indirect estimators.

3.6.2 Linearized domain estimators

As we see from the definitions of $\hat{t}^d_{greg\text{-}D}$ and $\hat{t}^d_{greg\text{-}P}$, these formulas differ from each other only by the estimator of $\mathbf{B}$ (which is equal to $\hat{\mathbf{B}}_d$ for the assisting D-model and $\hat{\mathbf{B}}_P$ for the P-model). The same holds for the synthetic estimators $\hat{t}^d_{syn\text{-}D}$ and $\hat{t}^d_{syn\text{-}P}$. To shorten further developments we use the unified form of domain estimators, covering both types of assisting models:

$$\hat{t}^d_{greg} = \left(\mathbf{X}_d\hat{\mathbf{B}}\right)'\mathbf{1} + \hat{\boldsymbol{r}}'_d\breve{\mathbf{I}}, \tag{3.66}$$

$$\hat{t}^d_{syn} = \left(\mathbf{X}_d\hat{\mathbf{B}}\right)'\mathbf{1}, \tag{3.67}$$

where

$$\hat{\boldsymbol{r}}_d = \boldsymbol{y}_d - \mathbf{X}_d\hat{\mathbf{B}}, \;\; \hat{\mathbf{B}} = \hat{\mathbf{B}}_1^{-1}\hat{\mathbf{B}}_2. \tag{3.68}$$

For the D-model $\hat{\mathbf{B}}_1 = \hat{\mathbf{T}}^d_{\mathbf{xx}}$ and $\hat{\mathbf{B}}_2 = \hat{\mathbf{T}}^d_{\mathbf{x}y}$, as given in (3.56)-(3.57). For the P-model $\hat{\mathbf{B}}_1 = \hat{\mathbf{T}}_{\mathbf{xx}}$ and $\hat{\mathbf{B}}_2 = \hat{\mathbf{T}}_{\mathbf{x}y}$, as given in (3.15) and (3.16).

In order to study properties of the estimators (3.66) and (3.67) we expand them into Taylor series up to the linear term.

Proposition 3.14 The linear terms of the Taylor expansions of the domain GREG and the SYN estimators are

$$\hat{t}^d_{greg,\,lin} = (\mathbf{X}_d\mathbf{B})'\mathbf{1} + \boldsymbol{r}'_d\breve{\mathbf{I}}, \tag{3.69}$$

$$\hat{t}^d_{syn,\,lin} = \left(\mathbf{B}' - \mathbf{B}'\hat{\mathbf{B}}_1\mathbf{B}_1^{-1} + \hat{\mathbf{B}}'_2\mathbf{B}_1^{-1}\right)\boldsymbol{t}^d_{\mathbf{x}}, \tag{3.70}$$

where $\mathbf{B} = \mathbf{B}_1^{-1}\mathbf{B}_2$, $\boldsymbol{r}_d = \boldsymbol{y}_d - \mathbf{X}_d\mathbf{B}$. The matrices $\mathbf{B}_1$ and $\mathbf{B}_2$ depend on the model: for the D-model $\mathbf{B}_1 = \mathbf{T}^d_{\mathbf{xx}}$ and $\mathbf{B}_2 = \mathbf{T}^d_{\mathbf{x}y}$, given in (3.53)-(3.54); for the P-model $\mathbf{B}_1 = \mathbf{T}_{\mathbf{xx}}$ and $\mathbf{B}_2 = \mathbf{T}_{\mathbf{x}y}$, given in (3.13)-(3.14).

Proof. The formulas for $\hat{t}^d_{greg,\,lin}$ and $\hat{t}^d_{syn,\,lin}$ follow directly from the corresponding estimators of the population total. One simply needs to replace the auxiliary matrix $\mathbf{X}$ by the matrix $\mathbf{X}_d$ in the linearized GREG (3.25) and in the linearized SYN (3.30).

□

Below we give another expression for the linearized synthetic estimator that shows its relationship with the GREG estimator. Contrary to Section 3.3, the expression is now for the domains case.

Proposition 3.15 The alternative form of the Taylor expansion (3.70) of the synthetic estimator for the domain d is

$$\hat{t}^d_{syn,\,lin} = (\mathbf{X}_d\mathbf{B})'\,\mathbf{1} + \boldsymbol{r}'_{d\bullet}\mathrm{diag}(\mathbf{S}_d)\breve{\mathbf{I}}, \tag{3.71}$$

where $\mathbf{S}_d = \boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{B}_1^{-1}\boldsymbol{t}^d_{\mathbf{x}}$, $\mathbf{B}_1$ and $\mathbf{B}$ depend on the model as defined in Proposition 3.14. The residual vector is $\boldsymbol{r}_{d\bullet} = \boldsymbol{y}_d - \mathbf{X}_d\mathbf{B}$ for the assisting D-model, and $\boldsymbol{r}_{d\bullet} = \boldsymbol{y} - \mathbf{X}\mathbf{B}$ for the assisting P-model.

Proof. We rewrite the Taylor expansion (3.70) in the following way:

$$\hat{t}^d_{syn,\,lin} = \mathbf{B}'\boldsymbol{t}^d_{\mathbf{x}} + \left(\hat{\mathbf{B}}_2 - \hat{\mathbf{B}}_1\mathbf{B}\right)'\hat{\mathbf{B}}_1^{-1}\boldsymbol{t}^d_{\mathbf{x}}. \tag{3.72}$$

The first term of (3.72) is equal to that of (3.71), $\mathbf{B}'\boldsymbol{t}^d_{\mathbf{x}} = (\mathbf{X}_d\mathbf{B})'\mathbf{1}$. Elaborating the second term, let us consider the assisting D- and P-models separately.

In the case of the D-model we rewrite $\hat{\mathbf{B}}_1 = \hat{\mathbf{T}}_{\mathbf{xx}}$ in a following way:

$$\begin{aligned} \hat{\mathbf{B}}_1 &= \mathbf{X}'_d\boldsymbol{\Sigma}^{-1}\mathrm{diag}(\breve{\mathbf{I}})\mathbf{X}_d = (\mathrm{diag}(\boldsymbol{\delta}_d)\mathbf{X})'\,\boldsymbol{\Sigma}^{-1}\mathrm{diag}(\breve{\mathbf{I}})\mathrm{diag}(\boldsymbol{\delta}_d)\mathbf{X} \\ &= \mathbf{X}'\mathrm{diag}(\boldsymbol{\delta}_d)\boldsymbol{\Sigma}^{-1}\mathrm{diag}(\breve{\mathbf{I}})\mathrm{diag}(\boldsymbol{\delta}_d)\mathbf{X}. \end{aligned}$$

Three matrices, $\mathrm{diag}(\boldsymbol{\delta}_d)$, $\boldsymbol{\Sigma}^{-1}$, and $\mathrm{diag}(\breve{\mathbf{I}})$, are diagonal, so their order can be exchanged. Since also, $\mathrm{diag}(\boldsymbol{\delta}_d)\cdot\mathrm{diag}(\boldsymbol{\delta}_d) = \mathrm{diag}(\boldsymbol{\delta}_d)$, we get for $\hat{\mathbf{B}}_1$ with the assisting D-model

$$\hat{\mathbf{B}}_1 = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathrm{diag}(\breve{\mathbf{I}})\mathbf{X}_d. \tag{3.73}$$

Analogically, it can be shown that $\hat{\mathbf{B}}_2 = \hat{\mathbf{T}}_{\mathbf{x}y}$ is equal to

$$\hat{\mathbf{B}}_2 = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathrm{diag}(\breve{\mathbf{I}})\boldsymbol{y}_d. \tag{3.74}$$

Consequently, in the case of D-model, we have for the second term in (3.72),

$$\begin{aligned} \left(\hat{\mathbf{B}}_2 - \hat{\mathbf{B}}_1\mathbf{B}\right)'\hat{\mathbf{B}}_1^{-1}\boldsymbol{t}^d_{\mathbf{x}} &= \left[\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathrm{diag}(\breve{\mathbf{I}})\,(\boldsymbol{y}_d - \mathbf{X}_d\mathbf{B})\right]'\mathbf{B}_1^{-1}\boldsymbol{t}^d_{\mathbf{x}} \\ &= \boldsymbol{r}'_{d\bullet}\cdot\mathrm{diag}(\breve{\mathbf{I}})\boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{B}_1^{-1}\boldsymbol{t}^d_{\mathbf{x}}. \end{aligned} \tag{3.75}$$

Denoting $\mathbf{S}_d = \boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{B}_1^{-1}\boldsymbol{t}^d_{\mathbf{x}}$, and using $\mathrm{diag}(\breve{\mathbf{I}})\mathbf{S}_d = \mathrm{diag}(\mathbf{S}_d)\breve{\mathbf{I}}$, we finally get

$$\left(\hat{\mathbf{B}}_2 - \hat{\mathbf{B}}_1\mathbf{B}\right)'\hat{\mathbf{B}}_1^{-1}\boldsymbol{t}^d_{\mathbf{x}} = \boldsymbol{r}'_{d\bullet}\mathrm{diag}(\mathbf{S}_d)\breve{\mathbf{I}},$$

which is equal to the second term in (3.72) for the assisting D-model.

In the case of the assisting P-model the second term of (3.72) is

$$\begin{aligned}\left(\hat{\mathbf{B}}_2 - \hat{\mathbf{B}}_1\mathbf{B}\right)' \hat{\mathbf{B}}_1^{-1}\boldsymbol{t}_{\mathbf{x}}^d &= \left(\hat{\mathbf{T}}_{\mathbf{x}y} - \mathbf{T}_{\mathbf{xx}}\mathbf{B}\right)' \mathbf{B}_1^{-1}\boldsymbol{t}_{\mathbf{x}}^d \\ &= \left[\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathrm{diag}(\breve{\mathbf{I}})(\boldsymbol{y} - \mathbf{XB})\right]' \mathbf{B}_1^{-1}\boldsymbol{t}_{\mathbf{x}}^d \\ &= \boldsymbol{r}_{d\bullet} \cdot \mathrm{diag}(\breve{\mathbf{I}})\boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{B}_1^{-1}\boldsymbol{t}_{\mathbf{x}}^d.\end{aligned}$$

Since $\boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{B}_1^{-1}\boldsymbol{t}_{\mathbf{x}}^d = \mathbf{S}_d$, we have shown that

$$\left(\hat{\mathbf{B}}_2 - \hat{\mathbf{B}}_1\mathbf{B}\right)' \hat{\mathbf{B}}_1^{-1}\boldsymbol{t}_{\mathbf{x}}^d = \boldsymbol{r}_{d\bullet} \cdot \mathrm{diag}(\breve{\mathbf{I}})\mathbf{S}_d,$$

which is equal to the second term in (3.72) in the case of the assisting P-model.

□

Remark 3.4 Comparing $\hat{t}^d_{greg,\,lin}$ and $\hat{t}^d_{syn,\,lin}$, respectively in (3.69) and (3.71), we see that their first terms coincide under both the D- and the P-model, but the residual terms differ. The term $\boldsymbol{r}'_{d\bullet}\mathrm{diag}(\mathbf{S}_d)\breve{\mathbf{I}}$ differs for the D- and P-models by the term $\mathbf{S}_d$ through the matrix $\mathbf{B}_1$ and also by the residual vector $\boldsymbol{r}_{d\bullet}$. Under the D-model $\boldsymbol{r}_{d\bullet} = \boldsymbol{r}_d$, i. e. the estimators $\hat{t}^d_{greg,\,lin}$ and $\hat{t}^d_{syn,\,lin}$ use the same residual vectors. Under the P-model $\boldsymbol{r}_{d\bullet} = \boldsymbol{y} - \mathbf{XB} \neq \boldsymbol{r}_d$, showing that the linearized GREG and SYN estimators use different residual vectors. The last feature of the GREG and SYN estimators occurs only with domain estimation; it was not present when estimating population totals.

3.6.3 Properties of the domain estimators

The next proposition shows an interesting feature of the linearized synthetic domain estimator. This feature affects both the variance and the bias of the estimator.

Proposition 3.16 The second term of (3.71), $\boldsymbol{r}'_{d\bullet}\mathrm{diag}(\mathbf{S}_d)\breve{\mathbf{I}}$, is zero on the average,

$$\mathbb{E}\left(\boldsymbol{r}'_{d\bullet}\mathrm{diag}(\mathbf{S}_d)\breve{\mathbf{I}}\right) = 0. \tag{3.76}$$

The equality (3.76) is valid for both the D- and the P-models.

Proof. We have,

$$\mathbb{E}\left(\boldsymbol{r}'_{d\bullet}\mathrm{diag}(\mathbf{S}_d)\breve{\mathbf{I}}\right) = \boldsymbol{r}'_{d\bullet}\mathrm{diag}(\mathbf{S}_d)\mathbf{1} = \boldsymbol{r}'_{d\bullet}\mathbf{S}_d = \mathbf{S}'_d\boldsymbol{r}_{d\bullet}.$$

Now, from definitions of $\mathbf{S}_d$ and $\boldsymbol{r}_{d\bullet}$ (Proposition 3.15) we get in the case of the D-model

$$\begin{aligned}\mathbf{S}'_d\boldsymbol{r}_{d\bullet} &= \left[\boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{T}^d_{\mathbf{xx}})^{-1}\mathbf{X}'_d\mathbf{1}\right]'(\boldsymbol{y}_d - \mathbf{X}_d\mathbf{B}) \\ &= \mathbf{1}'\mathbf{X}_d(\mathbf{T}^d_{\mathbf{xx}})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_d - \mathbf{1}'\mathbf{X}_d(\mathbf{T}^d_{\mathbf{xx}})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}_d\mathbf{B}.\end{aligned}$$

Since $\mathrm{diag}(\boldsymbol{\delta}_d)\boldsymbol{y}_d = \boldsymbol{y}_d$, therefore

$$\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_d = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathrm{diag}(\boldsymbol{\delta}_d)\boldsymbol{y}_d = \mathbf{X}'\mathrm{diag}(\boldsymbol{\delta}_d)\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_d = \mathbf{X}'_d\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_d = \mathbf{T}^d_{\mathbf{x}y}.$$

Analogically, it can be shown that $\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}_d = \mathbf{X}'_d\boldsymbol{\Sigma}^{-1}\mathbf{X}_d = \mathbf{T}^d_{\mathbf{xx}}$. So, we have

$$\mathbf{S}'_d\boldsymbol{r}_{d\bullet} = \mathbf{1}\mathbf{X}_d(\mathbf{T}^d_{\mathbf{xx}})^{-1}\mathbf{T}^d_{\mathbf{x}y} - \mathbf{1}\mathbf{X}_d(\mathbf{T}^d_{\mathbf{xx}})^{-1}\mathbf{T}^d_{\mathbf{xx}}\mathbf{B} = \mathbf{1}'\mathbf{X}_d\mathbf{B} - \mathbf{1}'\mathbf{X}_d\mathbf{B} = 0,$$

and, consequently, (3.76) is proved for the D-model.

In the case of the assisting P-model

$$\begin{aligned}\mathbf{S}'_d\boldsymbol{r}_{d\bullet} &= \left[\boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{T}^{-1}_{\mathbf{xx}}\mathbf{X}'_d\mathbf{1}\right]'(\boldsymbol{y} - \mathbf{X}\mathbf{B}) \\ &= \mathbf{1}'\mathbf{X}_d\mathbf{T}^{-1}_{\mathbf{xx}}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y} - \mathbf{1}'\mathbf{X}_d\mathbf{T}^{-1}_{\mathbf{xx}}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{B} \\ &= \mathbf{1}'\mathbf{X}_d\mathbf{T}^{-1}_{\mathbf{xx}}\mathbf{T}_{\mathbf{x}y} - \mathbf{1}'\mathbf{X}_d\mathbf{T}^{-1}_{\mathbf{xx}}\mathbf{T}_{\mathbf{xx}}\mathbf{B}.\end{aligned}$$

Since, for the P-model $\mathbf{B} = \mathbf{T}^{-1}_{\mathbf{xx}}\mathbf{T}_{\mathbf{x}y}$, we get $\mathbf{S}'_d\boldsymbol{r}_{d\bullet} = 0$, and the proof is completed.

□

In the following proposition we give formulas for the approximate bias and variance of domain estimators $\hat{t}^d_{greg}$ and $\hat{t}^d_{syn}$. They are obtained from the corresponding Taylor expansions.

Proposition 3.17 The approximate bias and variance of $\hat{t}^d_{greg}$ and $\hat{t}^d_{syn}$ are

$$A\boldsymbol{b}(\hat{t}^d_{greg}) = 0, \tag{3.77}$$

$$A\mathbb{V}(\hat{t}^d_{greg}) = \boldsymbol{r}'_d\breve{\boldsymbol{\Delta}}\boldsymbol{r}_d, \tag{3.78}$$

$$A\boldsymbol{b}(\hat{t}^d_{syn}) = -\boldsymbol{r}'_d\mathbf{1}, \tag{3.79}$$

$$A\mathbb{V}(\hat{t}^d_{syn}) = \boldsymbol{r}'_{d\bullet}\mathrm{diag}(\mathbf{S}_d)\breve{\boldsymbol{\Delta}}\mathrm{diag}(\mathbf{S}_d)\boldsymbol{r}_{d\bullet}, \tag{3.80}$$

where $\breve{\boldsymbol{\Delta}} = \mathbb{C}\mathrm{ov}(\breve{\mathbf{I}})$, $\mathbf{S}_d$ and $\boldsymbol{r}_{d\bullet}$ are given in Proposition 3.15, $\boldsymbol{r}_d = \boldsymbol{y}_d - \mathbf{X}_d\mathbf{B}$ with $\mathbf{B}$ depending on the model.

Proof. The approximate bias of the $\hat{t}^d_{greg}$ comes from the linearized domain GREG estimator (3.69),

$$\begin{aligned} A\boldsymbol{b}(\hat{t}^d_{greg}) &= \mathbb{E}(\hat{t}^d_{greg,\,lin}) - t^d_y = \Big[(\mathbf{X}_d\mathbf{B})'\mathbf{1} + \boldsymbol{r}'_d\mathbb{E}\breve{\mathbf{I}}\Big] - \boldsymbol{y}'_d\mathbf{1} \\ &= (\mathbf{X}_d\mathbf{B})'\mathbf{1} + (\boldsymbol{y}_d - \mathbf{X}_d\mathbf{B})'\mathbf{1} - \boldsymbol{y}'_d\mathbf{1} = 0. \end{aligned}$$

The approximate variance of $\hat{t}^d_{greg}$ from (3.69) is,

$$A\mathbb{V}(\hat{t}^d_{greg}) = \mathbb{V}(\hat{t}^d_{greg,\,lin}) = \mathbb{V}(\boldsymbol{r}'_d\breve{\mathbf{I}}) = \boldsymbol{r}'_d\mathbb{C}\mathrm{ov}(\breve{\mathbf{I}})\boldsymbol{r}_d.$$

The approximate bias of $\hat{t}^d_{syn}$ is obtained from (3.70),

$$\begin{aligned} A\boldsymbol{b}(\hat{t}^d_{syn}) &= \mathbb{E}(\hat{t}^d_{syn,lin}) - t^d_y = \mathbf{B}'\boldsymbol{t}^d_{\mathbf{x}} - \mathbf{B}'\mathbf{B}_1\mathbf{B}_1^{-1}\boldsymbol{t}^d_{\mathbf{x}} + \mathbf{B}'_2\mathbf{B}_1^{-1}\boldsymbol{t}^d_{\mathbf{x}} - t^d_y \\ &= \mathbf{B}'\boldsymbol{t}^d_{\mathbf{x}} - t^d_y = \mathbf{B}'\mathbf{X}'_d\mathbf{1} - \boldsymbol{y}'_d\mathbf{1} = -(\boldsymbol{y}_d - \mathbf{X}_d\mathbf{B})'\mathbf{1}, \end{aligned}$$

which is equal to (3.79).

The approximate variance of $\hat{t}^d_{syn}$ comes from its alternative Taylor expansion (3.71),

$$A\mathbb{V}(\hat{t}^d_{syn}) = \mathbb{V}(\hat{t}^d_{syn,\,lin}) = \mathbb{V}(\boldsymbol{r}'_{d\bullet}\mathrm{diag}(\mathbf{S}_d)\breve{\mathbf{I}}) = \boldsymbol{r}'_{d\bullet}\mathrm{diag}(\mathbf{S}_d)\mathbb{C}\mathrm{ov}(\breve{\mathbf{I}})\mathrm{diag}(\mathbf{S}_d)\boldsymbol{r}_{d\bullet}.$$

□

3.6.4 On the equality of domain GREG and SYN estimators

In Sections 3.2 and 3.3 we showed that if there exists a vector $\boldsymbol{\lambda}$ such $\boldsymbol{\Sigma} = \mathrm{diag}(\mathbf{X}\boldsymbol{\lambda})$, then $\hat{t}_{greg} = \hat{t}_{syn}$. Moreover, then $\hat{t}_{greg,\,lin} = \hat{t}_{syn,\,lin}$ and $\mathbf{S} = \boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{T}^{-1}_{\mathbf{XX}}\boldsymbol{t}^d_{\mathbf{X}} = \mathbf{1}$. Below we consider conditions for the equality of $\hat{t}^d_{greg}$ and $\hat{t}^d_{syn}$. The model plays an important role here.

