University of Southampton

Faculty of Social, Human & Mathematical Sciences

Department of Social Statistics and Demography

Empirical likelihood approach for estimation from multiple sources

by

Ewa Joanna Kabzińska

Thesis for the degree of Doctor of Philosophy

December 2017

University of Southampton

Abstract

Faculty of Social, Human & Mathematical Sciences
Social Statistics and Demography

Doctor of Philosophy

Empirical likelihood approach for estimation from multiple sources

by Ewa Joanna Kabzińska

Empirical likelihood is a non-parametric, likelihood-based inference approach. In the design-based empirical likelihood approach introduced by Berger and De La Riva Torres (2016), the parameter of interest is expressed as a solution to an estimating equation. The maximum empirical likelihood point estimator is obtained by maximising the empirical likelihood function under a system of constraints. A single vector of weights, which can be used to estimate various parameters, is created. Design-based empirical likelihood confidence intervals are based on the χ^2 approximation of the empirical likelihood ratio function. The confidence intervals are range-preserving and asymmetric, with the shape driven by the distribution of the data.

In this thesis we focus on the extension and application of design-based empirical likelihood methods to various problems occurring in survey inference. First, a design-based empirical likelihood methodology for parameter estimation in two surveys context, in presence of alignment and benchmark constraints, is developed. Second, a design-based empirical likelihood multiplicity adjusted estimator for multiple frame surveys is proposed. Third, design-based empirical likelihood is applied to a practical problem of census coverage estimation.

The main contribution of this thesis is defining the empirical likelihood methodology for the studied problems and showing that the aligned and multiplicity adjusted empirical likelihood estimators are $\sqrt{n}$-design-consistent. We also discuss how the original proofs presented by Berger and De La Riva Torres (2016) can be adjusted to show that the empirical likelihood ratio statistic is pivotal and follows a χ^2 distribution under alignment constraints and when the multiplicity adjustments are used.

We evaluate the asymptotic performance of the empirical likelihood estimators in a series of simulations on real and artificial data. We also discuss the computational aspects of the calculations necessary to obtain empirical likelihood point estimates and confidence intervals and propose a practical way to obtain empirical likelihood confidence intervals in situations when they might be difficult to obtain using standard approaches.

Contents

Introduction **1**

1 Empirical likelihood methods for inference from survey data **7**

1.1 Empirical likelihood for a mean of independent and identically distributed observations 7

1.2 Complex sampling designs 9

1.3 Auxiliary information 13

1.4 Design-based empirical likelihood 15

2 Empirical likelihood approach for aligning estimates from multiple surveys **21**

2.1 Introduction 21

2.2 Sampling design and data collected 25

2.3 Some existing approaches 30

2.4 Empirical likelihood approach proposed 37

2.5 Maximum empirical likelihood point estimator 41

2.5.1 Estimation of the mean of the function of the common variable 42

2.6 Asymptotic properties of the maximum empirical likelihood point estimator 44

2.6.1 Regularity conditions 45

2.6.2 Asymptotic equivalence of the maximum empirical likelihood point estimator to a generalized regression type estimator 47

2.6.3 Asymptotic design-consistency of the maximum empirical likelihood point estimator 50
2.7 Effect of a difference in sample sizes on the maximum empirical likelihood point estimator 51
2.8 The empirical likelihood ratio statistic 55
2.9 Tests and empirical likelihood confidence regions 57
2.10 Stratification 58
2.11 Without replacement sampling and large sampling fractions 59
2.12 Simulation studies 61
2.12.1 Point estimation 61
2.12.2 Samples of different sizes 67
2.12.3 Confidence intervals: British Labour Force Survey 73
2.12.4 Confidence intervals: quantiles 76
2.13 Conclusions 78

3 Empirical likelihood multiplicity adjusted estimator for multiple frame surveys 81

3.1 Introduction 81
3.2 Some existing dual frame estimators 83
3.2.1 Separate frame estimators 85
3.2.2 Combined frame estimators 87
3.3 Multiplicity estimation 90
3.4 Empirical likelihood approach 94
3.4.1 Maximum empirical likelihood point estimator 97
3.4.2 Relationship to a generalized regression type estimator . . . 99
3.4.3 Asymptotic design-consistency of the empirical likelihood multiplicity adjusted estimator 101
3.4.4 Empirical Likelihood confidence intervals 102
3.5 Extensions 103
3.5.1 Stratification 104
3.5.2 Domain-based constraints 105

3.5.3 Alignment constraints on the overlapping domain 106
3.6 Relationship to the aligned empirical likelihood estimator 108
3.7 Simulation study 111
3.7.1 Estimation of totals 112
3.7.2 Estimation of quantiles of distribution 118
3.8 Conclusions 121

4 Using empirical likelihood to obtain range-respecting confidence intervals for census coverage 123
4.1 Introduction 123
4.2 Sampling design of the census coverage survey and the current estimation procedure 125
4.2.1 Sample selection 125
4.2.2 Dual system population size estimation 126
4.2.3 Population size and census coverage estimation 128
4.3 Applying empirical likelihood to census coverage 129
4.4 Numerical illustration 134
4.5 Simulation study 137
4.6 Conclusions 145

5 Numerical aspects of empirical likelihood 147
5.1 Obtaining the vectors of adjusted weights $\widehat{m}_i(\boldsymbol{\varphi}_U)$ and $\widehat{m}_i^\star(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)$ . . 148
5.2 Obtaining the point estimate $\widehat{\boldsymbol{\theta}}$ 151
5.3 Obtaining the lower and upper bounds of confidence intervals . . . 152
5.4 Simulation study: execution times 156
5.5 Conclusions 160

Discussion 163

A Proofs of the results 167
A.1 Proofs of the results of Chapter 2 167
A.2 Proofs of the results of Chapter 3 176

List of Tables

2.1 Aligned empirical likelihood estimator: Relative absolute root mean square errors (%) for estimators of totals of the non-common variables in three populations of interest, with both samples of equal sizes. 64

2.2 Aligned empirical likelihood estimator: Relative absolute root mean square errors (%) for estimators of totals of the common variable in two populations of interest, with both samples of equal sizes. . . 66

2.3 Aligned empirical likelihood estimator: Relative absolute root mean square errors (%) for estimators of totals of the non-common variables in the artificial population *AMELIA*, with both samples of equal sizes. 67

2.4 Aligned empirical likelihood estimator: Relative absolute root mean square errors (%) for estimators of totals of the non-common variables with alignment of samples of different sizes. 69

2.5 Aligned empirical likelihood estimator: Relative absolute root mean square errors (%) for estimators of totals of the common variable with alignment of samples of different sizes. 72

2.6 Aligned empirical likelihood estimator: Coverages and tail error rates of confidence intervals for total number of hours worked per week per domains. No outliers introduced. 75

2.7 Aligned empirical likelihood estimator: Coverages and tail error rates of confidence intervals for total number of hours worked per week per domains. 5 % outliers introduced into the variable of interest. 75

2.8 Aligned empirical likelihood estimator: Coverages and tail error rates of confidence intervals for 80% and 90% quantiles 77