Proposition 3.18 If there exists such a vector $\boldsymbol{\lambda}$ that the condition $\boldsymbol{\Sigma} = \mathrm{diag}(\mathbf{X}\boldsymbol{\lambda})$ holds, then in (3.58)-(3.60)

$$\hat{\boldsymbol{r}}'_{d\text{-}D}\breve{\mathbf{I}} = 0, \tag{3.81}$$

and

$$\hat{t}^d_{greg\text{-}D} = \hat{t}^d_{syn\text{-}D}, \tag{3.82}$$

where $\hat{\boldsymbol{r}}_{d\text{-}D}$ is given in (3.60). At the same time in (3.63)-(3.64), in general,

$$\hat{\boldsymbol{r}}'_{d\text{-}P}\breve{\mathbf{I}} \neq 0,$$
$$\hat{t}^d_{greg\text{-}P} \neq \hat{t}^d_{syn\text{-}P},$$

where $\hat{\boldsymbol{r}}_{d\text{-}P}$ is given in (3.65).

Proof. We assume that there exists such $\boldsymbol{\lambda}$ that $\boldsymbol{\Sigma} = \text{diag}(\mathbf{X}\boldsymbol{\lambda})$, and show that (3.81) then holds. For D-model we have

$$\hat{\boldsymbol{r}}'_{d\text{-}D}\breve{\mathbf{I}} = (\boldsymbol{y}_d - \mathbf{X}_d\hat{\mathbf{B}}_d)'\breve{\mathbf{I}}.$$

Now, from equalities $\mathbf{X}_d = \mathbf{X}_d\text{diag}(\boldsymbol{\delta}_d)$ and $\boldsymbol{\Sigma}\breve{\mathbf{I}} = \text{diag}(\mathbf{X}\boldsymbol{\lambda})\breve{\mathbf{I}} = \text{diag}(\breve{\mathbf{I}})\mathbf{X}\boldsymbol{\lambda}$ we have for $\mathbf{X}'_d\breve{\mathbf{I}}$

$$\begin{aligned}\mathbf{X}'_d\breve{\mathbf{I}} &= \mathbf{X}'_d\text{diag}(\boldsymbol{\delta}_d)\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\breve{\mathbf{I}} \\ &= \mathbf{X}'_d\text{diag}(\boldsymbol{\delta}_d)\boldsymbol{\Sigma}^{-1}\text{diag}(\breve{\mathbf{I}})\mathbf{X}\boldsymbol{\lambda} \\ &= \mathbf{X}'_d\boldsymbol{\Sigma}^{-1}\text{diag}(\breve{\mathbf{I}})\mathbf{X}_d\boldsymbol{\lambda} \\ &= \hat{\mathbf{T}}^d_{\mathbf{xx}}\boldsymbol{\lambda}.\end{aligned}$$

Analogically, it can be shown that

$$\boldsymbol{y}'_d\breve{\mathbf{I}} = (\hat{\mathbf{T}}^d_{\mathbf{x}y})'\boldsymbol{\lambda}.$$

Finally, we have for $\hat{\boldsymbol{r}}'_{d\text{-}D}\breve{\mathbf{I}}$,

$$\begin{aligned}\hat{\boldsymbol{r}}'_{d\text{-}D}\breve{\mathbf{I}} &= \boldsymbol{y}'_d\breve{\mathbf{I}} - \hat{\mathbf{B}}'_d\mathbf{X}'_d\breve{\mathbf{I}} = (\hat{\mathbf{T}}^d_{\mathbf{x}y})'\boldsymbol{\lambda} - \hat{\mathbf{B}}'_d\hat{\mathbf{T}}^d_{\mathbf{xx}}\boldsymbol{\lambda} \\ &= \left[(\hat{\mathbf{T}}^d_{\mathbf{x}y})' - (\hat{\mathbf{T}}^d_{\mathbf{x}y})'(\hat{\mathbf{T}}^d_{\mathbf{xx}})^{-1}\hat{\mathbf{T}}^d_{\mathbf{xx}}\right]\boldsymbol{\lambda} = 0. \qquad (3.83)\end{aligned}$$

For the assisting P-model we express $\hat{\boldsymbol{r}}'_{d\text{-}P}\breve{\mathbf{I}}$ as

$$\begin{aligned}\hat{\boldsymbol{r}}'_{d\text{-}P}\breve{\mathbf{I}} &= (\boldsymbol{y}_d - \mathbf{X}_d\hat{\mathbf{B}})'\breve{\mathbf{I}} = (\text{diag}(\boldsymbol{\delta}_d)\boldsymbol{y} - \text{diag}(\boldsymbol{\delta}_d)\mathbf{X}\hat{\mathbf{B}})'\breve{\mathbf{I}} \\ &= (\boldsymbol{y} - \mathbf{X}\hat{\mathbf{B}})'\text{diag}(\boldsymbol{\delta}_d)\breve{\mathbf{I}} = \hat{\boldsymbol{r}}'\text{diag}(\boldsymbol{\delta}_d)\breve{\mathbf{I}}.\end{aligned}$$

From Proposition 3.11 we know that if $\boldsymbol{\Sigma} = \text{diag}(\mathbf{X}\boldsymbol{\lambda})$ holds, then $\hat{\boldsymbol{r}}'\breve{\mathbf{I}} = 0$. But it does not imply that $\hat{\boldsymbol{r}}'\text{diag}(\boldsymbol{\delta}_d)\breve{\mathbf{I}} = 0$, in general.

□

Remark 3.5 We see that $\hat{\boldsymbol{r}}'\text{diag}(\boldsymbol{\delta}_d)\breve{\mathbf{I}} = 0$, if no elements from domain d belong to the sample. In this case for that particular sample, however, the domain d total can be estimated by $\hat{t}^d_{greg\text{-}P} = \hat{t}^d_{syn\text{-}P} = (\mathbf{X}_d\hat{\mathbf{B}}_P)'\mathbf{1}$, since $\hat{\mathbf{B}}_P = \hat{\mathbf{T}}^{-1}_{\mathbf{xx}}\hat{\mathbf{T}}_{\mathbf{x}y}$ is computed from the whole sample s, thus, available. On the contrary, the empty domain can not be estimated under the D-model.

3.6.5 On the bias of SYN-P estimator

Remark 3.6 As it follows from (3.82) and (3.77), the estimator $\hat{t}^d_{syn\text{-}D}$ is approximately unbiased for t^d_y, if the assumption $\mathbf{\Sigma} = \text{diag}(\mathbf{X}\boldsymbol{\lambda})$ holds. However, in the case of the assisting P-model, the estimator $\hat{t}^d_{syn\text{-}P}$ has a bias, even if $\mathbf{\Sigma} = \text{diag}(\mathbf{X}\boldsymbol{\lambda})$.

The following proposition gives an expression for this bias.

Proposition 3.19 If there exists such a constant vector $\boldsymbol{\lambda}$, that $\mathbf{\Sigma} = \text{diag}(\mathbf{X}\boldsymbol{\lambda})$ holds, then the approximate bias of $\hat{t}^d_{syn\text{-}P}$ can be expressed as

$$A\boldsymbol{b}(\hat{t}^d_{syn\text{-}P}) = -(\mathbf{B}_d - \mathbf{B})'\boldsymbol{t}^d_{\mathbf{x}}, \tag{3.84}$$

where $\mathbf{B}_d = (\mathbf{T}^d_{\mathbf{xx}})^{-1}\mathbf{T}^d_{\mathbf{x}y}$ and $\mathbf{B} = (\mathbf{T}_{\mathbf{xx}})^{-1}\mathbf{T}_{\mathbf{x}y}$.

Proof. As it follows from (3.79), the synthetic estimator for the domain d has an approximate bias $-\boldsymbol{r}'_d\mathbf{1}$. For the P-model this bias takes a form,

$$A\boldsymbol{b}(\hat{t}^d_{syn\text{-}P}) = -(\boldsymbol{y}_d - \mathbf{X}_d\mathbf{B})'\mathbf{1}.$$

We express it in the following way,

$$A\boldsymbol{b}(\hat{t}^d_{syn\text{-}P}) = -\left[\boldsymbol{y}_d - \mathbf{X}_d\mathbf{B}_d + \mathbf{X}_d(\mathbf{B}_d - \mathbf{B})\right]'\mathbf{1} = -\left[\boldsymbol{r}_{d\text{-}D} + \mathbf{X}_d(\mathbf{B}_d - \mathbf{B})\right]'\mathbf{1}.$$

Now, let us assume that $\mathbf{\Sigma} = \text{diag}(\mathbf{X}\boldsymbol{\lambda})$ holds for some $\boldsymbol{\lambda}$. We show that in this case $\boldsymbol{r}'_{d\text{-}D}\mathbf{1} = 0$:

$$\begin{aligned}
\boldsymbol{r}'_{d\text{-}D}\mathbf{1} &= \mathbf{1}'\boldsymbol{r}_{d\text{-}D} = \mathbf{1}'\mathbf{\Sigma}\mathbf{\Sigma}^{-1}\boldsymbol{r}_{d\text{-}D} = \mathbf{1}'\text{diag}(\mathbf{X}\boldsymbol{\lambda})\mathbf{\Sigma}^{-1}\boldsymbol{r}_{d\text{-}D} \\
&= \left[\text{diag}(\mathbf{1})(\mathbf{X}\boldsymbol{\lambda})\right]'\mathbf{\Sigma}^{-1}\boldsymbol{r}_{d\text{-}D} = \boldsymbol{\lambda}'\mathbf{X}'\mathbf{\Sigma}^{-1}\boldsymbol{r}_{d\text{-}D} \\
&= \boldsymbol{\lambda}'\mathbf{X}'\mathbf{\Sigma}^{-1}(\boldsymbol{y}_d - \mathbf{X}_d\mathbf{B}_d).
\end{aligned}$$

Since $\mathbf{X}'\mathbf{\Sigma}^{-1}\boldsymbol{y}_d = \mathbf{X}'_d\mathbf{\Sigma}^{-1}\boldsymbol{y}_d$ and $\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X}_d = \mathbf{X}'_d\mathbf{\Sigma}^{-1}\mathbf{X}_d$, it follows

$$\boldsymbol{r}'_{d\text{-}D}\mathbf{1} = \boldsymbol{\lambda}'(\mathbf{X}'_d\mathbf{\Sigma}^{-1}\boldsymbol{y}_d - \mathbf{X}'_d\mathbf{\Sigma}^{-1}\mathbf{X}_d\mathbf{B}_d) = \boldsymbol{\lambda}'(\mathbf{T}^d_{\mathbf{x}y} - \mathbf{T}^d_{\mathbf{xx}}\mathbf{B}_d) = 0. \tag{3.85}$$

Thus, the approximate bias of $\hat{t}^d_{syn\text{-}P}$ is

$$Ab(\hat{t}^{d}_{syn\text{-}P}) = -\left[0 + \mathbf{X}_d(\mathbf{B}_d - \mathbf{B})\right]' \mathbf{1} = -(\mathbf{B}_d - \mathbf{B})'\mathbf{X}'_d\mathbf{1},$$

which is equal to (3.84).

$\square$

We see from (3.84) that for a given domain, the bias is negligible if the domain parameter-vector $\mathbf{B}_d$ is close to the population parameter $\mathbf{B}$, i.e. assisting model in the U_d is approximately the same as in the whole population. If this condition does not hold, then a substantial bias can be encountered. The same result for the P-model with one auxiliary variable and without intercept can be found in Lehtonen and Pahkinen (2004), p. 204.

3.7 Dependence characteristics of estimators

The ultimate interest of this thesis is on the behavior of the general restriction estimator, and on its properties. This estimator is built on the initial estimators that can be e. g. the linear, the GREG and the SYN estimators. In Chapter 2 three GR estimators were introduced that satisfy linear restrictions. All of them require knowledge of the matrix $\mathbf{M}$, which is the mean square error matrix of the vector of initial estimators. The main diagonal of $\mathbf{M}$ consists of the one-dimensional mean square errors of the initial estimators, they can be found in Sections 2.1-2.3. But elements outside of the main diagonal are so called cross-mean square errors between the estimators. Their expressions are given in this section. An one-dimensional study variable $\boldsymbol{y} = (y_1, y_2, ..., y_N)'$ is considered.

3.7.1 Approximate cross-mean square errors of estimators

The following proposition gives approximate MSEs of three estimators of the population total. They are obtained by using expressions of the linear estimator $\hat{t}_y = \boldsymbol{y}'\breve{\mathbf{I}}$, and of the linearized GREG and SYN estimators $\hat{t}_{greg,lin} = (\mathbf{XB})'\mathbf{1} + \boldsymbol{r}'\breve{\mathbf{I}}$ (3.25), $\hat{t}_{syn,lin} = (\mathbf{XB})'\mathbf{1} + \boldsymbol{r}'\text{diag}(\mathbf{S})\breve{\mathbf{I}}$ (3.31).

Proposition 3.20 The approximate cross-mean square errors of the linear estimator $\hat{t}_y$, the GREG estimator $\hat{t}_{greg}$ and the SYN estimator $\hat{t}_{syn}$ are

$$\begin{aligned} \mathbb{AMSE}(\hat{t}_y,\, \hat{t}_{greg}) &= \boldsymbol{y}'\breve{\Delta}\boldsymbol{r}, & (3.86)\\ \mathbb{AMSE}(\hat{t}_y,\, \hat{t}_{syn}) &= \boldsymbol{y}'\breve{\Delta}\,\mathrm{diag}(\mathbf{S})\boldsymbol{r}, & (3.87)\\ \mathbb{AMSE}(\hat{t}_{greg},\, \hat{t}_{syn}) &= \boldsymbol{r}'\breve{\Delta}\,\mathrm{diag}(\mathbf{S})\boldsymbol{r}, & (3.88) \end{aligned}$$

where $\breve{\boldsymbol{\Delta}} = \mathbb{C}\mathrm{ov}(\breve{\mathbf{I}})$, $\boldsymbol{r} = \boldsymbol{y} - \mathbf{XB}$ and $\mathbf{S} = \boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{1}$, as defined in Proposition 3.7.

Proof. All three AMSEs of the proposition contain either unbiased $\hat{t}_y$ or asymptotically unbiased $\hat{t}_{greg}$ estimator. In this case, according to (1.65), MSE equals to the covariance. Approximate covariances of the estimators follow straightforwardly from the expressions of linear estimator and linearized GREG and SYN estimators.

□

Next we consider the domain's case. We use sub-indexes d and g for two different domains. With obvious changes between d and g, the formulas expressed for domain d hold also for domain g. We skip the linear domain estimator $\hat{t}^d_y = \boldsymbol{y}'_d\breve{\boldsymbol{I}}$, and concentrate on the estimators that use auxiliary information, the GREG and the SYN.

In the following proposition approximate cross-MSEs between different domain estimators are given. They are derived from the linearized forms of estimators, $\hat{t}^d_{greg,\,lin} = (\mathbf{X}_d\mathbf{B})'\mathbf{1} + \boldsymbol{r}'_d\breve{\mathbf{I}}$ from (3.69) and $\hat{t}^d_{syn,\,lin} = (\mathbf{X}_d\mathbf{B})'\mathbf{1} + \boldsymbol{r}'_{d\bullet}\mathrm{diag}(\mathbf{S}_d)\breve{\mathbf{I}}$ from (3.71).

Proposition 3.21 The approximate cross-mean square errors of domain estimators are the following:

$$\begin{aligned} \mathbb{AMSE}(\hat{t}^d_{greg},\, \hat{t}^g_{greg}) &= \boldsymbol{r}'_d\breve{\Delta}\boldsymbol{r}_g, & (3.89)\\ \mathbb{AMSE}(\hat{t}^d_{greg},\, \hat{t}^g_{syn}) &= \boldsymbol{r}'_d\breve{\Delta}\,\mathrm{diag}(\mathbf{S}_g)\boldsymbol{r}_{g\bullet}, & (3.90)\\ \mathbb{AMSE}(\hat{t}^d_{syn},\, \hat{t}^g_{syn}) &= \boldsymbol{r}'_{d\bullet}\mathrm{diag}(\mathbf{S}_d)\breve{\Delta}\,\mathrm{diag}(\mathbf{S}_g)\boldsymbol{r}_{g\bullet} + \mathbf{1}'\boldsymbol{r}_d\boldsymbol{r}'_g\mathbf{1}, & (3.91) \end{aligned}$$

with

$$\boldsymbol{r}_d = \boldsymbol{y}_d - \mathbf{X}_d\mathbf{B},\ \mathbf{B} = \mathbf{B}_1^{-1}\mathbf{B}_2,$$

where in the case of the assisting D-model,

$$\mathbf{B}_1 = \mathbf{X}_d'\mathbf{\Sigma}^{-1}\mathbf{X}_d,\ \mathbf{B}_2 = \mathbf{X}_d'\mathbf{\Sigma}^{-1}\boldsymbol{y}_d,\ \boldsymbol{r}_{d\bullet} = \boldsymbol{r}_d,\ \mathbf{S}_d = \mathbf{\Sigma}^{-1}\mathbf{X}\mathbf{B}_1^{-1}\boldsymbol{t}_{\mathbf{x}}^d;$$

and in the case of the assisting P-model,

$$\mathbf{B}_1 = \mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X},\ \mathbf{B}_2 = \mathbf{X}'\mathbf{\Sigma}^{-1}\boldsymbol{y},\ \boldsymbol{r}_{d\bullet} = \boldsymbol{r}_P = \boldsymbol{y} - \mathbf{X}\mathbf{B},\ \mathbf{S}_d = \mathbf{\Sigma}^{-1}\mathbf{X}\mathbf{B}_1^{-1}\boldsymbol{t}_{\mathbf{x}}^d.$$

Proof. The estimator $\hat{t}_{greg}^d$ is asymptotically unbiased. According to (1.65), the approximate MSE in (3.89) and (3.90) is equal to the approximate covariance between the estimators. Therefore, these results come directly from the linearized expressions of the domain GREG and the SYN estimators.

For the AMSE in (3.91) we use

$$A\mathbb{MSE}(\hat{t}_{syn}^d, \hat{t}_{syn}^g) = A\mathbb{C}\mathrm{ov}(\hat{t}_{syn}^d, \hat{t}_{syn}^g) + \left[A\boldsymbol{b}(\hat{t}_{syn}^d)\right]' A\boldsymbol{b}(\hat{t}_{syn}^g),$$

where the approximate covariance comes from (3.71) and the approximate bias from (3.79).

□

The following proposition gives expressions for the approximate cross-MSEs between domain estimators and estimators of the population total.

Proposition 3.22 The approximate cross-MSEs between domain SYN and GREG estimators and population SYN and GREG estimators are the following:

$$A\mathbb{MSE}(\hat{t}_{syn}^d, \hat{t}_{syn}) = \boldsymbol{r}_{d\bullet}'\mathrm{diag}(\mathbf{S}_d)\breve{\Delta}\mathrm{diag}(\mathbf{S})\boldsymbol{r}_P + \mathbf{1}'\boldsymbol{r}_d\boldsymbol{r}_P'\mathbf{1}, \qquad (3.92)$$

$$A\mathbb{MSE}(\hat{t}_{syn}^d, \hat{t}_{greg}) = \boldsymbol{r}_{d\bullet}'\mathrm{diag}(\mathbf{S}_d)\breve{\Delta}\boldsymbol{r}_P, \qquad (3.93)$$

$$A\mathbb{MSE}(\hat{t}_{greg}^d, \hat{t}_{syn}) = \boldsymbol{r}_d'\breve{\Delta}\,\mathrm{diag}(\mathbf{S})\boldsymbol{r}_P, \qquad (3.94)$$

$$A\mathbb{MSE}(\hat{t}_{greg}^d, \hat{t}_{greg}) = \boldsymbol{r}_d'\breve{\Delta}\,\boldsymbol{r}_P, \qquad (3.95)$$

where $\boldsymbol{r}_P = \boldsymbol{y} - \mathbf{X}\mathbf{B}$, $\mathbf{B} = (\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}^{-1}\boldsymbol{y}$, $\mathbf{S} = \mathbf{\Sigma}^{-1}(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{1}$. The domain quantities $\boldsymbol{r}_{d\bullet}$, $\boldsymbol{r}_d$ and $\mathbf{S}_d$ depend on the model, which is explained in Proposition 3.21.

Proof. In the approximate MSE (3.92) both estimators are biased, therefore

$$A\mathbb{MSE}(\hat{t}^d_{syn},\, \hat{t}_{syn}) = A\mathbb{C}\text{ov}(\hat{t}^d_{syn},\, \hat{t}_{syn}) + \left[A\boldsymbol{b}(\hat{t}^d_{syn})\right]' A\boldsymbol{b}(\hat{t}_{syn}).$$

Using linearized estimators, $\hat{t}^d_{syn,\,lin}$ and $\hat{t}_{syn,\,lin}$ defined in (3.71) and (3.31), we see that $\mathbb{C}\text{ov}(\hat{t}^d_{syn,\,lin},\, \hat{t}_{syn,\,lin})$ corresponds to the first part of (3.92). The second part comes from the corresponding approximate bias expressions in (3.79) and (3.35).

In AMSEs (3.93)-(3.95) at least one estimator is approximately unbiased. Therefore, the expressions (3.93)-(3.95) are obtained as covariances of the linearized estimators.

$\square$

3.7.2 Dependence characteristics in a particular case

Here we make further specifications on the estimators and sampling designs to approach to the situation considered in our simulation study. Besides we prove some interesting results.

We use three different domain estimators, i. e. we assume that some of the domains are estimated by $\hat{t}^d_{greg\text{-}D}$ and $\hat{t}^d_{greg\text{-}P}$, while the rest by $\hat{t}^d_{syn\text{-}P}$. We use $\hat{t}_{greg}$ for the population total. Since additive consistency between these estimators does not hold, we want to construct a restriction estimator that satisfies summation restriction. In fact, under the condition $\boldsymbol{\Sigma} = \text{diag}(\mathbf{X}\boldsymbol{\lambda})$, $\hat{t}^d_{greg\text{-}D} = \hat{t}^d_{syn\text{-}D}$.