3.1 Multiplicity adjusted empirical likelihood estimator: 100 x percent relative mean squared error of the Empirical Likelihood Multiplicity adjusted estimator and the Generalized Multiplicity-adjusted Regression estimator based on *Population 1* 113

3.2 Multiplicity adjusted empirical likelihood estimator: 100 × percent relative mean squared error of the Empirical Likelihood Multiplicity adjusted estimator and the Generalized Multiplicity-adjusted Regression estimator based on *Population 2* 115

3.3 Multiplicity adjusted empirical likelihood estimator: Coverage of confidence intervals of the Empirical Likelihood Multiplicity adjusted estimator and the Generalized Multiplicity-adjusted Regression estimator, based on *Population 2* 116

3.4 Multiplicity adjusted empirical likelihood estimator: 100 x percent relative mean squared error of the Empirical Likelihood Multiplicity adjusted estimator and the Generalized Multiplicity-adjusted Regression estimator, y_i and π_i generated independently 117

3.5 Multiplicity adjusted empirical likelihood estimator:100 x percent relative mean squared error of the Empirical Likelihood Multiplicity adjusted estimator and the Generalized Multiplicity-adjusted Regression estimator, $cor(y_i, \pi_i) \approx 0.6$ 117

3.6 Multiplicity adjusted empirical likelihood estimator: Coverage of confidence intervals for the estimators of totals in *Population 3* . . 118

3.7 Multiplicity adjusted empirical likelihood estimator: Relative absolute root mean square errors (%), 100 × percent relative mean squared errors, left tail error rates, right tail error rates and coverages of confidence intervals of the Empirical Likelihood Multiplicity-Adjusted estimator based on *EUSILCP* data. Estimation of quantiles of distribution and the mean of equalised household income . . 120

5.1 Distribution of the user execution times in seconds for calculation of point estimates and confidence intervals using various methods. . 159

List of Figures

2.1 Aligned empirical likelihood estimator: Sample data and parameters of the samples $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ 26

3.1 Empirical likelihood multiplicity adjusted estimator: Illustration of the sampling frames within population U and samples selected . . . 84

4.1 Geographical entities in England and Wales used in the design of the Census Coverage Survey 126

4.2 Empirical likelihood 95% confidence intervals for the census coverage in different Estimation Areas and age-sex groups. 135

4.3 Empirical likelihood and Symmetric (jackknife) 95% confidence intervals for the census coverage in selected age-sex groups 136

4.4 Examples of the log-likelihood ratio plotted as a function of the point estimate for census coverage in some selected age-sex groups. 137

4.5 Coverage of empirical likelihood and symmetric confidence intervals in various age-sex groups, in population *synthIL06KENS* 140

4.6 Coverage of empirical likelihood and symmetric confidence intervals in various age-sex groups, in population *synthIL09SOUT* 141

4.7 Coverage of empirical likelihood and symmetric confidence intervals in various age-sex groups, in population *synthSW03CORN* 141

4.8 Coverage of empirical likelihood and symmetric confidence intervals in various age-sex groups, in population *synthNW06MERS* 142

4.9 Average length of confidence intervals in various age-sex groups, in population *synthIL06KENS* 142

4.10 Average length of confidence intervals in various age-sex groups, in population *synthIL09SOUT* 143
4.11 Average length of confidence intervals in various age-sex groups, in population *synthSW03CORN* 143
4.12 Average length of confidence intervals in various age-sex groups, in population *synthNW06MERS* 144

Acknowledgements

I would like to thank my supervisors, Dr Yves Berger and Prof. Li Chun Zhang, for their advice throughout this research. I am extremely grateful to Dr Yves Berger for introducing me to empirical likelihood and for his incredible support during my studies. I would also like to express special gratitude to Mr Paul Smith for suggesting the topic of chapter 4 and for his helpful comments and explanations during the research, as well as for numerous helpful comments on the first draft of a paper based on chapter 2. I am also grateful to Mr Owen Abbott and Mr Viktor Racinskij from the Office for National Statistics for several discussions related to the research presented in chapter 4 and for providing me with the opportunity to speak to the Census Advisory Board.

I am very grateful to Prof. Fulvia Mecatti and Dr Giovanna Rannali for introducing me to multiple frames estimation. I am also grateful to Dr Melike Oguz-Alper for her guidance and help during my studies, particularly for explaining the ins and outs of empirical likelihood estimation.

I would also like to thank the Department of Social Statistics and Demography for creating an inspiring and supportive environment where I could carry out my research and the Economic and Social Research Council for supporting my studies.

Finally, I would like to thank my family and friends for their support and help during my studies, and for bearing with me for the last four years.

Notation

We use the following notation:

$\log(\cdot)$	Natural logarithm of $(\cdot)$
$\|\boldsymbol{A}\|$	Frobenius (Euclidean) norm defined by $\|\boldsymbol{A}\| := \text{trace}(\boldsymbol{A}^\top \boldsymbol{A})^{1/2}$
$\boldsymbol{\mathcal{O}}_{\mathcal{P}}(\cdot)$	A matrix or a vector such that $\|\boldsymbol{\mathcal{O}}_{\mathcal{P}}(a)\| = O_{\mathcal{P}}(a)$
$\boldsymbol{o}_{\mathcal{P}}(\cdot)$	A matrix or a vector such that $\|\boldsymbol{o}_{\mathcal{P}}(a)\| = o_{\mathcal{P}}(a)$
$a = o_{\mathcal{P}}(b)$	a converges in probability to b
$a = O_{\mathcal{P}}(b)$	a is stochastically bounded by b
$a = O(b)$	a is bounded by b
$a = b$	a is equal to b
$a := b$	a is defined as b
$a \asymp_p b$	a converges in probability to b
$a \xrightarrow{d} b$	a converges in distribution to b
$\boldsymbol{x}$	A column vector
$\boldsymbol{x}^\top$	A row vector
x	A scalar
$\boldsymbol{\theta}_U$	A finite population parameter
$\boldsymbol{\theta}$	A candidate value for the population parameter
$\widehat{\boldsymbol{\theta}}$	An estimator of a finite population parameter $\boldsymbol{\theta}_U$
$\mathbf{0}_q$	A vector of zeros of dimension q

We take the convention that each chapter defines its specific notation, e.g. definition of a vector $\boldsymbol{C}$ given in chapter 1 applies each time this variable is used in chapter 1 and chapter 1 only.

Introduction

This document consists of five chapters and an Appendix. The first chapter gives a brief introduction to empirical likelihood and shows how the design-based empirical likelihood approach, used throughout this research, relates to other empirical likelihood methods.

The following three chapters focus on extension and application of design-based empirical likelihood to various problems occurring in survey inference. First, a methodology for parameter estimation in two surveys context, in presence of alignment and benchmark constraints, is developed. Second, an empirical likelihood multiplicity adjusted estimator for multiple frame surveys is proposed. Third, empirical likelihood is applied to a practical problem of census coverage estimation. The last chapter contains some details on the numerical operations necessary in empirical likelihood estimation. Each chapter includes conclusions specific to the studied problem. We finish with general conclusions from the work described in the previous chapters and a discussion of possible future direction of research. Proofs of the theoretical results are presented in the Appendix.

Empirical likelihood is a non-parametric, likelihood-based inference approach. The method was proposed by Owen (1988), for independent and identically distributed (*iid*) observations. This made it possible to use empirical likelihood in survey estimation under simple random sampling, when the sample is selected with replacement or when the sampling fraction is negligible (e.g. Chen and Qin, 1993). Qin and Lawless (1994) showed how confidence intervals can be obtained based on the χ^2 approximation of the log-likelihood ratio function in the *iid* case and how auxil-

iary information can be incorporated into empirical likelihood estimation. Subsequently an extension to stratified simple random sampling was proposed by Zhong and Rao (1996, 2000). Chen and Sitter (1999) proposed a census pseudoempirical likelihood approach, which was later developed into a pseudoempirical likelihood approach by Wu and Rao (2006). In the pseudoempirical likelihood methodology, unit sampling probabilities are incorporated into the empirical likelihood function. This allows empirical likelihood to be used under unequal probability sampling, but usually requires estimation of the design effect to obtain confidence intervals (Wu and Rao, 2006). Chen and Kim (2014) proposed population empirical likelihood, where the likelihood function is defined at the population level and the sampling probabilities are included in the estimating equation for the parameter of interest.

Berger and De La Riva Torres (2016) developed design-based empirical likelihood, where the likelihood function is defined at the sample level and sampling probabilities are included in an additional design constraint. This approach allows confidence intervals to be based directly on the χ^2 approximation of the empirical likelihood ratio function, without a design effect correction as in pseudo-empirical likelihood methods. The confidence intervals can be calculated without variance estimates. This is a very desirable feature, as variance estimators may be heavily biased when the variables of interest are skewed. The confidence intervals are asymmetric, with the shape driven by the distribution of the data. They are also range-preserving (Owen, 2001). The method can be used in complex, i.e., stratified and clustered, sampling designs.

In the design-based empirical likelihood approach, the parameter of interest is expressed as a solution to a population level estimating equation. The maximum empirical likelihood point estimator of the parameter of interest is obtained by maximising the empirical likelihood function under a system of constraints including the estimating function for the parameter of interest, optional benchmark constraints constructed around known population level parameters and design

constraints including information about the sampling design and unit selection probabilities. A single vector of weights, which can be used to estimate various parameters, is created. These weights are always positive.

Berger and De La Riva Torres's (2016) design-based empirical likelihood approach is used throughout the research presented in this document. In chapter 2 a design-based empirical likelihood methodology for parameter estimation in two surveys context, in presence of alignment and benchmark constraints, is developed. Alignment constraints, which require that each of the considered independent surveys gives the same point estimates for the common variables, are sometimes used in official statistics in order to ensure numerical consistency of estimates obtained from various sources. Alignment may also increase precision of other estimates. The standard methods either focus on means or totals and rely on composite regression estimators and variance estimates, or assume negligible sampling fractions. The proposed empirical likelihood approach ensures alignment and is not limited to means, as it can be used for a general class of complex parameters defined by estimating equations. It also allows to use various functions of the common variable in the alignment constraint. The proposed approach is well suited when the variables of interest are skewed. It can accommodate large sampling fractions, stratification and population level (auxiliary) information, and can be applied to estimation in domains. The confidence intervals are asymmetric and driven by the distribution of the data. They can be calculated without the need for variance estimates, joint selection probabilities or re-sampling.

The main contribution of chapter 2 is in defining the empirical likelihood framework for alignment of estimates, showing that the maximum empirical likelihood estimator is $\sqrt{n}$-design-consistent and deriving the empirical likelihood ratio test statistic, which can be used to test hypotheses and construct consistent confidence regions or intervals. We evaluate the proposed approach in a series of simulations on real and artificial datasets and conclude that the proposed aligned empirical likelihood estimator has good asymptotic properties across the designs tested. In

some cases, e.g. when there is a large difference in sample sizes and the distribution of the data is skewed, empirical likelihood estimates of totals may perform better than other available methods. The main purpose of the proposed approach, however, is not the efficiency gain, as this might vary depending on circumstances, but providing a practical method for estimation of more complex parameters than totals or means and for calculation of confidence intervals when variance estimation is difficult.

In chapter 3 an empirical likelihood methodology for parameter estimation from multiple frame surveys, based on the multiplicity approach, is proposed. Multiple frame surveys are commonly used for a variety of reasons, such as correcting for frame undercoverage, increasing precision of estimation of population parameters for groups of interest, targeting rare populations or reducing survey costs. Several approximately design unbiased estimators have been proposed for inference from multiple frame surveys. Singh and Mecatti (2011) and Mecatti and Singh (2014) generalized most of the existing estimators as a class of Generalized Multiplicity-Adjusted Horvitz-Thompson Estimators. We adopt the idea of the Multiplicity-Adjusted Estimation and develop an Empirical Likelihood based estimator. The proposed estimator is flexible in that it allows researchers to use the multiplicity adjustment of their choice, setting some standard regularity conditions on the multiplicity adjustment and the sampling design. It can handle auxiliary information and can be applied to a variety of parameters of interest expressed as solutions of estimating equations. As in the case of the aligned empirical likelihood estimator, Wilks (1938) type confidence intervals can be calculated without the intermediate step of variance estimation.

The main contribution of chapter 3 is an extension of the theoretical results of chapter 2 to the multiple frame case. We define a design-based empirical likelihood multiplicity adjusted estimator and show that under some regularity conditions this estimator is $\sqrt{n}$-design-consistent. We also show that the multiplicity adjusted empirical likelihood ratio function is pivotal and can be used to construct

confidence intervals. Through a series of simulations, we demonstrate that the proposed estimator performs well even in difficult conditions, e.g. with skewed data and when the size of the overlap between sampling frames is unknown. In these cases the empirical likelihood confidence intervals often have better coverage than symmetric confidence intervals, and the empirical likelihood point estimator may be more precise than regression estimators with the same multiplicity adjustment.

Chapter 4 shows how design-based empirical likelihood can be applied to estimation of census coverage from a census coverage survey. Currently census coverage is estimated using normality-based techniques and symmetric confidence intervals are reported. However, in areas with very high estimated coverage, the upper bound of the symmetric confidence intervals for the census coverage sometimes exceed 1. We show that the empirical likelihood confidence intervals do not have this problem as they always remain within the range of the parameter of interest and that they have comparable, acceptable coverage for moderate and large samples. The main contribution of this chapter is in the definition of the relevant estimating equations and constraints for the problem of census coverage estimation. We also perform a series of simulations showing that the empirical likelihood confidence intervals are within the desired range and that they have good asymptotic properties provided that the sample size is sufficient.

Finally, in chapter 5, we discuss the practical aspects of empirical likelihood estimation. In particular, we focus on the numerical methods involved and consider various ways of obtaining empirical likelihood adjusted weights, point estimates and confidence intervals. We propose some adjustments to the commonly used algorithms. The problems discussed in chapter 5 apply to empirical likelihood estimation in general, but they are particularly relevant when multiple samples and numerous constraints are used, which is often the case in the applications discussed in this piece of work.