For the restriction estimator we need the following approximate cross-MSEs in domains: $A\mathbb{MSE}(\hat{t}^d_{greg\text{-}D}, \hat{t}^g_{greg\text{-}P})$, $A\mathbb{MSE}(\hat{t}^d_{syn\text{-}P}, \hat{t}^g_{syn\text{-}P})$ and $A\mathbb{MSE}(\hat{t}^d_{greg\text{-}D}, \hat{t}^g_{syn\text{-}P})$. Corresponding general formulas are given in Proposition 3.21. Also we need the cross-AMSEs between domain estimators and the estimator of population total. They are $A\mathbb{MSE}(\hat{t}^d_{greg\text{-}D}, \hat{t}_{greg})$, $A\mathbb{MSE}(\hat{t}^d_{greg\text{-}P}, \hat{t}_{greg})$ and $A\mathbb{MSE}(\hat{t}^d_{syn\text{-}P}, \hat{t}_{greg})$. Formulas for them are given in Proposition 3.21.

All formulas in Propositions 3.20 - 3.22 require knowledge of the design covariance matrix $\breve{\boldsymbol{\Delta}}$. For some sampling designs this matrix has the form,

$$\breve{\boldsymbol{\Delta}} = \text{diag}(\breve{\boldsymbol{\Delta}}_0) + c\mathbf{1}\mathbf{1}' - c\mathbb{I}, \tag{3.96}$$

where $\breve{\mathbf{\Delta}}_0 = (\breve{\Delta}_{11}, \breve{\Delta}_{22}, ..., \breve{\Delta}_{NN})'$ with elements $\breve{\Delta}_{ii} = \mathbb{V}(I_i)/(\mathbb{E}I_i)^2$ and c is some constant. In other words, elements of the matrix $\breve{\mathbf{\Delta}}$ outside the main diagonal, $\breve{\Delta}_{ij} = \mathbb{C}\mathrm{ov}(I_i, I_j)/(\mathbb{E}I_i \cdot \mathbb{E}I_j)$, are equal to some constant c.

In this thesis we consider the SI and the MN sampling designs. For the SI design we get from (1.53) and (1.69):

$$c = -\frac{1-f}{f(N-1)}. \tag{3.97}$$

For the MN sampling design we get from (1.55) and (1.70):

$$c = -\frac{1}{n}. \tag{3.98}$$

We note that for the equal probability designs (like simple random sampling with and without replacement),

$$\breve{\Delta}_{ii} = \frac{N}{n} - 1 \equiv b, \tag{3.99}$$

is also a constant. Then the matrix $\breve{\mathbf{\Delta}}$ from (3.96) simplifies to

$$\breve{\mathbf{\Delta}} = (b - c)\mathbb{I} + c\mathbf{1}\mathbf{1}'. \tag{3.100}$$

For the sampling designs with $\breve{\mathbf{\Delta}}$ of structure (3.96), the $A\mathbb{M}\mathrm{SE}(\hat{t}^d_{greg}, \hat{t}^g_{greg})$ is zero. We show this in the following proposition.

Proposition 3.23 If the design covariance matrix $\breve{\mathbf{\Delta}}$ has the structure (3.96), then

$$A\mathbb{M}\mathrm{SE}(\hat{t}^d_{greg}, \hat{t}^g_{greg}) = 0. \tag{3.101}$$

This statement is true independently on the model (D- or P-model) for the domains d and g.

Proof. From (3.89) and (3.96) we have

$$\begin{aligned} A\mathbb{M}\mathrm{SE}(\hat{t}^d_{greg}, \hat{t}^g_{greg}) &= \boldsymbol{r}'_d \mathrm{diag}(\breve{\mathbf{\Delta}}_0)\boldsymbol{r}_g \\ &\quad + c\,\boldsymbol{r}'_d\mathbf{1}\mathbf{1}'\boldsymbol{r}_g - c\boldsymbol{r}'_d\boldsymbol{r}_g, \end{aligned} \tag{3.102}$$

where $\boldsymbol{r}_d$ and $\boldsymbol{r}_g$ are given in Proposition 3.21.

For the first term in (3.102) we first note that $\boldsymbol{r}_d$ and $\boldsymbol{r}_g$ can be written as

$$\boldsymbol{r}_d = \mathrm{diag}(\boldsymbol{\delta}_d)(\boldsymbol{y} - \mathbf{XB}), \text{ and } \boldsymbol{r}_g = \mathrm{diag}(\boldsymbol{\delta}_g)(\boldsymbol{y} - \mathbf{XB}),$$

where only $\mathbf{B}$ depends on the underlying model.

Since for $d \neq g$,

$$\mathrm{diag}(\boldsymbol{\delta}_d)\mathrm{diag}(\boldsymbol{\delta}_g) = 0, \tag{3.103}$$

and also

$$\mathrm{diag}(\boldsymbol{\delta}_d)\mathrm{diag}(\breve{\boldsymbol{\Delta}}_0)\mathrm{diag}(\boldsymbol{\delta}_g) = 0,$$

then the first summand in (3.102) is zero.

The second summand in (3.102) is zero, since

$$\boldsymbol{r}_d'\mathbf{1} = (\boldsymbol{y} - \mathbf{XB})'\mathrm{diag}(\boldsymbol{\delta}_d)\mathbf{1} = (\boldsymbol{y} - \mathbf{XB})'\boldsymbol{\delta}_d,$$

and then

$$\boldsymbol{r}_d'\mathbf{1}\mathbf{1}'\boldsymbol{r}_g = (\boldsymbol{y} - \mathbf{XB})'\boldsymbol{\delta}_d\boldsymbol{\delta}_g'(\boldsymbol{y} - \mathbf{XB}) = 0,$$

for $d \neq g$.

The third summand in (3.102) is also zero, due to (3.103). This proves (3.101).

$\square$

Proposition 3.23 claims that under stated conditions the GREG estimators for domains d and g are approximately uncorrelated.

The relationship (3.101) was shown in Sõstra (2007, p. 32) for the particular $\hat{t}^d_{greg\text{-}D}$, namely for the domain ratio estimator.

Remark 3.7 Since under the conditions $\boldsymbol{\Sigma} = \mathrm{diag}(\mathbf{X}\boldsymbol{\lambda})$, $\hat{t}^d_{greg\text{-}D} = \hat{t}^d_{syn\text{-}D}$, then the respective SYN estimators for domains d and g are also approximately uncorrelated for $\breve{\boldsymbol{\Delta}}$ in (3.96),

$$A\mathbb{C}\mathrm{ov}(\hat{t}^d_{syn\text{-}D}, \hat{t}^g_{syn\text{-}D}) = 0. \tag{3.104}$$

For the restriction estimator we also need the cross-MSEs between estimators of a domain and of the population total. The general formulas were given in Proposition 3.22. In some cases the formulas can be simplified.

Proposition 3.24 Let $\mathbf{\Sigma} = \sigma^2 \mathbb{I}$ and the assisting D-model has an intercept. Then for the equal probability designs with the covariance matrix $\breve{\mathbf{\Delta}}$ of structure (3.100), the following property takes place:

$$A\mathbb{M}\mathrm{SE}(\hat{t}^d_{greg\text{-}D}, \hat{t}_{greg}) = A\mathbb{V}(\hat{t}^d_{greg\text{-}D}). \tag{3.105}$$

Proof. Note, that this statement holds for the domain estimator $\hat{t}^d_{greg}$ under the D-model, denoted $\hat{t}^d_{greg\text{-}D}$. The AMSE for $\hat{t}^d_{greg}$, depending on the model, is given in Proposition 3.22. Using that and the matrix $\breve{\mathbf{\Delta}}$, given in (3.96), we have

$$A\mathbb{M}\mathrm{SE}(\hat{t}^d_{greg}, \hat{t}_{greg}) = (b-c)\boldsymbol{r}'_d \boldsymbol{r}_P + c\,\boldsymbol{r}'_d \mathbf{1}\mathbf{1}' \boldsymbol{r}_P, \tag{3.106}$$

where $\boldsymbol{r}_d = \boldsymbol{y}_d - \mathbf{X}_d \mathbf{B}_d$ with $\mathbf{B}_d = (\mathbf{X}'_d \mathbf{X}_d)^{-1} \mathbf{X}'_d \boldsymbol{y}_d$ under $\mathbf{\Sigma} = \sigma^2 \mathbb{I}$, and $\boldsymbol{r}_P = \boldsymbol{y} - \mathbf{X}\mathbf{B}$ with $\mathbf{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y}$ under the same $\mathbf{\Sigma}$.

For $\mathbf{\Sigma} = \sigma^2 \mathbb{I}$ and the model with intercept the condition $\mathbf{\Sigma} = \mathrm{diag}(\mathbf{X}\boldsymbol{\lambda})$ holds. Now, the second term in (3.106) is zero due to $\mathbf{1}'\boldsymbol{r}_P = 0$ (see Proposition 3.13).

For the first term of (3.106) we use

$$\boldsymbol{r}_d = \mathrm{diag}(\boldsymbol{\delta}_d)\boldsymbol{r}_d,$$

$$\mathrm{diag}(\boldsymbol{\delta}_d)\boldsymbol{r}_P = \boldsymbol{y}_d - \mathbf{X}_d \mathbf{B},$$

and we get

$$(b-c)\boldsymbol{r}'_d \boldsymbol{r}_P = (b-c)\boldsymbol{r}'_d \mathrm{diag}(\boldsymbol{\delta}_d)\boldsymbol{r}_P = (b-c)\boldsymbol{r}'_d(\boldsymbol{y}_d - \mathbf{X}_d \mathbf{B}).$$

With adding the term $\pm \mathbf{X}_d \mathbf{B}_d$ we have:

$$\begin{aligned} (b-c)\boldsymbol{r}'_d \boldsymbol{r}_P &= (b-c)\boldsymbol{r}'_d\left((\boldsymbol{y}_d - \mathbf{X}_d \mathbf{B}_d) + \mathbf{X}_d(\mathbf{B}_d - \mathbf{B})\right) \\ &= (b-c)\boldsymbol{r}'_d \boldsymbol{r}_d + (b-c)\boldsymbol{r}'_d \mathbf{X}_d(\mathbf{B}_d - \mathbf{B}). \end{aligned}$$

Now,

$$\begin{aligned} \boldsymbol{r}'_d \mathbf{X}_d &= (\boldsymbol{y}_d - \mathbf{X}_d \mathbf{B}_d)' \mathbf{X}_d \\ &= \boldsymbol{y}'_d \mathbf{X}_d - \mathbf{B}'_d \mathbf{X}'_d \mathbf{X}_d \\ &= \boldsymbol{y}'_d \mathbf{X}_d - \left(\boldsymbol{y}'_d \mathbf{X}_d (\mathbf{X}'_d \mathbf{X}_d)^{-1}\right) \mathbf{X}'_d \mathbf{X}_d = 0 \end{aligned}$$

Thus, the AMSE in (3.106) simplifies to

$$A\mathbb{M}\mathrm{SE}(\hat{t}^d_{greg\text{-}D}, \hat{t}_{greg}) = (b-c)\boldsymbol{r}'_d \boldsymbol{r}_d. \tag{3.107}$$

Now the $A\mathbb{V}(\hat{t}^{d}_{greg\text{-}D})$ for the sampling design in (3.100) is,

$$A\mathbb{V}(\hat{t}^{d}_{greg}) = (b-c)\boldsymbol{r}'_d\boldsymbol{r}_d + c\boldsymbol{r}'_d\mathbf{1}\mathbf{1}'\boldsymbol{r}_d. \tag{3.108}$$

Under $\boldsymbol{\Sigma} = \text{diag}(\mathbf{X}\boldsymbol{\lambda})$, $\boldsymbol{r}'_d\mathbf{1} = 0$ due to (3.85), and the approximate variance (3.108) equals to (3.107). This proves the statement (3.105) of the proposition.

$\square$

Special case of this result is proved in Sõstra (2007, p. 33) and in Sõstra and Traat (2009), where $\hat{t}^{d}_{greg}$ is a ratio estimator and $\mathbf{B}_d = \mathbf{B}$.

Chapter 4

Simulation study

In this thesis we use the real data of the healthcare personnel of Estonia, received from the Department of Health Statistics of the National Institute for Health Development. Sources of the data were health personnel hourly wages and other characteristics, received from economic reports of the healthcare providers and the statistical register of the healthcare providers.

4.1 Data description

The population of the healthcare personnel of Estonia (March 2009) consists of 21764 medical laborers from 1112 healthcare institutions (HI), such as clinics, family doctor offices, dental care centers, rehabilitation care centers and so on.

More precisely, the following study and auxiliary variables were included to the persons database, used by us as the population frame.

Two study variables:

– hourly wage (a continuous variables: in Estonian kroons),

– physician (a binary variable: 1- if the medical laborer is working as physician, for example surgeon, cardiologist, neurologist and so on; 0- otherwise).

The auxiliary variables:

- ID of healthcare institution (where the medical laborer belongs to),
- age of the medical laborer (in years) for the selection probabilities,
- sex of the medical laborer (0 for men and 1 for women),
- education level of the medical laborer (values from 1 to 5, where 5 is the highest level),
- domain indicator d ($d = 1, 2, 3, 4$).

The population is divided into 4 domains by the type of the healthcare institution:

- Domain 1: family doctor centres, emergency care, diagnostics providers, general and central hospitals;
- Domain 2: regional hospitals;
- Domain 3: special health care and dental care institutions;
- Domain 4: rehabilitation hospitals, nursing care hospitals, local hospitals.

The above division was made for the purposes of this thesis to achieve different domain sizes and the differences in study variable characteristics. Domain sizes are given in Table 4.1.

Table 4.1: Population and domain sizes

Domain	no. of laborers	%
1	10863	49.9
2	6742	31.0
3	3139	14.4
4	1020	4.7
Population	21764	100

The population characteristics of study variables are presented in Table 4.2. The study variables perform differently in the domains. For example, the smallest mean of the hourly wage is in the fourth, and the biggest is in the third domain. For the binary variable (physician) the fourth domain is also smallest in terms of means (proportions), whereas the first and the second domains are approximately equal. Differences in standard deviation (Std) are bigger for continuous variable.

Table 4.2: Population characteristics of study variables

Domain	Total	Mean	Min	Max	Std
		Continuous variable			
1	925580	85.20	13.64	705.88	47.54
2	628572	93.23	27.50	753.68	61.69
3	351135	111.86	3.10	1096.67	89.92
4	70673	69.29	27.69	455.92	39.61
Population	1975962	90.79	3.10	1096.67	60.41
		Binary variable			
1	2607	0.24	0	1	0.43
2	1512	0.22	0	1	0.42
3	600	0.19	0	1	0.39
4	154	0.15	0	1	0.36
Population	4873	0.22	0	1	0.42

Two sample designs were applied for the population frame, simple random sampling without replacement (SI) and the multinomial design (MN). The MN design, is an unequal probability sampling design with selection probabilities p_i, $i = 1, 2, ..., 21764$, found by the age of persons,

$$p_i = \frac{Age_i}{\sum_{i=1}^{21764} Age_i}.$$

This choice of p_i was not motivated by the precision of estimators. The aim was to check performance of our formulas under the design other than SI. It can be seen from Table 4.3 that selection probabilities are almost uncorrelated with our study variables in the population as a whole and also in each domain separately. The distribution of p_i is shown on Figure 4.1.

Table 4.3: Correlation between selection probabilities and study variables

Domain \ $Corr(p, \cdot)$	Continuous v.	Binary v.
1	0.051	0.068
2	0.043	0.063
3	-0.016	0.125
4	0.101	0.105
Population	0.022	0.073

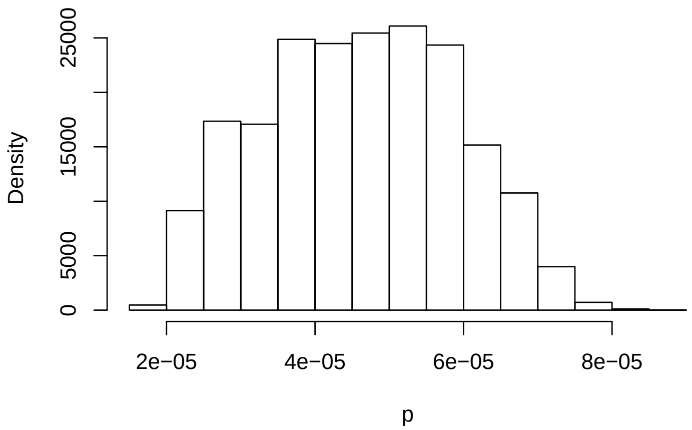

Figure 4.1: Distribution of the selection probabilities

For both designs 5000 independent samples were drawn with sample size 400 from the population of medical laborers.

The samples in the domains have different sizes over simulations (Table 4.4). There is no any empty domain sample through simulations.

Table 4.4: Sample sizes in the domains over simulations

Domain	Average	Minimum	Maximum
SI design, medical laborers			
1	199.8	159	235
2	123.9	90	158
3	57.6	37	84
4	18.7	3	34
Population	400.0	400	400
MN design, medical laborers			
1	204.8	168	238
2	116.9	87	148
3	57.8	34	84
4	20.5	4	38
Population	400	400	400

4.2 The choice of the initial estimators

We have four domains and two study variables. The natural restriction to require is summation of domain totals up to the population total. In our case the population total is not known (from the register or any other source). So, we are going to estimate four domain totals, and the population total as well, from the same sample. However, the initial estimators, chosen by us, do not satisfy the required summation restriction. The restriction matrix $\mathbf{R}$ and the restriction equation for the vector of true totals $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_P)'$, are:

$$\mathbf{R} = (1, 1, 1, 1, -1), \tag{4.1}$$

$$\mathbf{R}\boldsymbol{\theta} = 0. \tag{4.2}$$

In Chapter 3 the SYN and the GREG estimators for both the population and the domain total were studied. In the domain's case the underlying model (the population level P-model or the domain level D-model) plays an important role in the statistical properties of the estimator. Here we will study these properties empirically. More precisely, we observe the following estimators:

- the GREG estimator for the population total (3.22), $\hat{t}_{greg}$,

- the SYN estimator for the population total (3.29), $\hat{t}_{syn}$,
- the GREG (3.58) and the SYN (3.59) estimators for a domain total, both with the assisting D-model, $\hat{t}^d_{greg\text{-}D}$ and $\hat{t}^d_{syn\text{-}D}$ respectively,
- the GREG (3.63) and the SYN (3.64) estimators for a domain total, both with the assisting P-model, $\hat{t}^d_{greg\text{-}P}$ and $\hat{t}^d_{syn\text{-}P}$ respectively.

The auxiliary matrix consists of three variables: the vector of ones for the intercept, sex of a person and the level of his/her education. We have a domain indicator that allows to identify the auxiliary matrix at the domain level. Therefore, we can fit the model on both the population and the domain level.

The simple covariance structure, $\mathbf{\Sigma} = \boldsymbol{\sigma}^2\mathbb{I}$, and the model with intercept are assumed. The Remark 3.2 then tells that $\mathbf{\Sigma} = \text{diag}(\mathbf{X}\boldsymbol{\lambda})$ for some constant vector $\boldsymbol{\lambda}$. As a consequence, $\hat{t}_{greg} = \hat{t}_{syn}$; moreover, according to the Proposition 3.18, $\hat{t}^d_{greg\text{-}D} = \hat{t}^d_{syn\text{-}D}$.

For both designs, the four estimators $(\hat{t}_{greg}, \hat{t}^d_{greg\text{-}D}, \hat{t}^d_{greg\text{-}P}, \hat{t}^d_{syn\text{-}P})$ are computed from the $M = 5000$ samples. They are the building blocks for the restriction estimators. The following measures compare their performance over M simulations:

- the relative bias, $RB(\hat{\theta}) = \frac{\frac{1}{M}\sum_{m=1}^{M}\hat{\theta}^{(m)} - \theta}{\theta}$,
- the relative standard deviation, $RD(\hat{\theta}) = \frac{\sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(\hat{\theta}^{(m)} - \frac{1}{M}\sum_{m=1}^{M}\hat{\theta}^{(m)}\right)^2}}{\theta}$,
- the relative root mean square error, $RRMSE(\hat{\theta}) = \frac{\sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(\hat{\theta}^{(m)} - \theta\right)^2}}{\theta}$,

where $\hat{\theta}^{(m)}$ is the computed estimate of the population or domain total from the mth simulation and θ refers to the true total.

Performance measures of the four estimators under the SI case are given in Table 4.5, and under the MN case in Table 4.6.

The GREG estimator is nearly unbiased, also confirmed by Tables 4.5 and 4.6. Furthermore, we know that the GREG estimator for a domain is also nearly unbiased regardless of the model choice (Proposition 3.17).