Results presented in chapters 2 and 4 have been submitted for publication, in joint

authorship respectively with Dr. Yves Berger and with Mr Paul Smith and Dr Yves Berger. The scientific paper produced based on chapter 2 was considerably changed and enriched by Dr Yves Berger. It also includes some results, such as the derivation of a consistent estimator of the variance-covariance matrix of the regression estimator, used in the proof of the asymptotic distribution of the empirical log-likelihood ratio function, and extension of the results to sampling with large sampling fractions, which are based on previous research done by Dr Yves Berger. These results are cited from the paper. Also, the proofs of the theoretical results, except from the proof of the asymptotic design-consistency of the point estimator, are an adaptation of the results presented by Berger and De La Riva Torres (2016) for a single sample case. This is acknowledged in the Appendix. The scientific paper based on chapter 4 was prepared jointly with Mr Paul Smith and Dr Yves Berger. Some of the details related to the design of the census coverage survey, which were contributed to the paper by Mr Paul Smith, are cited in this document. These are referenced in the text. The review of empirical likelihood methods presented in chapter 1 is based on reviews of Rao and Wu (2009*a*), Rao (2006), Berger and De La Riva Torres (2016) and Berger (2018). The review of aligned estimators is based on the reviews presented by Merkouris (2004, 2010*a*). The review of multiple frame estimators is based on the reviews presented by Arcos et al. (2015), Ranalli et al. (2016), Singh and Mecatti (2011), Singh and Mecatti (2014) and Lohr (2007). During the course of research I received guidance from my supervisors, Dr Yves Berger and Prof. Li Chun Zhang, as well as from Mr Paul Smith, who offered advice on chapter 4. I also consulted Mr Owen Abbott and Mr Viktor Racinskij from the Office for National Statistics about the specifics related to the design of the census coverage survey and the current census coverage estimation practice.

Chapter 1

Empirical likelihood methods for inference from survey data

This chapter provides a brief overview of empirical likelihood methods for parameter estimation. We start with a summary of the origins of empirical likelihood and proceed to discuss the two crucial areas of development which made empirical likelihood applicable to social and business surveys: incorporation of auxiliary information and unequal probability sampling. We also describe the design-based empirical likelihood approach, which we rely on in subsequent chapters.

1.1 Empirical likelihood for a mean of independent and identically distributed observations

Empirical likelihood is a non-parametric, likelihood-based inference approach. The method derives from the scale-load approach introduced by Hartley and Rao (1968) for survey sampling. It was popularised and developed by Owen (1988), as a unified empirical likelihood methodology for independent and identically distributed (*iid*) observations.

Consider that a sample $\boldsymbol{S}$ of independent and identically distributed values

$y_1, y_2, ..., y_n$, is drawn through simple random sampling from a finite population U of size N. Let $p_i = Pr(y = y_i)$ be the probability mass associated with unit i. The sample level empirical log-likelihood function takes the following form (Owen, 1988):

$$\ell(\boldsymbol{p}) = \sum_{i \in \boldsymbol{S}} \log(p_i), \tag{1.1}$$

where $\boldsymbol{p} = (p_1, p_2, ..., p_n)$ and n is the size of the sample $\boldsymbol{S}$. The maximum empirical likelihood estimator $\hat{p}_i$ of p_i is defined as the value which maximises (1.1) under $p_i > 0$ and the normalising constraint

$$\sum_{i \in \boldsymbol{S}} p_i = 1\cdot \tag{1.2}$$

This gives $\hat{p}_i = n^{-1}$. The maximum empirical likelihood estimator of a population mean $\theta_U = N^{-1} \sum_{i=1}^{N} y_i$ is (Rao, 2006):

$$\hat{\theta} = \sum_{i \in \boldsymbol{S}} \hat{p}_i y_i\cdot \tag{1.3}$$

Empirical likelihood confidence intervals for the mean θ_U are obtained by maximising the empirical log-likelihood ratio function (Owen, 1988)

$$\hat{r}(\theta) = -2 \sum_{i \in \boldsymbol{S}} \log\{n\hat{p}_i^*(\theta)\}, \tag{1.4}$$

where $\hat{p}_i^*(\theta)$ are the values which maximise (1.1) under $p_i > 0$, the normalising constraint (1.2) and the constraint

$$\sum_{i \in \boldsymbol{S}} p_i y_i = \theta\cdot \tag{1.5}$$

Under simple random sampling, when the sample is selected with replacement or when the sampling fraction is negligible, when $\theta = \theta_U$, statistic (1.4) follows a $\chi^2_{df=1}$ distribution asymptotically. This property can be used to construct empirical like-

lihood confidence intervals for the mean θ_0 by selecting values $\theta : r(\theta) \leq \chi^2_{df=1;\alpha}$, where $\chi^2_{df=1;\alpha}$ is the upper α quantile of the $\chi^2_{df=1}$ distribution. The empirical likelihood confidence intervals are asymmetric and range and transformation preserving (Rao, 2006). As the sample size n approaches infinity, the coverage error of the empirical likelihood confidence intervals approaches zero at the rate n^{-1}. This is the same rate that applies to most parametric confidence intervals (Owen, 2001).

1.2 Complex sampling designs

An extension of empirical likelihood to stratified simple random sampling was proposed by Zhong and Rao (1996). The empirical log-likelihood function for a sample consisting of H strata is defined as

$$\ell(\boldsymbol{p}) = \sum_{h=1}^{H} \sum_{i \in \boldsymbol{S}_h} \log(p_{h;i}), \tag{1.6}$$

where $p_{h;i}$ is the sampling probability mass associated with unit i in strata h. The maximum likelihood estimator of the population mean θ_U equals

$$\hat{\theta} = \sum_{h=1}^{H} \sum_{i \in \boldsymbol{S}_h} \hat{p}_{h;i} y_{h;i}, \tag{1.7}$$

where $\hat{p}_{h;i}$ are the values which maximise (1.6) under the constraint $p_{h;i} > 0$ and the normalisation constraints defined for each stratum:

$$\sum_{i \in \boldsymbol{S}_h} p_i = 1, \quad h = 1, 2, ..., H\cdot \tag{1.8}$$

For empirical likelihood to be applicable to inference from commonly used survey sampling designs, it is crucial that it can handle unequal probability sampling. Consider a sample $\boldsymbol{S}$ of size n selected with unequal probabilities π_i from a finite

population U. The pseudoempirical likelihood approach of Wu and Rao (2006) defines the pseudoempirical log-likelihood function as

$$\ell(\boldsymbol{p}) = n \sum_{i \in \boldsymbol{S}} \tilde{d}_i \log(p_i), \tag{1.9}$$

where $\tilde{d}_i$ are the normalised weights $\tilde{d}_i = d_i(\sum_{i \in \boldsymbol{S}} d_i)^{-1}$ and $d_i = \pi_i^{-1}$. The pseudoempirical likelihood approach was inspired by an earlier formulation of the census pseudoempirical likelihood by Chen and Sitter (1999), which was based on a super-population model (Rao and Wu, 2009*a*).