Table 4.5: Characteristics of different estimators for the population and the domain totals, SI design

	True total	Mean	RB, %	RD, %	RRMSE, %
Continuous variable					
GREG					
Population	1975962	1976099	-0.01	2.65	2.65
SYN-P					
Dom 1	925580.8	971393.4	-4.95	2.75	5.66
Dom 2	628572.2	593646	5.56	2.39	6.05
Dom 3	351135.7	326783	6.94	2.89	7.51
Dom 4	70673	84023	-18.89	2.73	19.09
GREG-P					
Dom 1	925580.8	925658	-0.01	2.91	2.91
Dom 2	628572.2	627851	0.11	4.38	4.38
Dom 3	351135.7	351367	-0.07	9.52	9.52
Dom 4	70673	70798	-0.18	11.14	11.14
GREG-D					
Dom 1	925580.8	926278	-0.08	2.93	2.93
Dom 2	628572.2	628009	0.09	4.43	4.43
Dom 3	351135.7	350431	0.20	9.93	9.93
Dom 4	70673	69557	1.58	9.87	10.00
Binary variable					
GREG					
Population	4873	4875.89	-0.06	8.15	8.15
SYN-P					
Dom 1	2607	2349.33	9.88	7.39	12.34
Dom 2	1512	1405.21	7.06	7.66	10.42
Dom 3	600	935	-55.83	12.00	57.11
Dom 4	154	179	-13.97	8.94	19.27
GREG-P					
Dom 1	2607	2597	0.35	10.82	10.82
Dom 2	1512	1500	0.73	12.30	12.32
Dom 3	600	605	-0.83	28.50	28.51
Dom 4	154	153	0.00	55.84	55.84
GREG-D					
Dom 1	2607	2599	0.27	10.97	10.97
Dom 2	1512	1501	0.66	11.90	11.92
Dom 3	600	595	0.67	27.17	27.17
Dom 4	154	148	3.90	57.14	57.28

Table 4.6: Characteristics of different estimators for the population and the domain totals, MN design

	True total	Mean	RB, %	RD, %	RRMSE, %
Continuous variable					
GREG					
Population	1975962	1975115	0.04	2.77	2.77
SYN-P					
Dom 1	925580.8	970810.4	-4.89	2.86	5.66
Dom 2	628572.2	593697.7	5.55	2.53	6.10
Dom 3	351135.7	326572.8	7.00	3.03	7.62
Dom 4	70673	84034.14	-18.91	2.93	19.13
GREG-P					
Dom 1	925580.8	925407.7	0.02	3.03	3.03
Dom 2	628572.2	627943.7	0.10	4.66	4.67
Dom 3	351135.7	350678.9	0.13	10.06	10.06
Dom 4	70673	70782.9	-0.16	10.99	10.99
GREG-D					
Dom 1	925580.8	925118.4	0.05	2.96	2.96
Dom 2	628572.2	629157.5	-0.09	4.82	4.82
Dom 3	351135.7	351644.5	-0.14	10.59	10.59
Dom 4	70673	69894.8	1.10	10.08	10.14
Binary variable					
GREG					
Population	4873	4865.952	0.14	8.39	8.39
SYN-P					
Dom 1	2607	2345.494	10.03	7.65	12.62
Dom 2	1512	1405.913	7.02	8.03	10.66
Dom 3	600	935.0718	-55.85	12.53	57.23
Dom 4	154	179.4724	-16.54	11.14	19.94
GREG-P					
Dom 1	2607	2606.858	0.01	11.19	11.19
Dom 2	1512	1507.998	0.26	13.24	13.24
Dom 3	600	607.3754	-1.23	29.78	29.80
Dom 4	154	154.3719	-0.24	55.10	55.10
GREG-D					
Dom 1	2607	2598.228	0.34	11.44	11.45
Dom 2	1512	1508.435	0.24	13.13	13.13
Dom 3	600	603.7609	-0.63	27.54	27.55
Dom 4	154	151.6712	1.51	55.45	55.47

From Tables 4.5 and 4.6 (rows GREG-P and GREG-D) we see that the bias is ignorable for almost all domains. A small bias occurs in the fourth domain. This can be explained by small sample size in that domain, only 18.7 in average for SI case and 20.5 for MN case. Unbiasedness holds asymptotically.

From the same tables, we also see that the SYN-P is biased for the domain totals. For example, for both designs the bias is very large in the third domain (about 56%!). That means that the true regression model in this domain differs completely from the population regression model. The bias of SYN-P was notified in Remark 3.6 and the bias value was given in Proposition 3.19.

From Tables 4.5 and 4.6 we see that the variance of the SYN-P is smaller than that of the GREG-P. This illustrates the well known fact that the synthetic estimator has small variability (see e. g. Yang and Rubin-Bleuer, 2007). Looking on the RRMSE column, we see the opposite in most of the domains. The reason is in the substantial bias of the SYN-P estimator.

In Chapter 2 three different restriction estimators are described. All of them allow initial estimators (or some of them) to be biased. The following vector was chosen as the vector of initial estimators for four domains and the population total:

$$\hat{\boldsymbol{\theta}} = (\hat{t}^1_{greg\text{-}P}, \hat{t}^2_{greg\text{-}D}, \hat{t}^3_{syn\text{-}P}, \hat{t}^4_{syn\text{-}P}, \hat{t}_{greg})'. \tag{4.3}$$

In this vector, the estimators of the domains 3 and 4 ($\hat{t}^3_{syn\text{-}P}$ and $\hat{t}^4_{syn\text{-}P}$) are biased. The restriction equation (4.2) is not satisfied for (4.3), which means that sum of the domain estimators is not equal to the estimator of the population total. The estimators are not consistent.

We illustrate the non-consistency over simulations for both the SI and the MN designs. We calculate the difference and the relative difference:

$$\begin{aligned} Diff &= \sum_{d=1}^{4} \hat{\theta}^d - \hat{t}_{greg}, \\ RDiff &= Diff/t_{greg}, \end{aligned}$$

where $\hat{\theta}^d$ is the estimator of domain d in (4.3). In Table 4.7 mean, minimum and maximum difference are presented for both continuous and binary variable. Also the proportions of samples for which $|RDiff| > 0.01$ are calculated.

Table 4.7: Differences $Diff$ and its characteristics

Design	Mean	Minimum	Maximum	$\lvert RDiff \rvert > 0.01$
		Continuous variable		
SI	-10812.3	-152348.2	82927.4	0.495
MN	-8716.3	-152204.4	109970.9	0.515
		Binary variable		
SI	357.5	-304.3	971.1	0.973
MN	357.1	-385.9	944.5	0.967

We see that the consistency problem is quite serious, especially for the binary variable. For the continuous variable, about 50% of all samples have the relative difference bigger than 0.01, which means for half of the samples $|Diff| > 19759.6$ (0.01 of the true total). This difference is too large and can not be ignored.

For the binary variable about 97% of samples are inconsist to the extent $|RDiff| > 0.01$. The Figures 4.2 and 4.3 show the distribution of the relative difference for the SI case and Figures 4.4-4.5 for the MN case.

The distribution in Table 4.7 and on Figures 4.2, 4.4 indicates for the continuous variable that the sum of domain estimates tends to be smaller than the estimated population total. For the binary variable, as it can be seen from Figures 4.3, 4.5, this sum is almost always bigger.

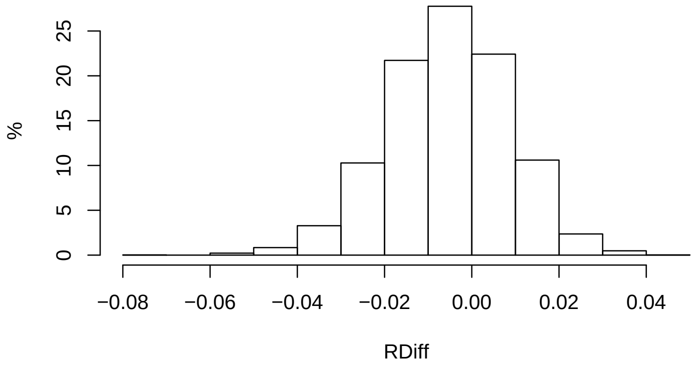

Figure 4.2: Distribution of relative difference, SI design, continuous variable

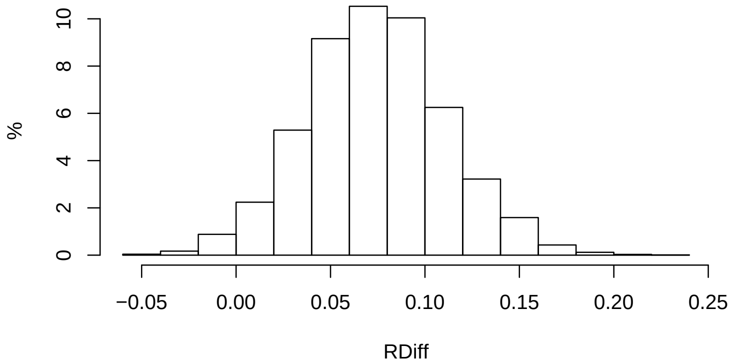

Figure 4.3: Distribution of relative difference, SI design, binary variable

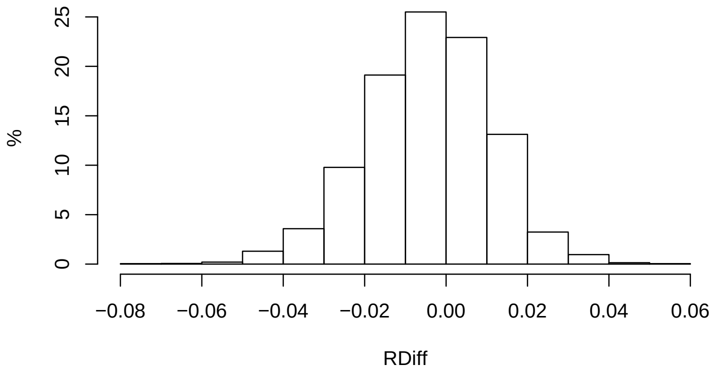

Figure 4.4: Distribution of relative difference, MN design, continuous variable

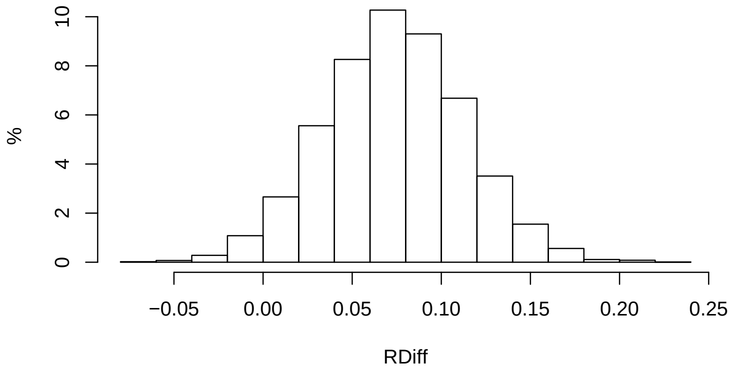

Figure 4.5: Distribution of relative difference, MN design, binary variable

4.3 MSE and bias of the initial estimators

For the GR estimators we need the MSE matrix $\mathbf{M}$ of the initial estimators, i. e. $\mathbf{M} = A\mathrm{MSE}(\hat{\boldsymbol{\theta}})$, $\hat{\boldsymbol{\theta}}$ given in (4.3). The components of $\mathbf{M}$ are calculated with formulas of Section 3.7.1, and are given in Table 4.8 for SI design and in Table 4.9 for MN design. The matrix $\mathbf{M}$ is positive definite in all cases as can be seen from the eigenvalues in Table 4.10.

Table 4.8: AMSE matrix $\mathbf{M}$ for the vector of initial estimators (4.3), SI case

Continuous variable, $\times 10^5$					
	$\hat{t}^1_{greg\text{-}P}$	$\hat{t}^2_{greg\text{-}D}$	$\hat{t}^3_{syn\text{-}P}$	$\hat{t}^4_{syn\text{-}P}$	$\hat{t}_{greg}$
$\hat{t}^1_{greg\text{-}P}$	7242.69	0	1330.84	283.31	7293.78
$\hat{t}^2_{greg\text{-}D}$	0	7586.72	1405.81	276.17	7586.72
$\hat{t}^3_{syn\text{-}P}$	1330.84	1405.81	6965.57	-3080.75	5165.77
$\hat{t}^4_{syn\text{-}P}$	283.31	276.17	-3080.75	1830.91	997.68
$\hat{t}_{greg}$	7293.78	7586.72	5165.77	997.68	27165.26
Binary variable					
	$\hat{t}^1_{greg\text{-}P}$	$\hat{t}^2_{greg\text{-}D}$	$\hat{t}^3_{syn\text{-}P}$	$\hat{t}^4_{syn\text{-}P}$	$\hat{t}_{greg}$
$\hat{t}^1_{greg\text{-}P}$	80889.04	0	13640.82	3442.92	81052.42
$\hat{t}^2_{greg\text{-}D}$	0	31821.16	6091.24	1219.67	31821.16
$\hat{t}^3_{syn\text{-}P}$	13640.82	6091.24	118052.3	9714.89	26821.95
$\hat{t}^4_{syn\text{-}P}$	3442.92	1219.67	9714.89	938.87	6226.87
$\hat{t}_{greg}$	81052.42	31821.16	26821.95	6226.87	152400.1

Table 4.9: AMSE matrix $\mathbf{M}$ for the vector of initial estimators (4.3), MN case

Continuous variable, $\times 10^5$					
	$\hat{t}^1_{greg\text{-}P}$	$\hat{t}^2_{greg\text{-}D}$	$\hat{t}^3_{syn\text{-}P}$	$\hat{t}^4_{syn\text{-}P}$	$\hat{t}_{greg}$
$\hat{t}^1_{greg\text{-}P}$	7809.75	0	1423.53	306.88	7861.8
$\hat{t}^2_{greg\text{-}D}$	0	8560.69	1546.3	311.56	8437.06
$\hat{t}^3_{syn\text{-}P}$	1423.53	1546.3	7070.47	-3060.53	5711.64
$\hat{t}^4_{syn\text{-}P}$	306.88	311.56	-3060.53	1835.61	1113.62
$\hat{t}_{greg}$	7861.8	8437.06	5711.64	1113.62	30161.26
Binary variable					
	$\hat{t}^1_{greg\text{-}P}$	$\hat{t}^2_{greg\text{-}D}$	$\hat{t}^3_{syn\text{-}P}$	$\hat{t}^4_{syn\text{-}P}$	$\hat{t}_{greg}$
$\hat{t}^1_{greg\text{-}P}$	87208.59	0	14588.88	3730.42	87375.03
$\hat{t}^2_{greg\text{-}D}$	0	38637.35	7266.21	1469.93	38215.83
$\hat{t}^3_{syn\text{-}P}$	14588.88	7266.21	118531.04	9818.22	29448.28
$\hat{t}^4_{syn\text{-}P}$	3730.42	1469.93	9818.22	966.72	6867.80
$\hat{t}_{greg}$	87375.03	38215.83	29448.28	6867.80	167846.21

It can be seen from all four $\mathbf{M}$ matrices that the domain estimators $\hat{t}^1_{greg\text{-}P}$ and $\hat{t}^2_{syn\text{-}D}$ are uncorrelated. This is in line with Proposition 3.23. The cross-AMSEs are generally positive. Only for the continuous variable it is negative between the estimators in the third and the fourth domains for both the SI and the MN cases. It can be explained by rather large negative bias in the fourth domain. For the third domain it is positive.

Table 4.10: Eigenvalues of the AMSE matrix $\mathbf{M}$

SI design		MN design	
Continuous variable	Binary variable	Continuous variable	Binary variable
3.29×10^9	2.19×10^5	3.64×10^9	2.39×10^5
7.62×10^8	1.09×10^5	8.18×10^8	1.09×10^5
7.41×10^8	4.50×10^4	7.64×10^8	5.24×10^4
2.86×10^8	1.07×10^4	3.23×10^8	1.19×10^4
4.85×10^5	27.86	5.44×10^5	29.96

Three diagonal elements of $\mathbf{M}$ that corresponds to the GREG estimators, are their approximate variances.

For the SI case we see that $A\mathbb{MSE}(\hat{t}^2_{greg\text{-}D}, \hat{t}_{greg}) = A\mathbb{V}(\hat{t}^2_{greg\text{-}D})$. For the MN case this does not hold. This illustrates Proposition 3.24.

The vector $\hat{\boldsymbol{\theta}}$ of initial estimators has two biased components, $\hat{t}^3_{syn\text{-}P}$ and $\hat{t}^4_{syn\text{-}P}$. Others are asymptotically unbiased (see Proposition 3.17). The approximate bias of these two components is calculated using (3.79), resulting in the following numerical vectors:

$$A\boldsymbol{b}(\hat{\boldsymbol{\theta}}) = (0, 0, -24376.56, 13388.26, 0)' \tag{4.4}$$

for the continuous variable, and

$$A\boldsymbol{b}(\hat{\boldsymbol{\theta}}) = (0, 0, 336.03, 25.87, 0)' \tag{4.5}$$

for the binary variable.

4.4 GR estimators

In Section 2.2 three different restriction estimators were described. They are:

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{GR1} &= (\mathbb{I} - \mathbf{KR})(\hat{\boldsymbol{\theta}} - \boldsymbol{b}),\\ \hat{\boldsymbol{\theta}}_{GR2} &= (\mathbb{I} - \mathbf{K}^*\mathbf{R})\hat{\boldsymbol{\theta}},\\ \hat{\boldsymbol{\theta}}_{GR3} &= (\mathbb{I} - \mathbf{K}^*\mathbf{R})(\hat{\boldsymbol{\theta}} - \boldsymbol{b}),\end{aligned}$$

where $\mathbf{K} = \mathbf{VR}'(\mathbf{RVR}')^{-1}$, $\mathbf{K}^* = \mathbf{MR}'(\mathbf{RMR}')^{-1}$, $\mathbf{V}$ and $\mathbf{M}$ are accordingly the covariance and the MSE matrices of the initial estimator-vector $\hat{\boldsymbol{\theta}}$ in (4.3). The GR estimators $\hat{\boldsymbol{\theta}}_{GR1}$ and $\hat{\boldsymbol{\theta}}_{GR3}$ are unbiased for $\boldsymbol{\theta}$. The bias of $\hat{\boldsymbol{\theta}}_{GR2}$ is

$$\boldsymbol{b}(\hat{\boldsymbol{\theta}}_{GR2}) = (\mathbb{I} - \mathbf{K}^*\mathbf{R})\boldsymbol{b}.$$

Numerically, in the SI case it is equal to

$$(1880.37, 2023.44, -22625.33, 11357.52, -7364.00)'$$

for the continuous variable, and to

$$(-38.16, -16.49, 63.88, 5.37, 14.61)'$$

for the binary variable. Under the MN design, values of the bias are accordingly,

$$(1800.13, 2125.25, -23016.43, 11543.36, -7547.70)'$$

and

$$(-40.27, -20.32, 68.13, 5.64, 13.18)'.$$

We see that the restriction estimator $\hat{\boldsymbol{\theta}}_{GR2}$ has all components biased. Compared to the initial estimator $\hat{\boldsymbol{\theta}}$, bias has appeared to the initially unbiased components, whereas it has decreased in the initially biased components. Despite of the bias, $\hat{\boldsymbol{\theta}}_{GR2}$ is more accurate than the initial $\hat{\boldsymbol{\theta}}$, in the MSE-terms.

The initial estimate-vector $\hat{\boldsymbol{\theta}}$ (4.3) is calculated from each of the 5000 samples, drawn separately by the SI and the MN designs. The estimates were computed with both study variables, the continuous and the binary variable. Then, in each sample three GR estimators were computed, based on $\hat{\boldsymbol{\theta}}$ in that sample and on the respective theoretical AMSE matrices $\mathbf{M}$ (given in

Tables 4.8 and 4.9). The covariance matrix $\mathbf{V}$ is computed from the relation $\mathbf{V} = \mathbf{M} - \boldsymbol{b}\boldsymbol{b}'$, where $\boldsymbol{b}$ is given in (4.4) and (4.5).

Table 4.11 shows results of the first three simulated samples to illustrate transformation of initial estimates to the GR estimates. All three GR estimators satisfy summation restriction to zero, while the initial estimator does not.

Table 4.11: The values of estimators of selected samples (SI and MN designs, binary variable)

Estimator	Sample	Domain d					
		$d = 1$	$d = 2$	$d = 3$	$d = 4$	Population	$\mathbf{R}\hat{\boldsymbol{\theta}} =$
				SI design			
Initial	1	2450.3	1428.3	884.1	178.3	4706.2	234.9
	2	2291.8	1878.7	1003.3	191.0	5160.7	204.1
	3	2777.1	1460.7	917.8	192.5	5019.6	328.6
GR1	1	2523.2	1459.8	544.0	151.2	4678.3	0.0
	2	2382.3	1917.8	662.2	163.7	5126.0	0.0
	3	2796.2	1469.0	580.7	166.4	5012.3	0.0
GR2	1	2425.6	1417.6	707.4	165.0	4715.7	0.0
	2	2270.2	1869.4	849.8	179.5	5168.9	0.0
	3	2742.5	1445.8	670.7	173.9	5032.9	0.0
GR3	1	2463.7	1434.1	643.6	159.6	4701.0	0.0
	2	2308.4	1885.9	785.9	174.1	5154.3	0.0
	3	2780.6	1462.3	606.8	168.6	5018.2	0.0
				MN design			
Initial	1	2666.0	1676.8	952.5	178.1	4883.7	589.7
	2	2419.3	1423.5	931.1	169.9	4725.0	218.8
	3	2674.4	1241.9	908.7	164.7	4604.8	384.8
GR1	1	2533.9	1613.3	622.7	154.0	4923.8	0.0
	2	2502.3	1463.4	591.2	142.9	4699.8	0.0
	3	2661.1	1235.5	573.3	139.0	4608.8	0.0
GR2	1	2600.0	1645.1	513.6	145.0	4903.7	0.0
	2	2394.8	1411.7	768.3	157.6	4732.4	0.0
	3	2631.3	1221.1	622.4	143.1	4617.9	0.0
GR3	1	2640.5	1664.6	446.9	139.5	4891.4	0.0
	2	2435.3	1431.2	701.5	152.1	4720.1	0.0
	3	2671.8	1240.6	555.6	137.5	4605.6	0.0

Besides the consistency of estimators, the variability and the bias of estimators are also of interest. From Proposition 2.5 we know that all three GR estimators should be more accurate than the initial estimator in the MSE terms. Theoretical AMSEs are calculated with formulas (2.11), (2.15) and (2.21). Results are included in Table 4.12 for the SI case and 4.13 for the MN case.

For better visualization, MSE matrices are presented in a vectorized form with AMSEs of the domain and the population estimators in the bold font. From the Löwner ordering of matrices (2.24), the same ordering holds for the diagonal elements of these matrices (see (1.12)). The AMSEs in bold in Tables 4.12 and 4.13 confirm this result for all domains and for the population in both SI and MN cases.

For the smallest fourth domain and the continuous variable, the AMSE of GR1 is about 50 times smaller than that of the initial estimator in that domain. This holds for both sampling designs. For the third domain the decrease is also large. It can be explained by very biased initial estimators (SYN-P) in these domains (see Tables 4.5 and 4.6). In the formula of $\hat{\boldsymbol{\theta}}_{GR1}$ (2.10) this bias is first subtracted and then the transformation by $(\mathbb{I} - \mathbf{KR})$ is made, which optimizes the accuracy of the estimator.

The estimator GR3 is the second best estimator, as can be seen from the AMSEs. But both estimators, GR1 and GR3, require knowledge of the initial bias. The estimator GR2 is free of this requirement, but is biased itself. Still, the AMSE of GR2 is smaller than that of the initial estimator in all domains and in the population. Tables 4.5 and 4.6 confirm this also. The decrease in AMSEs is especially remarkable in the case of the binary variable for the third and fourth domains, where initial estimators were substantially biased.

Table 4.12: Theoretical AMSEs between estimators for SI case

	$\hat{\boldsymbol{\theta}}_{\boldsymbol{GR1}}$	$\hat{\boldsymbol{\theta}}_{\boldsymbol{GR3}}$	$\hat{\boldsymbol{\theta}}_{\boldsymbol{GR2}}$	$\hat{\boldsymbol{\theta}}$
Continuous variable, $\times 10^5$				
Domain 1	**6934.5**	**6939.9**	**6975.2**	**7242.7**
Domain 1, Domain 2	-331.7	-325.9	-287.8	0
Domain 1, Domain 3	1572.0	1507.2	1081.7	1330.8
Domain 1, Domain 4	326.1	358.6	572.2	283.3
Domain 1, Population	8500.9	8479.8	8341.3	7293.8
Domain 2	**7229.8**	**7236.1**	**7277.0**	**7586.7**
Domain 2, Domain 3	1665.3	1595.6	1137.8	1405.8
Domain 2, Domain 4	322.2	357.2	587.0	276.2
Domain 2, Population	8885.6	8862.9	8713.9	7586.7
Domain 3	**834.7**	**1614.5**	**6733.6**	**6965.6**
Domain 3, Domain 4	149.4	-242.0	-2811.7	-3080.8
Domain 3, Population	4221.4	4475.2	6141.3	5165.8
Domain 4	**32.5**	**229.0**	**1518.9**	**1830.9**
Domain 4, Population	830.2	702.8	-133.6	997.7
Population	**22438.1**	**22520.7**	**23063.0**	**27165.3**
Binary variable				
Domain 1	**71185.0**	**77648.9**	**79105.0**	**80889.0**
Domain 1, Domain 2	-4192.9	-1399.9	-770.9	0
Domain 1, Domain 3	14176.4	3354.5	916.8	13640.8
Domain 1, Domain 4	3599.0	2689.4	2484.5	3442.9
Domain 1, Population	84767.5	82292.9	81735.4	81052.4
Domain 2	**30009.5**	**31216.3**	**31488.1**	**31821.2**
Domain 2, Domain 3	6322.7	1646.8	593.5	6091.2
Domain 2, Domain 4	1287.1	894.1	805.6	1219.7
Domain 2, Population	33426.4	32357.1	32116.3	31821.2
Domain 3	**5104.1**	**23222.1**	**27303.3**	**118052.3**
Domain 3, Domain 4	1013.7	2536.6	2879.6	9714.9
Domain 3, Population	26616.9	30759.9	31693.2	26822.0
Domain 4	**267.2**	**395.2**	**424.0**	**938.9**
Domain 4, Population	6167.1	6515.3	6593.8	6226.9
Population	**150977.9**	**151925.2**	**152138.6**	**152400.1**

Table 4.13: Theoretical AMSEs between estimators for MN case

Continuous variable, $\times 10^5$				
	$\hat{\boldsymbol{\theta}}_{\boldsymbol{GR1}}$	$\hat{\boldsymbol{\theta}}_{\boldsymbol{GR3}}$	$\hat{\boldsymbol{\theta}}_{\boldsymbol{GR2}}$	$\hat{\boldsymbol{\theta}}$
Domain 1	**7498.1**	**7502.4**	**7534.8**	**7809.7**
Domain 1, Domain 2	-368.0	-362.9	-324.6	0.0
Domain 1, Domain 3	1685.5	1630.1	1215.8	1423.5
Domain 1, Domain 4	353.1	380.9	588.7	306.9
Domain 1, Population	9168.7	9150.5	9014.6	7861.8
Domain 2	**8126.2**	**8132.3**	**8177.4**	**8560.7**
Domain 2, Domain 3	1855.5	1790.2	1301.0	1546.3
Domain 2, Domain 4	366.1	398.9	644.2	311.6
Domain 2, Population	9980.0	9958.5	9798.1	8437.1
Domain 3	**908.2**	**1615.9**	**6913.5**	**7070.5**
Domain 3, Domain 4	164.2	-190.7	-2847.6	-3060.5
Domain 3, Population	4613.4	4845.5	6582.7	5711.6
Domain 4	**36.3**	**214.3**	**1546.8**	**1835.6**
Domain 4, Population	919.8	803.4	-67.9	1113.6
Population	**24681.8**	**24757.9**	**25327.6**	**30161.3**
Binary variable				
Domain 1	**76959.1**	**83566.5**	**85188.5**	**87208.6**
Domain 1, Domain 2	-5170.6	-1837.3	-1019.1	0.0
Domain 1, Domain 3	15071.7	3894.5	1150.8	14588.9
Domain 1, Domain 4	3868.3	2942.9	2715.8	3730.4
Domain 1, Population	90728.5	88566.6	88036.0	87375.0
Domain 2	**36028.9**	**37710.5**	**38123.2**	**38637.4**
Domain 2, Domain 3	7509.8	1871.2	487.0	7266.2
Domain 2, Domain 4	1539.5	1072.7	958.1	1469.9
Domain 2, Population	39907.6	38817.0	38549.3	38215.8
Domain 3	**5589.7**	**24497.1**	**29138.4**	**118531.0**
Domain 3, Domain 4	1119.2	2684.5	3068.8	9818.2
Domain 3, Population	29290.3	32947.3	33845.0	29448.3
Domain 4	**295.7**	**425.3**	**457.1**	**966.7**
Domain 4, Population	6822.7	7125.4	7199.8	6867.8
Population	**166749.0**	**167456.3**	**167630.0**	**167846.2**

Since MSEs of the initial estimators were developed to hold asymptotically, then the MSEs (2.11), (2.15) and (2.21) of the GR estimators involving them will be also the approximate MSEs. These approximate theoretical MSEs were compared with the empirical ones. On Figure 4.6 this is done for the SI case, and on Figure 4.7 for the MN case. We see that for the SI design, empirical results are similar to the theoretical values. Only for the second domain (with the GREG-D as the initial estimator), and for the population (with the GREG as the initial), the theoretical AMSE of the continuous variable seems to underestimate a little the real MSE (the empirical). But for the binary variable and for the population case, the theoretical AMSE is bigger than the empirical one. Nevertheless, we cannot make any conclusion here about some tendentious overestimating or underestimating of MSEs. For the MN case, we see that theoretical AMSEs are smaller than corresponding empirical MSEs almost in all cases. But this difference is very small. Values of the empirical MSEs and cross-MSEs are given in the Appendix C, Tables 5.1 and 5.2.

Figures 4.6 and 4.7 confirm the findings of the previous tables of MSEs – the highest benefit of GR estimators is got for the third and for the fourth domains, which have the substantial initial bias.

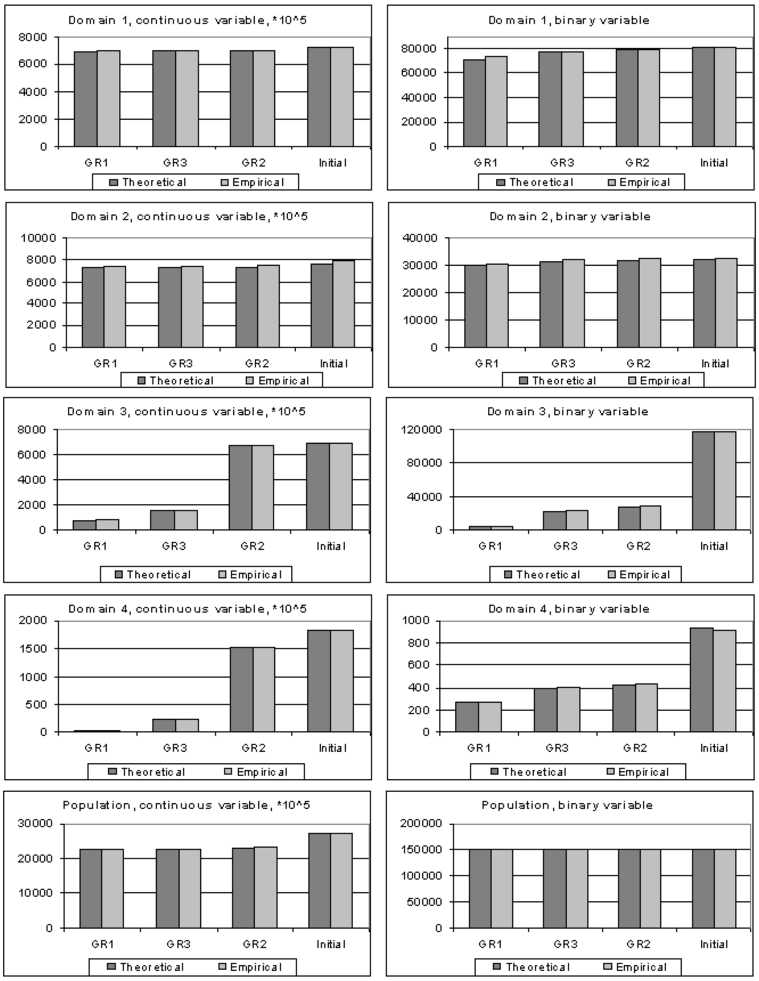

Figure 4.6: Empirical and theoretical MSEs, SI case.

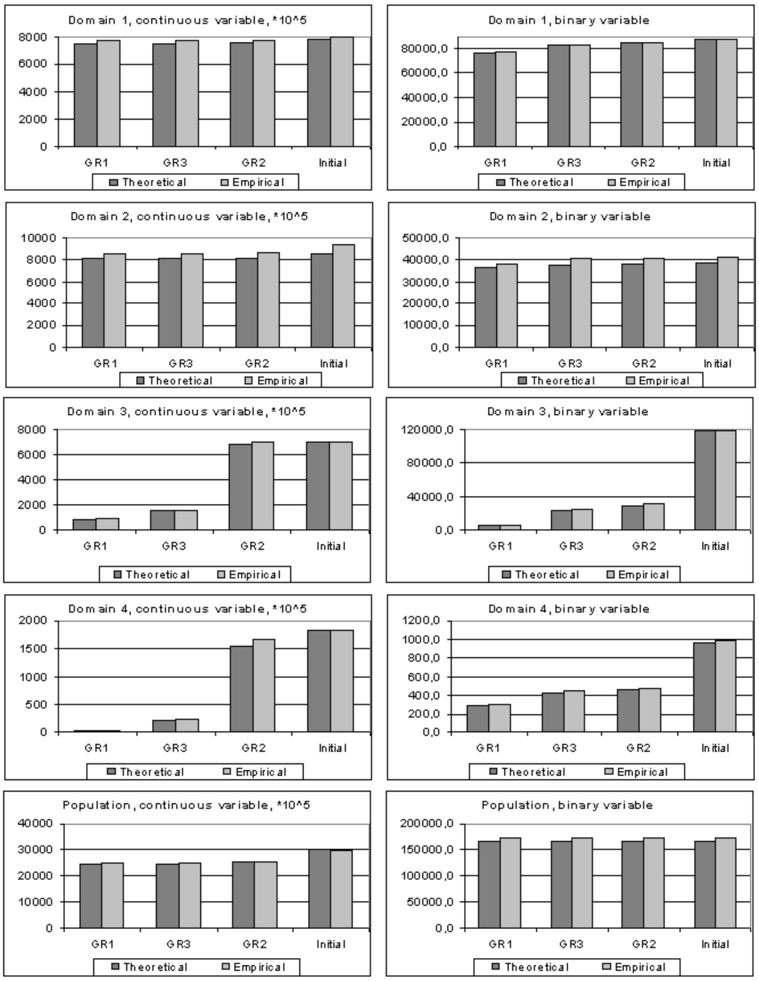

Figure 4.7: Empirical and theoretical MSEs, MN case.

4.5 Conclusions from simulations

The simulation study confirmed the theoretical results of this thesis. Though, many theoretical results were developed to hold asymptotically, they worked well also with rather small sample sizes in our examples.

Three different restriction estimators were considered for four domains with the requirement that these estimators sum up to the estimator of the population total, received from the same sample.

The biased estimator SYN-P was chosen as an initial estimator for two domains, while for other two domains the asymptotically unbiased estimators were taken. Behavior of these initial estimators was studied in terms of the relative bias and the RRMSE.

Known AMSE matrix of initial estimators was used in the restriction estimators. Simulation-based behavior of the restriction estimators was studied with special emphasis on their mean square errors. Samples were taken by two sampling design, the SI (as the equal probability design) and the MN (as the unequal probability design). Population size was 21764 and sample size 400 persons.

The main simulation results are summarized below.

- Different estimators (GREG and SYN) for the domains were studied. The SYN-P estimator showed smaller variance than the GREG estimator, but lead to the enormous bias.
- The GREG estimator is asymptotically unbiased for the domain total (irrespective of the underlying model). But for small sample sizes the minor bias may occur (like it happened in our example with GREG-D for the fourth domain).
- In the role of initial estimators, two asymptotically unbiased and two biased estimators were chosen for domain totals, and the asymptotically unbiased GREG for the population total. The AMSE matrix of the initial estimator-vector was found for both designs. Its structure illustrated theoretical results of the thesis. It was seen from the AMSE matrix that the domain GREG estimators were uncorrelated irrespective of the design. For the SI design, the AMSE of the GREG-D in

the second domain and the GREG in the population was equal to the asymptotic variance of the GREG-D. This property did not hold for the MN design.

– The extent of the inconsistency of the initial estimators was studied, and found to be rather large. The inconsistency disappeared with putting on restrictions and going over to the GR estimators.

– Two of the GR estimators (GR1 and GR3) were unbiased, but required knowledge of the bias of initial estimators. The estimator GR2 was free of this requirement, but biased itself. All three GR estimators showed smaller AMSEs than these of the initial estimators in both calculations – with theoretical formulas derived in this thesis, and also empirically over repeated samples. For the domains with biased initial estimators, the AMSE of the GR1 and GR3 showed high decrease when compared with the AMSE of initial estimators. As it was expected, the AMSE of the GR1 was smallest (for the continuous variable and fourth domain it was about 55 times smaller than the AMSE of the initial estimator). For the GR3 the biggest decrease compared with the initial AMSE was about 8 times. The AMSE of the GR2 was much bigger than that of the GR1 and the GR3, due to the bias, but still considerable smaller than that of the initial estimator. For the asymptotically unbiased domain estimators, the decrease in AMSEs was not so large.

– Empirically computed MSEs of the GR estimators were close to the asymptotic theoretical values for both designs and for both study variables.

Chapter 5

General conclusions

In this thesis estimation under linear restrictions was studied. More specifically, we focused on the domain estimation under summation restriction. Our starting point was that domains are initially estimated by conceptually different estimators with some of them being biased, and their sum is not equal to the estimated population total. We used the initial estimators from the GREG and the SYN families, specified under two different models for the domains - the population and the domain model.

The main goals of this thesis were achieved. The following results were received

1. The three GR estimators satisfying linear restrictions, generalized to allow biased initial estimators, were proposed. Though we concentrated more deeply on the summation restriction in the domains, all derived formulas for GR estimators in Chapter 2 hold for general linear restrictions.
2. The mean square error (MSE) and the bias expressions for the GR estimators were derived. This was done in matrix form since estimation under restrictions is a multivariate problem.
3. The MSE matrices of the three GR estimators were ordered (in the Löwner sense), and the GR estimator with the smallest MSE matrix was found. It appeared that the GR2 estimator was most accurate. It was shown that all tree GR estimators are never less accurate (in MSE terms) than the initial estimator.

4. Two important estimators, the GREG and the SYN, were elaborated in more detail, both for estimation of the population total and then for the domains case. Though GREG is much considered in the literature, and to some extent also SYN, we still discovered many new properties of these estimators, especially for SYN and especially for the mutual relationship of the GREG and the SYN.

5. The conditions for the equality of the GREG and the SYN were deeply studied, and novel results were established for the domains case.

6. The second-order Taylor expansion was derived for the GREG and the first-order for the SYN estimators. Based on these, the linearized forms of the estimators were derived, as for the population total, so for the domain totals. In domain's case the two different assisting models (population and domain models) were treated.

7. The expressions of the bias, the covariance and the mean square error of the estimators were found. They were the approximate (linearization-based) expressions.

8. The expressions for the approximate cross-mean square errors between the GREG and the SYN estimators were developed, for both assisting models in domains case.

9. Special cases of the design covariance matrix $\mathbf{\Delta}$ were also considered. These cases included the SI and the MN designs. Respectively, some approximate mean square errors simplified under assumptions for $\mathbf{\Delta}$. For example, the domain GREG estimators were uncorrelated irrespective of the model.

10. The theoretical results of this thesis were illustrated and tested in a simulation study, which confirmed the derived properties of the GR and the considered initial estimators. Conclusions on the simulation results are summarized in Section 4.5.

Appendix A

Proof of the Proposition 3.3

Proof. The estimator

$$\hat{t}_{greg} = \hat{t}_y - \hat{\mathbf{T}}'_{\mathbf{x}y}\hat{\mathbf{T}}^{-1}_{\mathbf{xx}}(\hat{\boldsymbol{t}}_{\mathbf{x}} - \boldsymbol{t}_{\mathbf{x}})$$

can be viewed as a function at four arguments, three of which are matrices,

$$\hat{t}_{greg} = \boldsymbol{f}(\mathbf{W}), \tag{A.1}$$

where $\mathbf{W} = (\hat{t}_y, \hat{\boldsymbol{t}}_{\mathbf{x}}, \hat{\mathbf{T}}_{\mathbf{x}y}, \hat{\mathbf{T}}_{\mathbf{xx}})$.