The maximum likelihood estimates of p_i are found by maximising (1.9) under $p_i > 0$ and the normalising constraint

$$\sum_{i \in \boldsymbol{S}} p_i = 1\cdot \tag{1.10}$$

This gives $\hat{p}_i = \tilde{d}_i$. The maximum pseudoempirical likelihood estimate of the population mean $\theta_U = N^{-1} \sum_{i=1}^{N} y_i$ equals (Rao, 2006):

$$\hat{\theta} = \sum_{i \in \boldsymbol{S}} \tilde{d}_i y_i\cdot \tag{1.11}$$

When the sample $\boldsymbol{S}$ is stratified, the pseudoempirical log-likelihood function takes the following form:

$$\ell(\boldsymbol{p}) = n \sum_{h=1}^{H} W_h \sum_{i \in \boldsymbol{S}} \tilde{d}_{h;i} \log(p_i), \tag{1.12}$$

where $W_h = N_h N^{-1}$, N_h is the population size of strata h and $\tilde{d}_{h;i} = d_{h;i}(\sum_{i \in \boldsymbol{S}_h} d_{h;i})^{-1}$, i.e., the design weights $\tilde{d}_{h;i}$ are normalised at the stratum level.

The pseudoempirical log-likelihood ratio function is defined as

$$\hat{r}(\theta) = -2\left[\ell\{\hat{\boldsymbol{p}}^*(\theta)\} - \ell\{\hat{\boldsymbol{p}}\}\right], \tag{1.13}$$

where $\ell(\cdot)$ is defined by (1.9) (or (1.12) if the sample $\boldsymbol{S}$ is stratified), $\hat{\boldsymbol{p}}$ are the values which maximise (1.9) under $p_i > 0$ and the normalising constraint (1.10) and $\hat{\boldsymbol{p}}^*(\theta)$ are the values which maximise (1.9) under $p_i > 0$, (1.10) and the constraint

$$\sum_{i \in \boldsymbol{S}} p_i y_i = \theta \cdot \tag{1.14}$$

Under simple random sampling, (1.13) follows a $\chi^2_{df=1}$ distribution asymptotically when $\theta = \theta_U$. For other sampling designs, the pseudoempirical log-likelihood ratio has to be adjusted by the design effect defined as

$$DEFF(\hat{\theta}) = V(\hat{\theta})\{V_{SRS}(\hat{\theta})\}^{-1}, \tag{1.15}$$

where $V(\hat{\theta})$ is the variance of the estimator $\hat{\theta}$ under the considered sampling design and $V_{SRS}(\hat{\theta})$ is the variance under simple random sampling. The function

$$\hat{r}_{ADJ}(\theta) = \hat{r}(\theta) \left\{ DEFF(\hat{\theta}) \right\}^{-1} \tag{1.16}$$

follows a $\chi^2_{df=1}$ distribution asymptotically when $\theta = \theta_U$. The pseudoempirical likelihood confidence intervals are constructed based on (1.16) in an analogous way to the empirical likelihood confidence intervals.

In practice the design effect (1.15) has to be estimated based on the sample data. As long as it is estimated consistently, the asymptotic distribution of (1.16) holds (Wu and Rao, 2006).

Chen and Kim (2014) proposed population empirical likelihood, which defines the empirical log-likelihood function at the population level:

$$\ell(\boldsymbol{p}) = \sum_{i=1}^{N} \log(p_i) \cdot \tag{1.17}$$

Under Poisson sampling, the weights p_i are estimated as the values which maximise

(1.17) subject to $\sum_{i=1}^{N} p_i = 1$. This gives the estimated weights $\hat{p}_i = N^{-1}$. The parameter of interest is estimated by solving

$$\sum_{i=1}^{N} p_i \delta_i \pi_i^{-1} g_i(\theta) = 0, \tag{1.18}$$

where δ_i is the sample inclusion indicator and $g_i(\theta)$ is an estimating function of the parameter of interest.

The population empirical log-likelihood ratio function is defined as

$$\hat{r}(\theta) = -2 \left[\ell\{\hat{\boldsymbol{p}}^*(\theta)\} - \ell\{\hat{\boldsymbol{p}}\} \right], \tag{1.19}$$

where $\ell\{\hat{\boldsymbol{p}}^*(\theta)\}$ is the population log-likelihood function (1.17), $\hat{p}_i^*(\theta)$ are estimated as the values which maximise (1.17) subject to $\sum_{i=1}^{N} p_i = 1$ and the parameter constraint (1.18); and values $\hat{\boldsymbol{p}}$ in $\ell\{\hat{\boldsymbol{p}}\}$ are estimated without the parameter constraint (1.18). Under Poisson sampling with a negligible sampling fraction, (1.19) follows a χ^2 distribution (Chen and Kim, 2014). This property can be used to obtain confidence intervals for the parameter of interest.

Population empirical likelihood has also been extended to rejective Poisson sampling with the Hájek's (1964) constraint

$$\sum_{i=1}^{N} \delta_i = \sum_{i=1}^{N} \pi_i, \tag{1.20}$$

where π_i are the sampling probabilities in the initial design and δ_i are the sampling indicators. This requires adding the design constraint: (Chen and Kim, 2014)

$$\sum_{i=1}^{N} p_i (I_i \pi_i^{-1} - 1) = 0\cdot \tag{1.21}$$

Extension to Fuller's (2009) rejection condition has also been proposed (Chen and Kim, 2014).

1.3 Auxiliary information

Auxiliary information is often used in survey inference. Typically these are known population parameters (e.g. means or totals) of variables which are also measured in the sample. These known parameters are included in the so-called *calibration* or *benchmark* constraints, which require that the adjusted sample weights reproduce the known population values. This might improve the precision of the estimator of the target variable, depending on the correlation between the target variable and the auxiliary variable.

Suppose that the total of a variable $\boldsymbol{x}$, $\boldsymbol{X_U} = \sum_{i=1}^{N} \boldsymbol{x_i}$ is known. Consider Horvitz and Thompson's (1952) estimator of $\boldsymbol{X_U}$:

$$\hat{\boldsymbol{X}} = \sum_{i \in \boldsymbol{S}} d_i \boldsymbol{x}_i, \tag{1.22}$$

where $d_i = \pi_i^{-1}$ are the design weights. In a general case, there is no guarantee that the estimator (1.22) reproduces the known value $\boldsymbol{X_U}$. The weights w_i are said to possess the generalized calibration property (Deville and Särndal, 1992*a*) if

$$\hat{\boldsymbol{X}} = \boldsymbol{X_U}, \tag{1.23}$$

where

$$\hat{\boldsymbol{X}} = \sum_{i \in \boldsymbol{S}} w_i \boldsymbol{x}_i. \tag{1.24}$$

The calibration weights w_i are calculated in such a way that the distance between the w_i and the design weights d_i is minimised and (1.23) is satisfied. Various distance measures can be used. Using the Euclidean distance leads to the generalized regression (GREG) estimator (Särndal, 2007).

Benchmark constraints are commonly used in survey practice, especially in official statistics. If the target variable y is highly correlated with the auxiliary variables $\boldsymbol{x}$, benchmark constraints on $\boldsymbol{X}$ might improve the precision of the estimator of a population parameter of y. Sometimes the benchmark constraints are also used for practical reasons, e.g. in order to obtain numerical consistency with values published from a census.