For the Taylor expansion we need derivatives of $\hat{t}_{greg}$ up to the second order in a neighborhood of a point $\mathbf{W}_0 = (t_y, \boldsymbol{t}_{\mathbf{x}}, \mathbf{T}_{\mathbf{x}y}, \mathbf{T}_{\mathbf{xx}})$ that consists of true totals. We use the expansion (1.51) for $\boldsymbol{f}(\mathbf{W})$, where $\mathbf{W}$ is replaced with its vectorized form,

$$\mathrm{vec}(\mathbf{W}) = (\hat{t}_y, \hat{\boldsymbol{t}}'_{\mathbf{x}}, \hat{\mathbf{T}}'_{\mathbf{x}y}, \mathrm{vec}'\hat{\mathbf{T}}_{\mathbf{xx}})'. \tag{A.2}$$

Properties of the Proposition 1.3 are used for the matrix derivatives. We denote the first derivatives with respect to each of the four arguments by I, II, III and IV.

$$\begin{aligned}
\mathrm{I} &= \frac{d\,\hat{t}_{greg}}{d\,\hat{t}_y} = 1.\\
\mathrm{II} &= \frac{d\,\hat{t}_{greg}}{d\,\hat{\boldsymbol{t}}_{\mathbf{x}}} \overset{(1.45)}{=} \frac{d\,(\hat{\boldsymbol{t}}_{\mathbf{x}} - \boldsymbol{t}_{\mathbf{x}})}{d\,\hat{\boldsymbol{t}}_{\mathbf{x}}} \cdot \frac{d\;\hat{t}_{greg}}{d\,(\hat{\boldsymbol{t}}_{\mathbf{x}} - \boldsymbol{t}_{\mathbf{x}})} \overset{(1.43)}{=} \mathbb{I}_p \cdot (-1 \otimes \hat{\mathbf{B}})\\
&= -\hat{\mathbf{T}}^{-1}_{\mathbf{xx}}\hat{\mathbf{T}}_{\mathbf{x}y}.
\end{aligned}$$

$$\begin{aligned}
\mathrm{III} &= \frac{d\,\hat{t}_{greg}}{d\,\hat{\mathbf{T}}_{\mathbf{x}y}} \overset{(1.45)}{=} \frac{d\,(\hat{\mathbf{T}}'_{\mathbf{x}y})}{d\,\hat{\mathbf{T}}_{\mathbf{x}y}} \cdot \frac{d\,\hat{t}_{greg}}{d\,(\hat{\mathbf{T}}'_{\mathbf{x}y})}\\
&\overset{(1.47)}{=} \mathbf{K}_{1,p} \cdot \frac{d\left[-\hat{\mathbf{T}}'_{\mathbf{x}y}\hat{\mathbf{T}}^{-1}_{\mathbf{xx}}(\hat{\boldsymbol{t}}_{\mathbf{x}} - \boldsymbol{t}_{\mathbf{x}})\right]}{d\,\hat{\mathbf{T}}'_{\mathbf{x}y}}\\
&\overset{(1.43)}{=} -\mathbb{I}_p\left[\hat{\mathbf{T}}^{-1}_{\mathbf{xx}}(\hat{\boldsymbol{t}}_{\mathbf{x}} - \boldsymbol{t}_{\mathbf{x}}) \otimes 1\right] = -\hat{\mathbf{T}}^{-1}_{\mathbf{xx}}(\hat{\boldsymbol{t}}_{\mathbf{x}} - \boldsymbol{t}_{\mathbf{x}}).
\end{aligned}$$

$$
\begin{aligned}
\mathrm{IV} \quad &= \quad \frac{d\,\hat{t}_{greg}}{d\,\hat{\mathbf{T}}_{\mathbf{xx}}} \overset{(1.45)}{=} \frac{d\,(\hat{\mathbf{T}}_{\mathbf{xx}}^{-1})}{\hat{\mathbf{T}}_{\mathbf{xx}}} \cdot \frac{d\,\hat{t}_{greg}}{d\,(\hat{\mathbf{T}}_{\mathbf{xx}}^{-1})} \\
&\overset{(1.44)}{=} \quad (\hat{\mathbf{T}}_{\mathbf{xx}}^{-1} \otimes \hat{\mathbf{T}}_{\mathbf{xx}}^{-1}) \left[(\hat{\boldsymbol{t}}_{\mathbf{x}} - \boldsymbol{t}_{\mathbf{x}}) \otimes \hat{\mathbf{T}}_{\mathbf{x}y}\right] \\
&= \quad \hat{\mathbf{T}}_{\mathbf{xx}}^{-1}(\hat{\boldsymbol{t}}_{\mathbf{x}} - \boldsymbol{t}_{\mathbf{x}}) \otimes \hat{\mathbf{T}}_{\mathbf{xx}}^{-1}\hat{\mathbf{T}}_{\mathbf{x}y}.
\end{aligned}
$$

Now we derive the necessary second order derivatives. Trivially, we get zero matrices for all second derivatives of I. The second derivatives of II, III and IV by $\hat{\boldsymbol{t}}_{\mathbf{x}}$ are the following:

$$
\begin{aligned}
\frac{d}{d\,\hat{\boldsymbol{t}}_{\mathbf{x}}}(\mathrm{II}) \quad &= \quad \mathbf{0}; \\
\frac{d}{d\,\hat{\boldsymbol{t}}_{\mathbf{x}}}(\mathrm{III}) \quad &\overset{(1.45)}{=} \quad -\frac{d\,(\hat{\boldsymbol{t}}_{\mathbf{x}} - \boldsymbol{t}_{\mathbf{x}})}{d\,\hat{\boldsymbol{t}}_{\mathbf{x}}} \cdot \frac{d\ (\mathrm{III})}{d\,(\hat{\boldsymbol{t}}_{\mathbf{x}} - \boldsymbol{t}_{\mathbf{x}})} \overset{(1.43)}{=} -\mathbb{I}_p \left(1 \otimes \hat{\mathbf{T}}_{\mathbf{xx}}^{-1}\right) = \hat{\mathbf{T}}_{\mathbf{xx}}^{-1}; \\
\frac{d}{d\,\hat{\boldsymbol{t}}_{\mathbf{x}}}(\mathrm{IV}) \quad &= \quad \frac{d\,(\mathrm{IV})}{d\,(\hat{\boldsymbol{t}}_{\mathbf{x}} - \boldsymbol{t}_{\mathbf{x}})} \\
&\overset{(1.48)}{=} \quad \left\{ \frac{d\,\left(\hat{\mathbf{T}}_{\mathbf{xx}}^{-1}(\hat{\boldsymbol{t}}_{\mathbf{x}} - \boldsymbol{t}_{\mathbf{x}})\right)}{d\,(\hat{\boldsymbol{t}}_{\mathbf{x}} - \boldsymbol{t}_{\mathbf{x}})} \otimes \mathrm{vec}'(\hat{\mathbf{T}}_{\mathbf{xx}}^{-1}\hat{\mathbf{T}}_{\mathbf{x}y}) + \mathbf{0} \right\} (1 \otimes \mathbf{K}_{p,1} \otimes \mathbb{I}_p) \\
&\overset{(1.43)}{=} \quad (1 \otimes \hat{\mathbf{T}}_{\mathbf{xx}}^{-1}) \otimes \mathrm{vec}'(\hat{\mathbf{T}}_{\mathbf{xx}}^{-1}\hat{\mathbf{T}}_{\mathbf{x}y}) = \hat{\mathbf{T}}_{\mathbf{xx}}^{-1} \otimes \hat{\mathbf{T}}'_{\mathbf{x}y}\hat{\mathbf{T}}_{\mathbf{xx}}^{-1},
\end{aligned}
$$

where the vec-operator disappeared due to $\hat{\mathbf{T}}_{\mathbf{xx}}^{-1}\hat{\mathbf{T}}_{\mathbf{x}y}$ being a vector itself.

The second derivatives by $\hat{\mathbf{T}}_{\mathbf{x}y}$ have the following expressions:

$$
\begin{aligned}
\frac{d}{d\,\hat{\mathbf{T}}_{\mathbf{x}y}}(\mathrm{II}) \quad &\overset{(1.43)}{=} \quad -1 \otimes \hat{\mathbf{T}}_{\mathbf{xx}}^{-1}; \\
\frac{d}{d\,\hat{\mathbf{T}}_{\mathbf{x}y}}(\mathrm{III}) \quad &= \quad \mathbf{0}; \\
\frac{d}{d\,\hat{\mathbf{T}}_{\mathbf{x}y}}(\mathrm{IV}) \quad &= \quad \mathbf{0}, \text{ at the point } \hat{\boldsymbol{t}}_{\mathbf{x}} = \boldsymbol{t}_{\mathbf{x}}.
\end{aligned}
$$

The second derivatives by $\hat{\mathbf{T}}_{\mathbf{xx}}$ of III and IV will lead to zero at the point $\hat{\boldsymbol{t}}_{\mathbf{x}} = \boldsymbol{t}_{\mathbf{x}}$. So, we have only one derivative different from $\mathbf{0}$ at this point:

$$
\frac{d}{d\,\hat{\mathbf{T}}_{\mathbf{xx}}}(\mathrm{II}) \quad \overset{(1.45)}{=} \quad \frac{d\,\hat{\mathbf{T}}_{\mathbf{xx}}^{-1}}{\hat{\mathbf{T}}_{\mathbf{xx}}} \cdot \frac{d(\mathrm{II})}{d\,\hat{\mathbf{T}}_{\mathbf{xx}}^{-1}}
$$

$$\overset{(1.44)}{=} \quad (\hat{\mathbf{T}}_{\mathbf{xx}}^{-1} \otimes \hat{\mathbf{T}}_{\mathbf{xx}}^{-1})(\hat{\mathbf{T}}_{\mathbf{x}y} \otimes \mathbb{I}_p) = \hat{\mathbf{T}}_{\mathbf{xx}}^{-1}\hat{\mathbf{T}}_{\mathbf{x}y} \otimes \hat{\mathbf{T}}_{\mathbf{xx}}^{-1}.$$

All necessary partial derivatives are found. The first matrix derivative of $\hat{t}_{greg}$ with respect to $\mathbf{W}$ at the point $\mathbf{W}_0 = (t_y, \boldsymbol{t}_{\mathbf{x}}, \mathbf{T}_{\mathbf{x}y}, \mathbf{T}_{\mathbf{xx}})$ is

$$\left.\frac{d\,\hat{t}_{greg}}{d\,\mathbf{W}}\right|_{\mathbf{W}=\mathbf{W}_0} = \left.\begin{pmatrix} \frac{d\,\hat{t}_{greg}}{d\,\hat{t}_y} \\ \frac{d\,\hat{t}_{greg}}{d\,\hat{\boldsymbol{t}}_{\mathbf{x}}} \\ \frac{d\,\hat{t}_{greg}}{d\,\hat{\mathbf{T}}_{\mathbf{x}y}} \\ \frac{d\,\hat{t}_{greg}}{d\,\hat{\mathbf{T}}_{\mathbf{xx}}} \end{pmatrix}\right|_{\mathbf{W}=\mathbf{W}_0} = \begin{pmatrix} 1 \\ -\mathbf{T}_{\mathbf{xx}}^{-1}\mathbf{T}_{\mathbf{x}y} \\ \mathbf{0}_{p\times 1} \\ \mathbf{0}_{p^2\times 1} \end{pmatrix}. \qquad A.3$$

The second matrix derivative with respect to $\mathbf{W}$ has the structure

$$\frac{d^2\,\hat{t}_{greg}}{d\,\mathbf{W}^2} = \begin{pmatrix} \frac{d}{d\,\hat{t}_y}(\mathrm{I}) & \frac{d}{d\,\hat{t}_y}(\mathrm{II}) & \frac{d}{d\,\hat{t}_y}(\mathrm{III}) & \frac{d}{d\,\hat{t}_y}(\mathrm{IV}) \\ \frac{d}{d\,\hat{\boldsymbol{t}}_{\mathbf{x}}}(\mathrm{I}) & \frac{d}{d\,\hat{\boldsymbol{t}}_{\mathbf{x}}}(\mathrm{II}) & \frac{d}{d\,\hat{\boldsymbol{t}}_{\mathbf{x}}}(\mathrm{III}) & \frac{d}{d\,\hat{\boldsymbol{t}}_{\mathbf{x}}}(\mathrm{IV}) \\ \frac{d}{d\,\hat{\mathbf{T}}_{\mathbf{x}y}}(\mathrm{I}) & \frac{d}{d\,\hat{\mathbf{T}}_{\mathbf{x}y}}(\mathrm{II}) & \frac{d}{d\,\hat{\mathbf{T}}_{\mathbf{x}y}}(\mathrm{III}) & \frac{d}{d\,\hat{\mathbf{T}}_{\mathbf{x}y}}(\mathrm{IV}) \\ \frac{d}{d\,\hat{\mathbf{T}}_{\mathbf{xx}}}(\mathrm{I}) & \frac{d}{d\,\hat{\mathbf{T}}_{\mathbf{xx}}}(\mathrm{II}) & \frac{d}{d\,\hat{\mathbf{T}}_{\mathbf{xx}}}(\mathrm{III}) & \frac{d}{d\,\hat{\mathbf{T}}_{\mathbf{xx}}}(\mathrm{IV}) \end{pmatrix},$$

and at the point $\mathbf{W}_0$, it is

$$\left.\frac{d^2\,\hat{t}_{greg}}{d\,\mathbf{W}^2}\right|_{\mathbf{W}=\mathbf{W}_0} = \begin{pmatrix} 0 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{T}_{\mathbf{xx}}^{-1} & \mathbf{T}_{\mathbf{xx}}^{-1} \otimes \mathbf{T}_{\mathbf{x}y}'\mathbf{T}_{\mathbf{xx}}^{-1} \\ \mathbf{0} & -\mathbf{T}_{\mathbf{xx}}^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_{\mathbf{xx}}^{-1}\mathbf{T}_{\mathbf{x}y} \otimes \mathbf{T}_{\mathbf{xx}}^{-1} & \mathbf{0} & \mathbf{0} \end{pmatrix},$$

where $\mathbf{0}$ denotes a zero-block of a suitable dimension.

According to (1.51) the Taylor expansion of $\hat{t}_{greg}$ up to the second term is

$$\hat{t}_{greg,\,sec} = \boldsymbol{f}(\mathbf{W}_0) + \left.\frac{d\,\hat{t}_{greg}}{d\,\mathbf{W}}\right|_{\mathbf{W}=\mathbf{W}_0}' \operatorname{vec}(\mathbf{W} - \mathbf{W}_0)$$

$$+\frac{1}{2}\operatorname{vec}'(\mathbf{W} - \mathbf{W}_0)\left.\frac{d^2\,\hat{t}_{greg}}{d\,\mathbf{W}^2}\right|_{\mathbf{W}=\mathbf{W}_0}' \operatorname{vec}(\mathbf{W} - \mathbf{W}_0). \qquad A.4$$

For simplicity, let us denote $\operatorname{vec}(\mathbf{W} - \mathbf{W}_0) = (w_1, \boldsymbol{w}_2', \boldsymbol{w}_3', \boldsymbol{w}_4')'$, where $\boldsymbol{w}_2$ and $\boldsymbol{w}_3$ are of size $p \times 1$ and $\boldsymbol{w}_4$ is of size $p^2 \times 1$, with definitions derived from $(A.2)$.

We use the derivative $(A.3)$ for the second summand in $(A.4)$,

$$\left.\frac{d\,\hat{t}_{greg}}{d\,\mathbf{W}}\right|'_{\mathbf{W}=\mathbf{W}_0} \cdot \mathrm{vec}(\mathbf{W}-\mathbf{W}_0) = w_1 - \mathbf{T}'_{\mathbf{x}y}\mathbf{T}^{-1}_{\mathbf{xx}}\boldsymbol{w}_2$$

$$= (\hat{t}_y - t_y) - \mathbf{T}'_{\mathbf{x}y}\mathbf{T}^{-1}_{\mathbf{xx}}(\hat{\boldsymbol{t}}_{\mathbf{x}} - \boldsymbol{t}_{\mathbf{x}}). \qquad A.5$$

The second derivative of $\hat{t}_{greg}$ has many zero-blocks, therefore the third summand in $(A.4)$ simplifies a lot after multiplication with $\mathrm{vec}(\mathbf{W}-\mathbf{W}_0)$. What remains is

$$L = \frac{1}{2}\left\{-\boldsymbol{w}_3'\mathbf{T}^{-1}_{\mathbf{xx}}\boldsymbol{w}_2 + \boldsymbol{w}_4'\left(\mathbf{T}^{-1}_{\mathbf{xx}} \otimes \mathbf{T}^{-1}_{\mathbf{xx}}\mathbf{T}_{\mathbf{x}y}\right)\boldsymbol{w}_2\right.$$

$$\left. -\boldsymbol{w}_2'\mathbf{T}^{-1}_{\mathbf{xx}}\boldsymbol{w}_3 + \boldsymbol{w}_2'\left(\mathbf{T}'_{\mathbf{x}y}\mathbf{T}^{-1}_{\mathbf{xx}} \otimes \mathbf{T}^{-1}_{\mathbf{xx}}\right)\boldsymbol{w}_4\right\}. \qquad A.6$$

It simplifies further, since its terms are pairwise equal. To see this, note that the terms of L are scalars and can be transposed. Therefore,

$$\boldsymbol{w}_3'\mathbf{T}^{-1}_{\mathbf{xx}}\boldsymbol{w}_2 = \boldsymbol{w}_2'\mathbf{T}^{-1}_{\mathbf{xx}}\boldsymbol{w}_3.$$

For transposition of the fourth term in L we use (1.6), and since the Kronecker product is not commutative, we use (1.21):

$$\left[\boldsymbol{w}_2'\left(\mathbf{T}'_{\mathbf{x}y}\mathbf{T}^{-1}_{\mathbf{xx}} \otimes \mathbf{T}^{-1}_{\mathbf{xx}}\right)\boldsymbol{w}_4\right]' = \boldsymbol{w}_4'\left(\mathbf{T}^{-1}_{\mathbf{xx}}\mathbf{T}_{\mathbf{x}y} \otimes \mathbf{T}^{-1}_{\mathbf{xx}}\right)\boldsymbol{w}_2$$

$$= \boldsymbol{w}_4'\mathbf{K}_{p,p}\left(\mathbf{T}^{-1}_{\mathbf{xx}} \otimes \mathbf{T}^{-1}_{\mathbf{xx}}\mathbf{T}_{\mathbf{x}y}\right)\mathbf{K}_{1,p}\boldsymbol{w}_2. \qquad A.7$$

Since $\mathbf{K}_{1,p} = \mathbb{I}_p$, and according to the property (1.20),

$$\begin{aligned} \boldsymbol{w}_4'\mathbf{K}_{p,p} &= \mathrm{vec}'(\hat{\mathbf{T}}_{\mathbf{xx}} - \mathbf{T}_{\mathbf{xx}})\mathbf{K}_{p,p} = (\mathbf{K}_{p,p}\mathrm{vec}(\hat{\mathbf{T}}_{\mathbf{xx}} - \mathbf{T}_{\mathbf{xx}}))' \\ &= [\mathrm{vec}(\hat{\mathbf{T}}_{\mathbf{xx}} - \mathbf{T}_{\mathbf{xx}})']' = \mathrm{vec}'(\hat{\mathbf{T}}_{\mathbf{xx}} - \mathbf{T}_{\mathbf{xx}}) = \boldsymbol{w}_4', \end{aligned}$$

we have that the fourth and the second term are equal.

Now we can write L in $(A.6)$ as

$$\begin{aligned} L &= -\boldsymbol{w}_2'\mathbf{T}_{\mathbf{xx}}^{-1}\boldsymbol{w}_3 + \boldsymbol{w}_2'\left(\mathbf{T}_{\mathbf{x}y}'\mathbf{T}_{\mathbf{xx}}^{-1} \otimes \mathbf{T}_{\mathbf{xx}}^{-1}\right)\boldsymbol{w}_4 \\ &= -(\hat{\boldsymbol{t}}_{\mathbf{x}} - \boldsymbol{t}_{\mathbf{x}})'\mathbf{T}_{\mathbf{xx}}^{-1}(\hat{\mathbf{T}}_{\mathbf{x}y} - \mathbf{T}_{\mathbf{x}y}) \\ &\quad +(\hat{\boldsymbol{t}}_{\mathbf{x}} - \boldsymbol{t}_{\mathbf{x}})'\left(\mathbf{T}_{\mathbf{x}y}'\mathbf{T}_{\mathbf{xx}}^{-1} \otimes \mathbf{T}_{\mathbf{xx}}^{-1}\right)\mathrm{vec}(\hat{\mathbf{T}}_{\mathbf{xx}} - \mathbf{T}_{\mathbf{xx}}). \end{aligned} \qquad A.8$$

Finally, we can put together the Taylor expansion $(A.4)$ from $(A.5)$ and $(A.8)$,

$$\begin{aligned} \hat{t}_{greg,\,sec} &= t_y + (\hat{t}_y - t_y - \mathbf{T}_{\mathbf{x}y}'\mathbf{T}_{\mathbf{xx}}^{-1}(\hat{\boldsymbol{t}}_{\mathbf{x}} - \boldsymbol{t}_{\mathbf{x}})) - (\hat{\boldsymbol{t}}_{\mathbf{x}} - \boldsymbol{t}_{\mathbf{x}})'\mathbf{T}_{\mathbf{xx}}^{-1}(\hat{\mathbf{T}}_{\mathbf{x}y} - \mathbf{T}_{\mathbf{x}y}) \\ &+ (\hat{\boldsymbol{t}}_{\mathbf{x}} - \boldsymbol{t}_{\mathbf{x}})'\left(\boldsymbol{t}_{\mathbf{x}y}'\mathbf{T}_{\mathbf{xx}}^{-1} \otimes \mathbf{T}_{\mathbf{xx}}^{-1}\right)\mathrm{vec}(\hat{\mathbf{T}}_{\mathbf{xx}} - \mathbf{T}_{\mathbf{xx}}). \end{aligned}$$

Recalling that $\hat{\mathbf{T}}_{\mathbf{x}y}'\hat{\mathbf{T}}_{\mathbf{xx}}^{-1} = \mathbf{B}'$, the expression (3.24) in Proposition 3.3 is proved.