Chen and Qin (1993) showed how auxiliary information on known population means can be incorporated into empirical likelihood estimators. Suppose that a population mean $\bar{x}$ is known. Imposing the additional constraint

$$\sum_{i \in \boldsymbol{S}} p_i (x_i - \bar{x}) = 0 \tag{1.25}$$

on the adjusted weights p_i ensures that the calibration property (1.23), with $w_i = p_i$, holds. For the estimates $\hat{p}_i$ to exist, $\bar{x}$ has to be an inner point of the convex hull formed by the values $\{x_i\}$, where $i \in \boldsymbol{S}$ (Rao, 2006).

The known population parameters may be included in the pseudoempirical likelihood estimators through an additional constraint. The constraint for a non-stratified sample takes the form (Rao and Wu, 2009*a*):

$$\sum_{i \in \boldsymbol{S}} p_i x_i = \bar{x}\cdot \tag{1.26}$$

For stratified samples, when the population level parameter is known, the constraint is defined as (Rao and Wu, 2009*a*)

$$\sum_{h=1}^{H} \sum_{i \in \boldsymbol{S}_h} p_{h;i} x_{h;i} = \bar{x}\cdot \tag{1.27}$$

Note that when a calibration constraint is used, this has to be considered in the calculation of the design effect used to construct pseudoempirical likelihood

confidence intervals and hence equation (1.15) becomes (Rao and Wu, 2009*a*):

$$DEFF^{GR}(\hat{\theta}) = V^{GR}(\hat{\theta})\{V^{GR}_{SRS}(\hat{\theta})\}^{-1}, \tag{1.28}$$

where $V^{GR}(\hat{\theta})$ and $V^{GR}_{SRS}(\hat{\theta})$ are variances of residuals in a regression on the known parameter. The adjusted pseudo empirical log-likelihood ratio function

$$\hat{r}^{GR}_{ADJ}(\theta) = \hat{r}(\theta)\left\{DEFF^{GR}(\hat{\theta})\right\}^{-1} \tag{1.29}$$

follows a $\chi^2_{df=1}$ distribution asymptotically, when $\theta = \theta_U$ (Rao and Wu, 2009*a*).

The population empirical likelihood defines the benchmark constraints at the population level. A constraint based on a known mean $\bar{h}_N = \sum_{i=1}^{N} h_i(x_i)$ takes the following form:

$$\sum_{i=1}^{N} p_i \delta_i \pi_i^{-1}(h_i(x_i) - \bar{h}_N) = 0, \tag{1.30}$$

where δ_i is the sample inclusion indicator.

1.4 Design-based empirical likelihood

Design-based empirical likelihood (Berger and De La Riva Torres, 2016) was developed as an alternative to the pseudoempirical likelihood and population empirical likelihood methods. It overcomes the need for estimation of design effects by incorporating sampling probabilities into the constraints system rather than into the likelihood function. This leads to a likelihood ratio function which asymptotically follows a χ^2 distribution when the parameter equals the true population parameter of interest. This is particularly useful for estimation of complex and multivariate parameters.

The (potentially multivariate) parameter of interest $\boldsymbol{\theta}_U$ is defined as the solution

to a population estimating equation of the following form:

$$\sum_{i \in U} \boldsymbol{g}_i(\boldsymbol{\theta}) = \mathbf{0}_\nu, \tag{1.31}$$

where $\mathbf{0}_\nu$ is a vector of zeros of dimension ν, ν is the dimension of the parameter $\boldsymbol{\theta}$ and $\boldsymbol{g}_i(\boldsymbol{\theta})$ is a p-vector function of the parameter of interest and the sample variables.

Estimating equations are a flexible way of representing a wide class of parameters, such as means, totals, quantiles, ratios or generalised regression coefficients. Examples of estimating equations for various parameters can be found in Binder and Patak (1994), Qin and Lawless (1994) and Godambe and Thompson (2009). For example, the estimating function $\boldsymbol{g}_i(\boldsymbol{\theta}) = \boldsymbol{y}_i - \boldsymbol{\theta}$, can be used to estimate the population mean $N^{-1} \sum_{i \in U} \boldsymbol{y}_i$, leading to a Hájek (1964) estimator (Berger and Tillé, 2009).

The design-based empirical log-likelihood function is defined as

$$\ell(\boldsymbol{m}) = \sum_{i \in \boldsymbol{S}} \log(m_i), \tag{1.32}$$

where $\boldsymbol{m} = (m_1, m_2, ..., m_n)$ and m_i are the unit mass loads estimated under $m_i > 0$ and the design constraint

$$\sum_{i \in \boldsymbol{S}} m_i \pi_i = n, \tag{1.33}$$

where π_i is the sampling probability for unit i. Constraint (1.33) is different from the normalisation constraint used in other empirical likelihood approaches, where the scale loads are required to sum to 1.

Note that (1.32) can be re-parametrised as a function of the probability masses p_i. If we express m_i as $m_i = n p_i \pi_i^{-1}$ we have that (Berger and De La Riva Torres,

2016, Addendum)

$$\ell(\boldsymbol{m}) = \ell(\boldsymbol{p}) + \sum_{i \in \boldsymbol{S}} \log(n\pi_i^{-1}), \tag{1.34}$$

where $\ell(\boldsymbol{p}) = \sum_{i \in \boldsymbol{S}} \log(p_i)$ and the values p_i are estimated by maximising $\ell(\boldsymbol{p})$ under $p_i > 0$ and $\sum_{i \in \boldsymbol{S}} p_i = 1$. Maximising (1.32) and (1.34) is equivalent, as the offset $\sum_{i \in \boldsymbol{S}} \log(n\pi_i^{-1})$ does not depend on θ or on p_i.

The design-based empirical log-likelihood ratio function (Berger and De La Riva Torres, 2016) for the parameter $\boldsymbol{\theta}$ is defined as

$$\widehat{r}(\boldsymbol{\theta}) = 2\left[\ell(\hat{\boldsymbol{m}}) - \ell\{\hat{\boldsymbol{m}}^*(\boldsymbol{\theta})\}\right], \tag{1.35}$$

where $\ell(.)$ is defined by (1.32), $\hat{\boldsymbol{m}}$ are the values $\hat{m}_i$ which maximise (1.32) under $m_i > 0$ and the design constraint (1.33) and the $\hat{\boldsymbol{m}}^*(\boldsymbol{\theta})$ are the values $\hat{m}_i^*(\boldsymbol{\theta})$ which maximise (1.32) under $m_i > 0$, (1.33) and the additional constraint

$$\sum_{i \in \boldsymbol{S}} m_i \boldsymbol{g}_i(\boldsymbol{\theta}) = \mathbf{0}\cdot \tag{1.36}$$

The maximum design-based empirical likelihood estimator of $\boldsymbol{\theta}$ is defined as the value $\widehat{\boldsymbol{\theta}}$ which minimises (1.35). In high entropy sampling designs (Hájek, 1981) and under some regularity conditions, the design-based empirical log-likelihood ratio function (1.35) follows a χ^2 distribution asymptotically, with the number of degrees of freedom depending on the dimension of $\boldsymbol{\theta}$ (Berger and De La Riva Torres, 2016). This allows us to obtain confidence regions by selecting values

$$\left\{\boldsymbol{\theta} : \widehat{r}(\boldsymbol{\theta}) \leqslant \chi^2_{df=p}(\alpha)\right\}, \tag{1.37}$$

where $\chi^2_{df=p}(\alpha)$ is the upper α quantile of the $\chi^2_{df=p}$ distribution. The univariate confidence intervals can be obtained directly if the univariate parameter of interest is completely defined by a single (univariate) estimating equation. If the parameter

of interest depends on other parameters, profiling may be used (Oguz-Alper and Berger, 2016).