$\square$

Appendix B

Proof of the Proposition 3.6

Proof. The proof of the expression (3.30) is similar to the derivation in Proposition 3.3. The estimator $\hat{t}_{syn}$ is a linear function of two random matrix arguments,

$$\hat{t}_{syn} = \boldsymbol{f}(\hat{\mathbf{T}}_{\mathbf{x}y}, \hat{\mathbf{T}}_{\mathbf{xx}}).$$

The first order derivatives are the following:

$$\begin{aligned}
\frac{d\,\hat{t}_{syn}}{d\,\hat{\mathbf{T}}_{\mathbf{xx}}} &= \frac{d\left(\hat{\mathbf{T}}'_{\mathbf{x}y}\hat{\mathbf{T}}^{-1}_{\mathbf{xx}}\boldsymbol{t}_{\mathbf{x}}\right)}{d\,\hat{\mathbf{T}}_{\mathbf{xx}}} \overset{(1.45)}{=} \frac{d\,\hat{\mathbf{T}}^{-1}_{\mathbf{xx}}}{d\,\hat{\mathbf{T}}_{\mathbf{xx}}} \cdot \frac{d\left(\hat{\mathbf{T}}'_{\mathbf{x}y}\hat{\mathbf{T}}^{-1}_{\mathbf{xx}}\boldsymbol{t}_{\mathbf{x}}\right)}{d\,\hat{\mathbf{T}}^{-1}_{\mathbf{xx}}} \\
&= (-\hat{\mathbf{T}}^{-1}_{\mathbf{xx}} \otimes \hat{\mathbf{T}}^{-1}_{\mathbf{xx}})(\boldsymbol{t}_{\mathbf{x}} \otimes \hat{\mathbf{T}}'_{\mathbf{x}y}) = -\hat{\mathbf{T}}^{-1}_{\mathbf{xx}}\boldsymbol{t}_{\mathbf{x}} \otimes \hat{\mathbf{T}}^{-1}_{\mathbf{xx}}\hat{\mathbf{T}}'_{\mathbf{x}y} \\
&= -\hat{\mathbf{T}}^{-1}_{\mathbf{xx}}\boldsymbol{t}_{\mathbf{x}} \otimes \hat{\mathbf{B}}. \\
\frac{d\,\hat{t}_{syn}}{d\,\hat{\mathbf{T}}_{\mathbf{x}y}} &= \frac{d\left(\hat{\mathbf{T}}'_{\mathbf{x}y}\hat{\mathbf{T}}^{-1}_{\mathbf{xx}}\boldsymbol{t}_{\mathbf{x}}\right)}{d\,\hat{\mathbf{T}}_{\mathbf{x}y}} = \frac{d\,\hat{\mathbf{T}}'_{\mathbf{x}y}}{d\,\hat{\mathbf{T}}_{\mathbf{x}y}} \cdot \frac{d\left(\hat{\mathbf{T}}'_{\mathbf{x}y}\hat{\mathbf{T}}^{-1}_{\mathbf{xx}}\boldsymbol{t}_{\mathbf{x}}\right)}{d\,\hat{\mathbf{T}}'_{\mathbf{x}y}} \\
&\overset{(1.47)}{=} \mathbf{K}_{1,p}(\hat{\mathbf{T}}^{-1}_{\mathbf{xx}}\boldsymbol{t}_{\mathbf{x}} \otimes 1) = \hat{\mathbf{T}}^{-1}_{\mathbf{xx}}\boldsymbol{t}_{\mathbf{x}}.
\end{aligned}$$

Denoting $\mathbf{W}_0 = (\mathbf{T}_{\mathbf{x}y}, \mathbf{T}_{\mathbf{xx}})$, the first order derivative of $\boldsymbol{f}$ at the point $\mathbf{W}_0$ is

$$\left.\frac{d\,\hat{t}_{syn}}{d\,\mathbf{W}}\right|_{\mathbf{W}=\mathbf{W}_0} = \left.\begin{pmatrix} \frac{d\,\hat{t}_{syn}}{d\,\hat{\mathbf{T}}_{\mathbf{x}y}} \\ \frac{d\,\hat{t}_{syn}}{d\,\hat{\mathbf{T}}_{\mathbf{xx}}} \end{pmatrix}\right|_{\mathbf{W}=\mathbf{W}_0} = \begin{pmatrix} -\mathbf{T}^{-1}_{\mathbf{xx}}\boldsymbol{t}_{\mathbf{x}} \otimes \mathbf{B} \\ \mathbf{T}^{-1}_{\mathbf{xx}}\boldsymbol{t}_{\mathbf{x}} \end{pmatrix}.$$

The vectorized form of $\mathbf{W} - \mathbf{W}_0$ is

$$\operatorname{vec}(\mathbf{W} - \mathbf{W}_0) = \begin{pmatrix} \operatorname{vec}(\hat{\mathbf{T}}_{\mathbf{xx}} - \mathbf{T}_{\mathbf{xx}}) \\ \hat{\mathbf{T}}_{\mathbf{x}y} - \mathbf{T}_{\mathbf{x}y} \end{pmatrix}.$$

Now the formula (1.51) of the Taylor expansion results in

$$\begin{aligned}
\hat{t}_{syn,\,lin} &= \mathbf{B}'\boldsymbol{t}_{\mathbf{x}} + \left.\left(\frac{d\,\hat{t}_{syn}}{d\,\mathbf{W}}\right)'\right|_{\mathbf{W}=\mathbf{W}_0} \operatorname{vec}(\mathbf{W} - \mathbf{W}_0) \\
&= \mathbf{B}'\boldsymbol{t}_{\mathbf{x}} + (-\boldsymbol{t}'_{\mathbf{x}}\mathbf{T}^{-1}_{\mathbf{xx}} \otimes \mathbf{B}',\ \boldsymbol{t}'_{\mathbf{x}}\mathbf{T}^{-1}_{\mathbf{xx}}) \begin{pmatrix} \operatorname{vec}(\hat{\mathbf{T}}_{\mathbf{xx}} - \mathbf{T}_{\mathbf{xx}}) \\ \hat{\mathbf{T}}_{\mathbf{x}y} - \mathbf{T}_{\mathbf{x}y} \end{pmatrix}.
\end{aligned}$$

After multiplying matrices, we get

$$\begin{aligned}\hat{t}_{syn,\,lin} &= \mathbf{B}'\boldsymbol{t}_{\mathbf{x}} - (\boldsymbol{t}'_{\mathbf{x}}\mathbf{T}^{-1}_{\mathbf{xx}} \otimes \mathbf{B}')\text{vec}(\hat{\mathbf{T}}_{\mathbf{xx}} - \mathbf{T}_{\mathbf{xx}}) \\ &+ \boldsymbol{t}'_{\mathbf{x}}\mathbf{T}^{-1}_{\mathbf{xx}}\hat{\mathbf{T}}_{\mathbf{x}y} - \boldsymbol{t}'_{\mathbf{x}}\mathbf{T}^{-1}_{\mathbf{xx}}\mathbf{T}_{\mathbf{x}y}. \qquad B.1\end{aligned}$$

Since $\mathbf{B}'\boldsymbol{t}_{\mathbf{x}} = \boldsymbol{t}'_{\mathbf{x}}\mathbf{B} = \boldsymbol{t}'_{\mathbf{x}}\mathbf{T}^{-1}_{\mathbf{xx}}\mathbf{T}_{\mathbf{x}y}$, the terms $\mathbf{B}'\boldsymbol{t}_{\mathbf{x}}$ and $\boldsymbol{t}'_{\mathbf{x}}\mathbf{T}^{-1}_{\mathbf{xx}}\mathbf{T}_{\mathbf{x}y}$ cancel out in (B.1). The second term in (B.1) simplifies with the property (1.14) of the vec() operator, the vec() can be omitted if applied to a scalar:

$$\begin{aligned}(\boldsymbol{t}'_{\mathbf{x}}\mathbf{T}^{-1}_{\mathbf{xx}} \otimes \mathbf{B}')\text{vec}(\hat{\mathbf{T}}_{\mathbf{xx}} - \mathbf{T}_{\mathbf{xx}}) &= \text{vec}\left[\mathbf{B}'(\hat{\mathbf{T}}_{\mathbf{xx}} - \mathbf{T}_{\mathbf{xx}})\mathbf{T}^{-1}_{\mathbf{xx}}\boldsymbol{t}_{\mathbf{x}}\right] \\ &= \mathbf{B}'(\hat{\mathbf{T}}_{\mathbf{xx}} - \mathbf{T}_{\mathbf{xx}})\mathbf{T}^{-1}_{\mathbf{xx}}\boldsymbol{t}_{\mathbf{x}} \\ &= \mathbf{B}'\hat{\mathbf{T}}_{\mathbf{xx}}\mathbf{T}^{-1}_{\mathbf{xx}}\boldsymbol{t}_{\mathbf{x}} - \mathbf{B}'\boldsymbol{t}_{\mathbf{x}}.\end{aligned}$$

Finally, we get

$$\hat{t}_{syn,\,lin} = -\mathbf{B}'\hat{\mathbf{T}}_{\mathbf{xx}}\mathbf{T}^{-1}_{\mathbf{xx}}\boldsymbol{t}_{\mathbf{x}} + \mathbf{B}'\boldsymbol{t}_{\mathbf{x}} + \boldsymbol{t}'_{\mathbf{x}}\mathbf{T}^{-1}_{\mathbf{xx}}\hat{\mathbf{T}}_{\mathbf{x}y},$$

which is equivalent to (3.30).

$\square$

Appendix C

Empirical MSEs of the initial and GR estimators

Table 5.1: Empirical MSEs between estimators for SI case

	$\hat{\theta}_{GR1}$	$\hat{\theta}_{GR3}$	$\hat{\theta}_{GR2}$	$\hat{\theta}$
Continuous variable, $\times 10^5$				
Domain 1	**6958.4**	**6954.1**	**7007.6**	**7191.7**
Domain 1, Domain 2	-316.2	-307.1	-259.2	39.2
Domain 1, Domain 3	1578.8	1573.8	1039.6	1249.0
Domain 1, Domain 4	328.1	331.3	600.7	360.8
Domain 1, Population	8549.1	8552.0	8388.7	7527.3
Domain 2	**7428.7**	**7453.3**	**7494.4**	**7923.4**
Domain 2, Domain 3	1723.2	1551.9	1094.0	1412.7
Domain 2, Domain 4	330.9	417.6	648.8	282.4
Domain 2, Population	9166.6	9115.6	8978.0	7658.4
Domain 3	**854.1**	**1608.6**	**6707.6**	**6937.9**
Domain 3, Domain 4	152.2	-234.3	-2808.8	-3073.2
Domain 3, Population	4308.3	4500.0	6032.4	5081.5
Domain 4	**33.0**	**230.8**	**1530.7**	**1834.0**
Domain 4, Population	844.1	745.4	-28.7	1062.1
Population	**22868.2**	**22913.0**	**23370.3**	**27293.2**
Binary variable				
Domain 1	**73406.9**	**77928.0**	**79677.8**	**80744.9**
Domain 1, Domain 2	-5716.7	-2919.1	-1931.7	-1511.7
Domain 1, Domain 3	14305.3	4905.7	2140.7	12185.9
Domain 1, Domain 4	3636.7	2830.2	2611.8	3350.7
Domain 1, Population	85632.2	82744.8	82498.6	81405.9
Domain 2	**30487.7**	**32061.2**	**32587.8**	**32751.5**
Domain 2, Domain 3	6052.5	577.8	-1004.2	3042.6
Domain 2, Domain 4	1232.7	765.5	638.5	935.7
Domain 2, Population	32056.2	30485.4	30290.5	29834.1
Domain 3	**5061.2**	**23862.5**	**28216.2**	**117218.1**
Domain 3, Domain 4	1006.5	2614.3	2956.9	9534.2
Domain 3, Population	26425.6	31960.4	32309.7	23584.7
Domain 4	**266.5**	**404.0**	**430.8**	**916.7**
Domain 4, Population	6142.5	6614.0	6638.1	5987.7
Population	**150256.5**	**151804.7**	**151736.8**	**152417.1**

Table 5.2: Empirical MSEs between estimators for MB case

Continuous variable, $\times 10^5$				
	$\hat{\boldsymbol{\theta}}_{\boldsymbol{GR1}}$	$\hat{\boldsymbol{\theta}}_{\boldsymbol{GR3}}$	$\hat{\boldsymbol{\theta}}_{\boldsymbol{GR2}}$	$\hat{\boldsymbol{\theta}}$
Domain 1	**7674.7**	**7675.0**	**7717.4**	**7966.4**
Domain 1, Domain 2	-514.1	-492.5	-430.8	10.4
Domain 1, Domain 3	1683.9	1656.6	1172.4	1294.5
Domain 1, Domain 4	355.2	371.5	618.1	403.6
Domain 1, Population	9199.7	9210.5	9077.1	8077.4
Domain 2	**8522.9**	**8573.4**	**8660.0**	**9354.6**
Domain 2, Domain 3	1919.6	1616.0	895.5	1150.8
Domain 2, Domain 4	379.5	534.8	900.6	496.5
Domain 2, Population	10307.9	10231.7	10025.3	8228.0
Domain 3	**926.4**	**1582.2**	**7027.6**	**7069.9**
Domain 3, Domain 4	167.4	-194.8	-2973.8	-3068.2
Domain 3, Population	4697.3	4660.0	6121.6	5643.2
Domain 4	**37.2**	**235.5**	**1653.3**	**1831.5**
Domain 4, Population	939.3	947.1	198.3	1052.3
Population	**25144.2**	**25049.3**	**25422.3**	**29428.0**
Binary variable				
Domain 1	**77557.5**	**83687.2**	**85150.2**	**87316.8**
Domain 1, Domain 2	-4648.4	-780.8	150.9	1240.6
Domain 1, Domain 3	15188.2	4108.4	1332.6	15167.1
Domain 1, Domain 4	3940.6	3038.9	2809.0	3857.4
Domain 1, Population	92037.8	90053.6	89442.6	88871.3
Domain 2	**37970.3**	**40312.6**	**40880.3**	**41428.3**
Domain 2, Domain 3	8180.5	1279.4	-448.5	6508.7
Domain 2, Domain 4	1658.3	1094.8	951.7	1479.0
Domain 2, Population	43160.6	41906.0	41534.4	41247.3
Domain 3	**5758.0**	**25703.0**	**30907.8**	**119092.0**
Domain 3, Domain 4	1161.8	2786.6	3217.7	9901.2
Domain 3, Population	30288.4	33877.4	35009.5	31398.5
Domain 4	**305.9**	**438.2**	**473.9**	**980.5**
Domain 4, Population	7066.5	7358.6	7452.3	7178.4
Population	**172553.4**	**173195.6**	**173438.8**	**173580.7**

Bibliography

[1] Cassel, C.M., Särndal, C.E., and Wretman, J.H. (1976). Some results on generalized difference estimation and generalized regression estimation for finite populations. *Biometrica*, vol. 63, pp. 615-620

[2] Cochran, W.G. (1977) *Sampling Techniques. Third Edition.* New York: Wiley

[3] Deng, L.-Y., Chhikara, R.S. (1990) On the ratio and regression estimator in finite population sampling. *The American statistician* vol. 44, pp. 282-284

[4] Dever, J.A., Valliant, R.L. (2010) A comparison of variance estimators for poststratification to estimated control totals. *Survey Methodology*, 36(1), pp. 45-56.

[5] Estevao, V.M., Särndal, C.-E. (2004) Borrowing strength is not the best technique within a wide class of design-consistent domain estimators. *Journal of Official Statistics* vol. 20, pp. 645-669

[6] Harville, D.A. (1997) *Matrix algebra from a statistician's perspective.* Springer-Verlag, Berlin, Heidelberg, New York

[7] Isaki, C.T., and Fuller, W.A. (1982) Survey design under the regression superpopulation model. *Journal of the American Statistical Association* vol. 77, pp. 89-96

[8] Knottnerus, P. (2003) *Sample Survey Theory. Some Pythagorean Perspectives.* Wiley, New York

[9] Knottnerus, P., van Duin, C. (2006). Variances in Repeated Weighting With an Application to the Dutch Labour Force Survey. *Journal of Official Statistics*, 22, pp. 565-584.

[10] Kollo, T., von Rosen, D. (2005) *Advanced Multivariate Statistics with Matrices.* Springer, Berlin, Heidelberg, New York

[11] Kroese, A.H., Renssen, R.H. (1999). Weighting and Imputation at Statistics Netherland. *Proceedings of the IASS Conference on Small Area Estimation*, Riga, 109-120.

[12] Lehtonen, R., Pahkinen E. (2004) *Practical Methods for Design and Analysis of Complex Surveys. Second Edition.* John Wiley & Sons

[13] Lehtonen, R., Särndal, C.-E., Veijanen, A. (2003) The effect of model choice in estimation for domains, including small domains. *Survey Methodology*, 29, 33-44

[14] Lehtonen, R., Särndal, C.-E., Veijanen, A. (2005) Does the model matter? Comparising model-assisted and model-dependent estimators of class frequencies for domains. *Statistics in Transition*, 7, 649-673

[15] Lepik, N. (2007) On the bias of the generalized regression estimator in survey sampling. *Acta Applicandae Mathematicae* vol. 97, pp. 41-52

[16] Lütkepohl, H. (1996) *Handbook of Matrices.* John Wiley & Sons, Chichester, New York

[17] Meister, K. (2004) *On Methods for Real Time Sampling and Distributions in Sampling. Doctoral Dissertation.* Umea, 2004

[18] Puntanen, S., Styan, G.P.H. (2004) *Matrix trics for linear statistical models: our personal Top Thirteen.* Tampere

[19] Rajaleid, K. (2004) Multivariate finite population inference under the assumption of linear pattern in the population. *Acta et Commentationes Universitatis Tartuensis de Mathematica* vol. 8, pp. 235-242

[20] Rao, C.R., Rao, M.B. (1998) *Matrix Algebra and Its Applications to Statistics and Econometrics.* World Scientific, Singapore

[21] Rao, J. N. K. (2003) *Small area estimation.* John Wiley & Sons, Hoboken, New Jersey

[22] Renssen, R.H., Nieuwenbroek, N.J. (1997), Aligning Estimates for Common Variables in two or More Sample Surveys, *Journal of the American Statistical Association*, 92, 368-374.

[23] Sõstra, K. (2007) *Restriction estimation for domains. Doctoral Dissertation.* Tartu

[24] Sõstra, K., Traat, I. (2009) Optimal domain estimation under summation restriction. *Journal of Statistical Planning and Inference* vol. 139, pp. 3928-3941

[25] Särndal, C.-E. (1980) On π inverse weighting versus best linear unbiased weighting in probability sampling. *Biometrica* vol. 67, pp. 639-650

[26] Särndal, C.-E. (1982) Implications of survey design for generalized regression estimation of linear funcions. *Journal of Statistical Planning and Inference* vol. 7, pp. 155-170

[27] Särndal, C.-E. (2007) The calibration approach in survey theory and practicw *Survey Methodology* vol.33, pp. 99-119

[28] Särndal, C.-E., Swensson, B., Wretman, J. (1992) *Model Assisted Survey Sampling.* New York: Springer-Verlag

[29] Tillé, Y. (2006) *Sampling Algorithms.* New-York: Springer-Verlag

[30] Traat, I. (2000) Sampling design as a multivariate distribution. *New trends in Probability and Statistics Processes*, vol. 7(23), pp. 301-316

[31] Traat, I., Ilves, M. (2007) The hypergeometric sampling design, theory and practice. *Acta Applicandae Mathematicae* vol. 97, pp. 311-321

[32] Traat, I., Bondesson, L., Meister, K. (2004) Sampling design and sample selection through distribution theory. *Journal of Statistical Planning and Inference*, vol. 123, 395-413

[33] Traat, I., Meister, K., Sõstra, K. (2001) Statistical inference in sampling theory. *Theory of Stochastic Processes*, vol. 7(23), pp. 301-316

[34] Traat, I., Särndal, C.E. (2009). Domain Estimators Calibrated on Information from Other Surveys. *Research Report* No. 2009-1, Vol. 15, Department of Mathematics and Mathematical Statistics, Umea University, Sweden.

[35] Yung, W., Rubin-Bleuer, S. (2007) The Survey of Employment, Payrolls and Hours: Improving the Stability of the GREG Estimator. *Advisory Committee on Statistical Methods*, Meeting No. 45

[36] Zieschang, K.D. (1990), Sample Weighting Methods and Estimation of Totals in the Consumer Expenditure Survey. *Journal of the American Statistical Association*, 85, 986-1001.

Osakogumite hindamine kitsenduste olemasolul baseerudes üldistatud regressioon- ja sünteetilisele hinnangule

Kokkuvõte

Tänapäeval on nõudmine usaldusväärse statistika järele oluliselt kasvanud. Seejuures vajatakse näitajaid üha detailsemal tasemel – mitmesuguste osakogumite tasemel. Vajalikke näitajaid saadakse nii valikuuringutest kui ka erinevatest registritest. Soovitavateks näitajateks on tunnuseväärtuste kogusummad. Registrite korral võib juhtuda, et isegi kui nad sisaldavad huvipakkuvaid tunnuseid, puuduvad sealt huvipakkuvate osakogumite identifikaatorid, mistõttu ei saa osakogumite kogusummasid sealt arvutada. Teisalt võib sama tunnus olla küsitletud valikuuringus ja seda koos osakogumi identifikaatoritega. Võimalus leida osakogumi hinnangud valikuuringust tekitab kooskõlalisuse probleemi: valikuuringust saadud hinnangud ei summeeru üldkogumi või vastavate suuremate osakogumite summadeks, mis on välja võetud registrist. Kooskõlalisuse probleem kerkib esile ka mitme paralleelselt läbiviidava uuringu korral, mis sisaldavad ühiseid uuritavaid tunnuseid. Veel üks ebakooskõlalisuse olukord tekib siis, kui ühe ja sama uuringu raames on erinevate osakogumite parameetrid hinnatud erinevate hinnangufunktsioonide või meetodite abil, põhjuseks kas täpsuse tõstmine osakogumis või olemasolevad praktilised võimalused. Ka sellisel juhul võib osakogumi hinnanguid summeerides saada erineva tulemuse vastava suurema osakogumi või üldkogumi hinnangust.

Teoreetikute jaoks on hinnangute ebakooskõlalisus loomulik nähtus; seosed, mis kehtivad üldkogumi parameetrite jaoks, ei pruugi kehtida valimist saadud hinnangute jaoks, seda viimaste juhusliku loomu tõttu. See nähtus häirib aga statistiliste näitajate tarbijaid. Samas on teada, et lisainformatsiooni haaramine hinnangusse võib tõsta selle täpsust. Ka seoseid üldkogumi parameetrite vahel võib vaadelda lisateabena. Siit tekkis mõte, et kui kaasata antud lisateave hinnangufunktsiooni konstrueerimisse, siis ehk oleks võimalik saavutada kaks eesmärki korraga: tõsta hinnangute täpsust ja lahendada kooskõlalisuse probleem hinnangute vahel. Antud töö põhiteemaks ongi osakogumite hinnangute väljatöötamine, mis on kooskõlalised ja parema täpsusega võrreldes esialgsete hinnangutega.

Kooskõlalisuse probleem pole valikuuringute valdkonnas uus, seda on uuritud juba mõnda aega. Hinnangute kooskõla kahe erineva uuringu vahel või valikuuringu ja registri vahel on püütud saavutada kalibreerimismeetoditega (Zieschang 1990, Renssen ja Nieuwenbroek 1997, Traat ja Särndal 2009, Dever ja Valliant 2010). Nendes meetodites on kooskõlalisuse nõue lisatud kalibreerimise kitsendustesse. Teised autorid (Kroese ja Renssen 1999, Knottnerus ja Van Duin 2006) kasutasid teistsugust kalibreerimistehnikat, nn korduvkaalumise tehnikat, kus juba leitud hinnangud kalibreeritakse uue informatsiooni ilmumisel ümber.