Benchmark constraints constructed around the known population parameters are handled in a similar way as in the pseudoempirical likelihood approach, however the parameters are defined through estimating equations, which gives more flexibility in the choice of the constraint. Suppose that a q-vector of population parameters $\boldsymbol{\varphi}_U$ is known, where $\boldsymbol{\varphi}_U$ is defined as the unique solution of

$$\sum_{i \in U} \mathbf{f}_i(\boldsymbol{x}_i, \boldsymbol{\varphi}) = \mathbf{0}_q, \tag{1.38}$$

and where the vector $\mathbf{f}_i(\boldsymbol{x}_i, \boldsymbol{\varphi})$ is a q-vector function of $\boldsymbol{x}_i$ and $\boldsymbol{\varphi}_U$ (e.g. Owen, 1991; Chaudhuri et al., 2008; Lesage, 2011). The benchmark constraint on $\boldsymbol{\varphi}_U$ takes the following form

$$\sum_{i \in \boldsymbol{S}} m_i \mathbf{f}_i(\boldsymbol{x}_i, \boldsymbol{\varphi}) = \mathbf{0}_q \cdot \tag{1.39}$$

For the generalised calibration property to hold on the design-based empirical likelihood adjusted weights $\hat{m}_i$, constraint (1.39) is imposed on the values $\hat{m}_i$ and $\hat{m}_i^*(\boldsymbol{\theta})$, alongside the design constraint (1.33) and the constraint (1.36) (Berger and De La Riva Torres, 2016).

Design-based empirical likelihood handles stratification by defining the design constraint (1.33) separately for each strata:

$$\sum_{i \in \boldsymbol{S}_h} m_{h;i} \pi_{h;i} = n_h, \quad h = 1, 2, ..., H \cdot \tag{1.40}$$

For cluster sampling designs, Oguz-Alper and Berger (2016) propose to use the ultimate cluster approach (Hansen et al., 1953), which defines the empirical likelihood function at the ultimate cluster level. In chapter 4 we show how the ultimate cluster approach is used in clustered and stratified samples from census coverage

survey.

The design-based empirical likelihood method has also been extended to handle non-replacement sampling with large sampling fractions. This requires using the penalised empirical log-likelihood function (Berger and De La Riva Torres, 2016)

$$\widetilde{\ell}(\boldsymbol{m}) := \sum_{i \in \boldsymbol{S}} \{\log(m_i) + 1 - \pi_i m_i\} \tag{1.41}$$

and adding finite population correction factors (Hájek, 1964) into constraints (1.33), (1.36) and (1.39). With these adjustments, the asymptotic distribution of the design-based empirical log-likelihood ratio function holds for large sampling fractions. We omit the details of the specification of the constraints system here for brevity. However, in chapter 2, we show how this approach can be used to handle large sampling fractions in the special case of aligning estimates from multiple samples.

The design-based empirical likelihood approach has several practical advantages. In particular, it allows to construct confidence intervals without the need for variance estimation. This is a very desirable feature, as variance estimators may be biased when the variables of interest are skewed. The confidence intervals are range-preserving and defined by the shape of the sample data. The design-based empirical likelihood approach can be used in complex, i.e., stratified and clustered, sampling designs, as long as these are high entropy designs (Hájek, 1981). These features of the method are utilised across the following chapters, where empirical likelihood methodology for several survey inference problems is defined.

Design-based empirical likelihood can be seen as an alternative to pseudoempirical likelihood or population empirical likelihood approaches, in that it also handles unequal probabilities and complex sampling designs. The use of estimating equations, incorporation of the sampling probabilities in the constraint system rather than in the log-likelihood function and the construction of confidence intervals based on a χ^2 approximation of the log-likelihood ratio function makes design-

based empirical likelihood more similar to population empirical likelihood than to pseudoempirical likelihood. The difference between the two approaches is in the definition of the log-likelihood function (which is defined at the sample level in design-based empirical likelihood) and in the design constraint used. Both approaches can handle complex parameters, defined as solutions of estimating equations. Design-based empirical likelihood, however, can be applied to samples selected with non-negligible sampling fraction and is closer to Owen's (1988) original formulation of the sample level empirical likelihood.

In the following chapters we extend the design-based empirical likelihood methodology to handle some specific problems in survey inference. For brevity, we use the term empirical likelihood to denote the design-based empirical likelihood. Whenever we refer to other empirical likelihood approaches, these are clearly referenced as either pseudoempirical likelihood or population empirical likelihood.

Chapter 2

Empirical likelihood approach for aligning estimates from multiple surveys

2.1 Introduction

Suppose that two independent samples, $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$, are selected from the same finite population U of size N. Let $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ be vectors of variables observed respectively in $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$. Let $\boldsymbol{w}$ denote a vector of common variables observed in both samples, which constitutes the key feature of the considered approach. A population parameter of a variable $\boldsymbol{w}$, e.g. the mean of $\boldsymbol{w}$, can be estimated either from $\boldsymbol{S}_1$ or $\boldsymbol{S}_2$. It can, however, be inconvenient to obtain different estimates for the same parameter, especially if other estimates are based on them. For example, suppose that $\boldsymbol{w}$ is a vector of age-sex categories measured in both samples. The two samples may not give the same estimates for the proportion within each category. A similar situation occurs if totals of turnover for various industries are estimated from sample $\boldsymbol{S}_1$, while sample $\boldsymbol{S}_2$ is used to estimate the overall population turnover. These domain-specific estimates from $\boldsymbol{S}_1$ do not necessarily add up to the overall total estimated from $\boldsymbol{S}_2$.

It is, of course, possible to obtain a composite estimate for the common parameter by taking a weighted average of the estimates obtained from two samples. The

weight applied to each survey's estimate might be selected based on an efficiency argument, e.g. inversely proportional to the estimated variance or proportional to the sample size. However, in practice it is desirable to have a single vector of weights for each survey which can be applied to all survey variables. Furthermore, it is a common practice that auxiliary variables are measured in surveys and survey weights are calibrated on known population parameters (see Deville and Särndal (1992*a*)).

The problem can therefore be summarised as follows: how to adjust the design weights of both surveys so that both calibration constraints (i.e., benchmarking on known population parameters) and alignment constraints (i.e., numerical consistency of common parameters) are respected, and inference about the common and non-common variables is possible. Apart from providing numerical consistency of estimates, alignment constraints might as well improve precision of the estimates of the non-common parameters, if the common and non-common variables are highly correlated. Specifically, when one of the samples is smaller, imposing alignment constraints on the variables shared with a larger sample is likely to improve precision of the smaller sample estimates. This property is exploited in the split questionnaire design or non nested two-phase sampling, where a subset of variables is measured for a large sample, and the whole set of variables is collected from another, smaller sample (see e.g. Hidiroglou (2001)).