Antud dissertatsiooni ideed pärinevad meetodist, mis on esitatud raamatus Knottnerus (2003). Sealne üldine kitsendustega hinnang (General Restriction estimator, lühidalt GR) baseerub nihketa esialgsetele hinnangutele ja rahuldab lineaarseid kitsendusi. Saadud GR-hinnangul on mitmeid häid omadusi, mille hulgas on hinnangu nihketus ja väiksem dispersioon võrreldes esialgsete hinnangutega. GR-hinnang on ka optimaalne teatud hinnangute klassis, mis rahuldavad antud kitsendusi. Knottneruse GR-hinnang ei olnud välja töötatud osakogumite jaoks. Dissertatsioonis Sõstra (2007) arendati GR-hinnang välja osakogumite kooskõlaliseks hindamiseks ning hiljem täiendati seda artiklis Sõstra ja Traat (2009).

Ülalnimetatud töödes keskendutakse nihketa hindamisele, see tähendab et aluseks on nihketa lähtehinnangud ja tulemuseks on nihketa kuid kitsendusi rahuldavad hinnangud. Käesolevas dissertatsioonis on GR-hinnangut üldistatud nii, et ta on rakendatav ka nihkega alghinnangutele. Osakogumite hindamiseks kasutatakse sageli hinnanguid, mis võivad omada nihet. Näiteks kasutatakse (seda eriti just väikeste osakogumite korral) mudelipõhiseid hinnangud (Rao, 2003), sünteetilist ehk projektsioonhinnangut (Särndal jt. 1992, lk. 408-412, Yung ja Rubin-Bleuer, 2007). Kuigi need hinnangud on nihkega, on nende positiivseks omaduseks väike varieeruvus. Senini pole uuritud, kuidas nihkega lähtehinnangud mõjutavad kitsendusi rahuldavat GR-hinnangut.

Antud töös keskendutakse osakogumite ja üldkogumi hindamisele summeeruvuskitsenduse olemasolul. Nii nihketa kui ka nihkega lähtehinnangud on lubatud. Lähtehinnangute rolli on valitud üldistatud regressioon- (Generalized Regression, lühidalt GREG) ja sünteetiline (SYN) hinnang. Mõlemad hinnangud võivad osakogumites olla üles ehitatud erinevate mudelite eeldusel (Lehtonen ja Pahkinen 2004, lk. 187-213). Vaadeldavateks mudeliteks on üldkogumitaseme ehk P-mudel ja osakogumitaseme ehk D-mudel. Hinnanguid, mis on konstrueeritud D-mudeli abil, nimetatakse kirjanduses otsesteks hinnanguteks, ja P-mudeli abil - kaudseteks hinnanguteks. Antud töös on vaadeldud GREG ja SYN hinnangute omadusi, on tuletatud nende hinnangute nihked ja ruutkeskmised vead (Mean Square Error, lühidalt MSE), samuti vastastikused ruutkeskmised vead.

Käesoleva dissertatsiooni eesmärgid, mis töö käigus ka realiseeriti, olid järgmised.

1. Tuletada uued üldisemad osakogumite GR-hinnangud, mis rahuldavad summeeruvuskitsendust. Üldistusena lubati nihkega lähtehinnangud.
2. Tuletada saadud GR-hinnangute nihked ja MSEd. Tuginedes saadud avaldistele näidata, et GR-hinnangud on täpsemad kui esialgsed.
3. Uurida osakogumite GREG ja SYN hinnangute, mis on valitud lähtehinnangute rolli, omadusi.
4. Tuletada GREG ja SYN hinnangute lineariseeritud kujud, millest tuletada hinnangute nihke ja MSE ligikaudsed avaldised, seda nii osakogumite kui ka üldkogumi korral. Osakogumite juhul arvestada nii D- kui ka P-mudelitega.
5. Tuletada vastastikused MSEd erinevate hinnangute vahel.
6. Uurida piisavat tingimust GREG ja SYN hinnangute kokkulangemiseks nii üldkogumi kui ka osakogumite korral.
7. Hinnangute omadused töötada välja üldkujul, mis kehtiksid suvalise valikudisaini jaoks. Erijuhtudena vaadelda kahte valikudisaini, lihtsat juhuvalikut ja multinomiaalset disaini.

8. Illustreerida teoreetilisi tulemusi simuleerimisülesandes reaalsete andmete põhjal, ja veenduda tulemuste rakendatavuses.

Osakogumite hindamine on mitmemõõtmeline probleem, vaatluse all on korraga palju osakogumeid ja hinnang on tegelikult hinnangute vektor. Seepärast on püstitatud ülesannete lahendamiseks loomulik kasutada maatriksaparatuuri, mida ongi käesolevas töös tehtud. Hinnangute vektori täpsust on mõõdetud MSE-maatriksi abil.

Dissertatsioonis on kasutatud disainipõhist lähenemist, mille kohaselt on hinnangute omadused määratud valikudisaini poolt ja ka uuritava tunnuse väärtuste poolt lõplikus üldkogumis. Valikudisaini on käsitletud mitmemõõtmelise jaotusena ja valim on realisatsioon sellest jaotusest. Töös on eeldatud, et osakogumite valimid pole liiga väikesed.

Esimeses peatükis antakse maatriksite teooria vajalikud mõisted ja omadused, samuti valikuuringute teooria disaini-põhise lähenemise alused. Siin on toodud ka hinnangute kovariatsiooni- ja MSE-maatriksite definitsioonid ning omadused.

Teises peatükis on toodud käesoleva dissertatsiooni põhitulemused, mis puudutavad osakogumite hindamist nihkega lähtehinnangute ja summeeruvuskitsenduse olemasolul. Näite abil demonstreeritakse, et kui Knottneruse (2003) GR-hinnangu konstruktsiooni rakendada nihkega lähtehinnangutele, siis tulemusena saadud hinnang pole täpsem kui esialgne. Selles peatükis pakutakse välja kolm uut GR-hinnangut ja näidatakse, et nende MSEd ei ole suuremad kui lähtehinnangu oma. Tuletatakse ka GR-hinnangute ja lähtehinnangu järjestus MSE- maatriksite suhtes ja on leitud hinnangute hulgas parim hinnang.

Kolmanda peatüki tähelepanu on keskendunud lähtehinnangute klassidele, milleks on GREG ja SYN hinnangud, nii üldkogumile kui ka osakogumitele. Töös on kirjeldatud situatsioone ja tingimusi, millal need kaks hinnangut on võrdsed. Samuti on tuletatud hinnangute Taylori read. Lineariseeritud hinnangutele tuginedes on leitud hinnangute ligikaudsed nihked, kovariatsiooni- ja MSE-maatriksid. Osakogumite korral on uuritud nii D- kui ka P-mudelid. Antud töö seisukohalt oli oluline leida lähtehinnangute MSE-maatriks. Seepärast tuletati ka vastastikused MSEd erinevate hinnangute vahel. Näidati, et teatud tingimuste korral need MSEd lihtsustuvad, näiteks olid osakogumite GREG-D hinnangud ligikaudu mittekorreleeritud.

Neljandas peatükis on kirjeldatud simuleerimiseksperimenti ja on esitatud selle tulemused Kasutati Eesti meditsiiniasutuste personali andmed. Eesmärgiks oli illustreerida GR-hinnangute käitumist praktilises situatsioonis. Üldkogum mahuga 21764 inimest jaotati nelja erineva suurusega osakogumisse ning seejärel võeti kogu üldkogumist korduvalt valimeid mahuga 400 kasutades lihtsat juhuvalikut (tagasipanekuta) ning multinomiaaldisaini. Uuritavaid tunnuseid oli kaks - tunnitasu pideva tunnuse, ja arst/mitte binaarse tunnuse esindajana. Kõigepealt on töös illustreeritud lähtehinnangute ebakooskõla probleemi nii tabeli kui ka graafikute abil. Seejärel on demonstreeritud, et kõik kolm GR-hinnangut rahuldavad kitsendusi. Samuti on leitud esialgsete ja GR-hinnangute MSE-maatriksid, seda nii tuletatud valemite abil kui ka empiiriliselt üle korduvate valimite. Tuletame meelde, et tuletatud valemid kehtivad asümptootiliselt. Graafikute abil on näidatud, et asümptootilised MSEd on väga lähedased empiirilistele, seda isegi üsna väikese osakogumivalimi korral. Tabelite abil on illustreeritud töö teoreetilises osas tõestatud MSE-võrratuste kehtivust. Samuti ka näidatud, et kõige kolme GR-hinnangu MSEd on väiksemad kui esialgse hinnangu oma.

Simuleerimiskatse käigus sai kinnitust ka valikuuringute valdkonnas teadaolev fakt, et SYN-P hinnang on tunduvalt väiksema varieeruvusega kui GREG, kuid võib kaasa tuua suure nihke. Näitasime (Proposition 3.19), et see juhtub nendes osakogumites, kus üldkogumi mudel erineb osakogumi omast.

Simuleerimisülesanne näitas, et osakogumites, kus nihe polnud eriti suur, on hinnangute GR1 ja GR3 MSEd tunduvalt väiksemad kui hinnangul GR2. Samas nõutakse mõlema hinnangu valemis esialgse nihke teadmist, mis aga pole praktikas teada. Hinnang GR2 on vaba sellest eeldusest ja demonstreeris väga häid tulemusi (MSE mõttes) just nendes osakogumites, kus esialgne nihe oli suur.

Töös tuletatud kolm GR-hinnangut on uudsed, need lubavad kasutada nihkega lähtehinnanguid. Neil on mitmeid häid omadusi. Lisaks kooskõlalisuse saavutamisele osakogumite hindamisel, on nad täpsemad (MSE mõttes) kui esialgsed hinnangud. Neid saab rakendada praktikas. Uusi tulemusi saadi ka osakogumite GREG ja SYN hinnangute korral. Nende hinnangute vastastikused MSE avaldised on uued, samuti SYN hinnangu üldkujulised nihke ja MSE avaldised. Töös esitatud süstemaatilist käsitlust GREG ja SYN hinnangute võrdumise tingimuste kohta ja seonduvate omaduste kohta ei ole ka mujal kirjanduses leida.

Curriculum Vitae

Natalja Lepik

Citizenship: Estonian Republic
Born: February, 17, 1977, Tartu Estonia
Marital Status: married, two daughters
Address: Aliise tee 3, Veibri küla, Luunja vald, Tartumaa 62220, Estonia
Contacts: e-mail: natalja.lepik@ut.ee

Education

1994 - 1999 University of Tartu, Faculty of Mathematics and Computer Science, Institute of Mathematical Statistics; bachelor's degree
2000-2003 University of Tartu, Faculty of Mathematics and Computer Science, Institute of Mathematical Statistics; master degree
2004 - 2011 University of Tartu, Faculty of Mathematics and Computer Science, Institute of Mathematical Statistics; PhD study

Professional employment

1999 - 2005 Tartu Vocational Educational Center (Tartu Kutsehariduskeskus), teacher of statistics and IT
2005 - 2008 University of Tartu, Faculty of Mathematics and Computer Science, Institute of Mathematical Statistics; Assistant
Since 2008 University of Tartu, Faculty of Mathematics and Computer Science, Institute of Mathematical Statistics; Lecturer (1.00)

Curriculum Vitae

Natalja Lepik

Kodakondsus: Eesti Vabariik
Sünniaeg ja -koht: 17. veebruar, 1977, Tartu, Eesti
Perekonnaseis: abielus, kaks tütart
Aadress: Aliise tee 3, Veibri küla, Luunja vald, Tartumaa 62220, Eesti
Kontaktandmed: e-mail: natalja.lepik@ut.ee

Hariduskäik

1994 - 1999 Tartu Ülikool, Matemaatika-informaatikateaduskond, bakalaureusekraad
2000-2003 Tartu Ülikool, Matemaatika-informaatikateaduskond, matemaatilise statistika instituut, magistrikraad
2004 - 2011 Tartu Ülikool, Matemaatika-informaatikateaduskond, doktoriõpingud matemaatilise statistika erialal

Erialane teenistuskäik

1999 - 2005 Tartu Kutsehariduskeskus, statistika ja informaatika õpetaja
2005 - 2008 Tartu Ülikool, Matemaatika-informaatikateaduskond, matemaatilise statistika instituut, assistent
Alates 2008 Tartu Ülikool, Matemaatika-informaatikateaduskond, matemaatilise statistika instituut, lektor (1.00)

DISSERTATIONES MATHEMATICAE UNIVERSITATIS TARTUENSIS

1. **Mati Heinloo.** The design of nonhomogeneous spherical vessels, cylindrical tubes and circular discs. Tartu, 1991, 23 p.
2. **Boris Komrakov.** Primitive actions and the Sophus Lie problem. Tartu, 1991, 14 p.
3. **Jaak Heinloo.** Phenomenological (continuum) theory of turbulence. Tartu, 1992, 47 p.
4. **Ants Tauts.** Infinite formulae in intuitionistic logic of higher order. Tartu, 1992, 15 p.
5. **Tarmo Soomere.** Kinetic theory of Rossby waves. Tartu, 1992, 32 p.
6. **Jüri Majak.** Optimization of plastic axisymmetric plates and shells in the case of Von Mises yield condition. Tartu, 1992, 32 p.
7. **Ants Aasma.** Matrix transformations of summability and absolute summability fields of matrix methods. Tartu, 1993, 32 p.
8. **Helle Hein.** Optimization of plastic axisymmetric plates and shells with piece-wise constant thickness. Tartu, 1993, 28 p.
9. **Toomas Kiho.** Study of optimality of iterated Lavrentiev method and its generalizations. Tartu, 1994, 23 p.
10. **Arne Kokk.** Joint spectral theory and extension of non-trivial multiplicative linear functionals. Tartu, 1995, 165 p.
11. **Toomas Lepikult.** Automated calculation of dynamically loaded rigid-plastic structures. Tartu, 1995, 93 p, (in Russian).
12. **Sander Hannus.** Parametrical optimization of the plastic cylindrical shells by taking into account geometrical and physical nonlinearities. Tartu, 1995, 74 p, (in Russian).
13. **Sergei Tupailo.** Hilbert's epsilon-symbol in predicative subsystems of analysis. Tartu, 1996, 134 p.
14. **Enno Saks.** Analysis and optimization of elastic-plastic shafts in torsion. Tartu, 1996, 96 p.
15. **Valdis Laan.** Pullbacks and flatness properties of acts. Tartu, 1999, 90 p.
16. **Märt Põldvere.** Subspaces of Banach spaces having Phelps' uniqueness property. Tartu, 1999, 74 p.
17. **Jelena Ausekle.** Compactness of operators in Lorentz and Orlicz sequence spaces. Tartu, 1999, 72 p.
18. **Krista Fischer.** Structural mean models for analyzing the effect of compliance in clinical trials. Tartu, 1999, 124 p.

19. **Helger Lipmaa.** Secure and efficient time-stamping systems. Tartu, 1999, 56 p.
20. **Jüri Lember.** Consistency of empirical k-centres. Tartu, 1999, 148 p.
21. **Ella Puman.** Optimization of plastic conical shells. Tartu, 2000, 102 p.
22. **Kaili Müürisep.** Eesti keele arvutigrammatika: süntaks. Tartu, 2000, 107 lk.
23. **Varmo Vene.** Categorical programming with inductive and coinductive types. Tartu, 2000, 116 p.
24. **Olga Sokratova.** Ω-rings, their flat and projective acts with some applications. Tartu, 2000, 120 p.
25. **Maria Zeltser.** Investigation of double sequence spaces by soft and hard analitical methods. Tartu, 2001, 154 p.
26. **Ernst Tungel.** Optimization of plastic spherical shells. Tartu, 2001, 90 p.
27. **Tiina Puolakainen.** Eesti keele arvutigrammatika: morfoloogiline ühestamine. Tartu, 2001, 138 p.
28. **Rainis Haller.** *M(r,s)*-inequalities. Tartu, 2002, 78 p.
29. **Jan Villemson.** Size-efficient interval time stamps. Tartu, 2002, 82 p.
30. **Eno Tõnisson.** Solving of expession manipulation exercises in computer algebra systems. Tartu, 2002, 92 p.
31. **Mart Abel.** Structure of Gelfand-Mazur algebras. Tartu, 2003. 94 p.
32. **Vladimir Kuchmei.** Affine completeness of some ockham algebras. Tartu, 2003. 100 p.
33. **Olga Dunajeva.** Asymptotic matrix methods in statistical inference problems. Tartu 2003. 78 p.
34. **Mare Tarang.** Stability of the spline collocation method for volterra integro-differential equations. Tartu 2004. 90 p.
35. **Tatjana Nahtman.** Permutation invariance and reparameterizations in linear models. Tartu 2004. 91 p.
36. **Märt Möls.** Linear mixed models with equivalent predictors. Tartu 2004. 70 p.
37. **Kristiina Hakk.** Approximation methods for weakly singular integral equations with discontinuous coefficients. Tartu 2004, 137 p.
38. **Meelis Käärik.** Fitting sets to probability distributions. Tartu 2005, 90 p.
39. **Inga Parts.** Piecewise polynomial collocation methods for solving weakly singular integro-differential equations. Tartu 2005, 140 p.
40. **Natalia Saealle.** Convergence and summability with speed of functional series. Tartu 2005, 91 p.
41. **Tanel Kaart.** The reliability of linear mixed models in genetic studies. Tartu 2006, 124 p.

42. **Kadre Torn.** Shear and bending response of inelastic structures to dynamic load. Tartu 2006, 142 p.
43. **Kristel Mikkor.** Uniform factorisation for compact subsets of Banach spaces of operators. Tartu 2006, 72 p.
44. **Darja Saveljeva.** Quadratic and cubic spline collocation for Volterra integral equations. Tartu 2006, 117 p.
45. **Kristo Heero.** Path planning and learning strategies for mobile robots in dynamic partially unknown environments. Tartu 2006, 123 p.
46. **Annely Mürk.** Optimization of inelastic plates with cracks. Tartu 2006. 137 p.
47. **Annemai Raidjõe.** Sequence spaces defined by modulus functions and superposition operators. Tartu 2006, 97 p.
48. **Olga Panova.** Real Gelfand-Mazur algebras. Tartu 2006, 82 p.
49. **Härmel Nestra.** Iteratively defined transfinite trace semantics and program slicing with respect to them. Tartu 2006, 116 p.
50. **Margus Pihlak.** Approximation of multivariate distribution functions. Tartu 2007, 82 p.
51. **Ene Käärik.** Handling dropouts in repeated measurements using copulas. Tartu 2007, 99 p.
52. **Artur Sepp.** Affine models in mathematical finance: an analytical approach. Tartu 2007, 147 p.
53. **Marina Issakova.** Solving of linear equations, linear inequalities and systems of linear equations in interactive learning environment. Tartu 2007, 170 p.
54. **Kaja Sõstra.** Restriction estimator for domains. Tartu 2007, 104 p.
55. **Kaarel Kaljurand.** Attempto controlled English as a Semantic Web language. Tartu 2007, 162 p.
56. **Mart Anton.** Mechanical modeling of IPMC actuators at large deformations. Tartu 2008, 123 p.
57. **Evely Leetma.** Solution of smoothing problems with obstacles. Tartu 2009, 81 p.
58. **Ants Kaasik.** Estimating ruin probabilities in the Cramér-Lundberg model with heavy-tailed claims. Tartu 2009, 139 p.
59. **Reimo Palm.** Numerical Comparison of Regularization Algorithms for Solving Ill-Posed Problems. Tartu 2010, 105 p.
60. **Indrek Zolk.** The commuting bounded approximation property of Banach spaces. Tartu 2010, 107 p.
61. **Jüri Reimand.** Functional analysis of gene lists, networks and regulatory systems. Tartu 2010, 153 p.
62. **Ahti Peder.** Superpositional Graphs and Finding the Description of Structure by Counting Method. Tartu 2010, 87 p.

63. **Marek Kolk.** Piecewise Polynomial Collocation for Volterra Integral Equations with Singularities. Tartu 2010, 134 p.
64. **Vesal Vojdani.** Static Data Race Analysis of Heap-Manipulating C Programs. Tartu 2010, 137 p.
65. **Larissa Roots.** Free vibrations of stepped cylindrical shells containing cracks. Tartu 2010, 94 p.
66. **Mark Fišel.** Optimizing Statistical Machine Translation via Input Modification. Tartu 2011, 104 p.
67. **Margus Niitsoo**. Black-box Oracle Separation Techniques with Applications in Time-stamping. Tartu 2011, 174 p.
68. **Olga Liivapuu.** Graded q-differential algebras and algebraic models in noncommutative geometry. Tartu 2011, 112 p.
69. **Aleksei Lissitsin.** Convex approximation properties of Banach spaces. Tartu 2011, 107 p.
70. **Lauri Tart.** Morita equivalence of partially ordered semigroups. Tartu 2011, 101 p.
71. **Siim Karus.** Maintainability of XML Transformations. Tartu 2011, 142 p.
72. **Margus Treumuth.** A Framework for Asynchronous Dialogue Systems: Concepts, Issues and Design Aspects. Tartu 2011, 95 p.
73. **Dmitri Lepp.** Solving simplification problems in the domain of exponents, monomials and polynomials in interactive learning environment T-algebra. Tartu 2011, 202 p.
74. **Meelis Kull.** Statistical enrichment analysis in algorithms for studying gene regulation. Tartu 2011, 151 p.
75. **Nadežda Bazunova.** Differential calculus $d^3 = 0$ on binary and ternary associative algebras. Tartu 2011, 99 p.