The procedure of adjusting survey weights so that estimates of two surveys agree with each other is often referred to as 'alignment'. The traditional methods used to include auxiliary information on known population quantities in the single sample case cannot be directly applied to aligning estimates from two or more surveys. Certain adjustments, which account for the added complexity, need to be made. This special situation has been studied extensively and several design-based methods have been proposed. Zieschang's (1990) and Renssen and Nieuwenbroek's (1997), as well as Merkouris's (2004) methods are based on the generalized calibration estimator. Zieschang (1990) and Renssen and Nieuwenbroek (1997) estimate

the unknown population mean of $\boldsymbol{w}$ by a linear combination of two estimates calculated from $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$. This linear combination is then used as a benchmark parameter in a regression estimator. Merkouris (2004) proposed '*composite regression estimator*' of a total of $\mathbf{y}$, which is based on a simultaneous regression using data of $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ pooled together, avoiding the estimation of the means of $\boldsymbol{w}$ as an intermediate step. In Zieschang's (1990), Renssen and Nieuwenbroek's (1997) and Merkouris (2004) approaches, symmetric confidence intervals are constructed based on suitably adjusted variance estimates.

Wu (2004*a*) proposed an estimator for means of $\mathbf{y}_1$ and $\mathbf{y}_2$ based on aligned pseudoempirical likelihood weights. Symmetric confidence intervals are created using the variance estimate for the asymptotically equivalent regression estimator. Chen and Kim (2014) developed an aligned population empirical likelihood approach, based on an empirical likelihood function defined at population level, and proposed an empirical log-likelihood ratio statistic, which is pivotal under Poisson sampling with negligible sampling fraction. Methods outside of the design-based paradigm have also been proposed, see e.g. Kim and Rao (2012) for a model-assited approach, Kim et al. (2015) for a model based small area application and Dong et al. (2014) for a bayesian bootstrap approach.

We propose a new aligned design-based empirical likelihood approach. The proposed approach is different from Zieschang's (1990), Renssen and Nieuwenbroek's (1997), Merkouris's (2004) and Wu's (2004*a*) methods as it considers a general class of parameters which are defined by estimating equations, rather than means or totals, and allows for construction of Wilks (1938) type confidence intervals. It also differs from Chen and Kim's (2014) approach in that it is defined at the sample level, does not require the population size to be known and can easily be applied to designs with large sampling fractions and stratification.

The proposed approach treats the empirical likelihood function as a standard likelihood. Point estimates are obtained by maximising this function. Confidence intervals are obtained from an empirical log-likelihood ratio function rather than

through variance estimation. The proposed method does not require knowledge of the population size and does not rely on the estimation of the population mean of $\boldsymbol{w}$. It is valid under without-replacement stratified sampling with small or large sampling fractions.

The presented method has some practical advantages. Confidence intervals are range-preserving and their construction does not require variance estimates, unlike the pseudoempirical likelihood and the composite regression confidence intervals. Simulation studies presented in chapter 2.12 show that the empirical likelihood confidence intervals have good coverage across a range of scenarios. The proposed approach can accommodate different functions of the common parameter (as opposed to just a mean or a total), making it possible to choose the function that is highly correlated with the parameter of interest. The empirical likelihood weights are always positive.

The proposed method is derived from Berger and De La Riva Torres's (2016) empirical likelihood methodology for construction of confidence intervals in a single sample case, in presence of benchmark constraints and under complex sampling designs (see chapter 1 for a brief summary of this approach). However, the core problem tackled here is different. Berger and De La Riva Torres (2016) deal with a traditional setup when a single sample is considered and benchmark constraints involve only known population parameters. We focus on alignment of two samples and allow for constraints including unknown (yet not necessarily nuisance) parameters.

The following chapters introduce the proposed empirical likelihood approach for aligning information from multiple surveys. Chapter 2.2 explains the sampling design and variables measured. Chapter 2.3 describes some alternative approaches to parameter estimation under alignment constraints. Chapters 2.4 - 2.11 introduce the proposed aligned empirical likelihood estimator and discuss its properties. Numerical results from Monte Carlo simulations performed on artificial and real datasets are presented in chapter 2.12.

2.2 Sampling design and data collected

Suppose that two independent surveys are carried out in a finite population U of size N. The samples $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are selected independently, where $\boldsymbol{S}_t$ denotes the sample selected after n_t independent random draws from population U. For simplicity, we start with the assumption that units are selected with unequal probabilities, with replacement, or without replacement with negligible sampling fractions, i.e., $n_1 N^{-1} \to 0$ and $n_2 N^{-1} \to 0$. In chapter 2.11 we discuss how the proposed method can be applied in the commonly used without replacement sampling designs with large sampling fractions. We consider non stratified sampling designs first. In chapter 2.10 we show how the proposed method can accommodate stratified sampling designs.

Let $\boldsymbol{y}_{t;i}$ be the value of the variable $\boldsymbol{y}$ measured in the t-th survey for the i-th unit, $i = 1, ..., n_t$ and $\pi_{t;i}$ be the first order selection probability for the i-th unit in the t-th survey. The samples may or may not overlap, because same population units may or may not be selected in both samples.

Let $\boldsymbol{S}$ of size $n = \sum_{t=1}^{T} n_t$ be the collection of labels of all units selected in all the T samples, i.e., a 'pooled' multiset of labels of $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$. If a unit is selected k times, its label appears k times in $\boldsymbol{S}$.

Suppose that the values of a set of variables, denoted by $\boldsymbol{v}_t$, are collected from the sample $\boldsymbol{S}_t$ and that $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ contain at least one common variable. The set $\boldsymbol{v}_t$ is composed of four types of variables: $\boldsymbol{z}_t$, $\boldsymbol{y}_t$, $\boldsymbol{x}_t$ and $\boldsymbol{w}$; that is, $\boldsymbol{v}_1 \equiv \{\boldsymbol{z}_1, \boldsymbol{y}_1, \boldsymbol{x}_1, \boldsymbol{w}\}$ and $\boldsymbol{v}_2 \equiv \{\boldsymbol{z}_2, \boldsymbol{y}_2, \boldsymbol{x}_2, \boldsymbol{w}\}$. The variables $\boldsymbol{z}_t$ denote the design variables, which include unit sampling probabilities. The variables $\boldsymbol{x}_t$ denote auxiliary variables. The variables $\boldsymbol{w}$ denote the common variables which are included in both $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ (see Figure 2.2). Other variables in the sample are denoted by $\boldsymbol{y}_t$. Some of them might be the variables of interest. The existence of at least one common variable is the key aspect of the considered problem.

Figure 2.2 shows a visualisation of the variables and population units. The horizontal axis represents the variables and the vertical axis represents the population units $\{1, \ldots, N\}$. The shaded areas represent the two samples. In Figure 2.2, there is no overlap between the two samples. In practice, some overlap is possible, as the samples are selected independently. The overlap does not play any role in inference and does not need to be known.

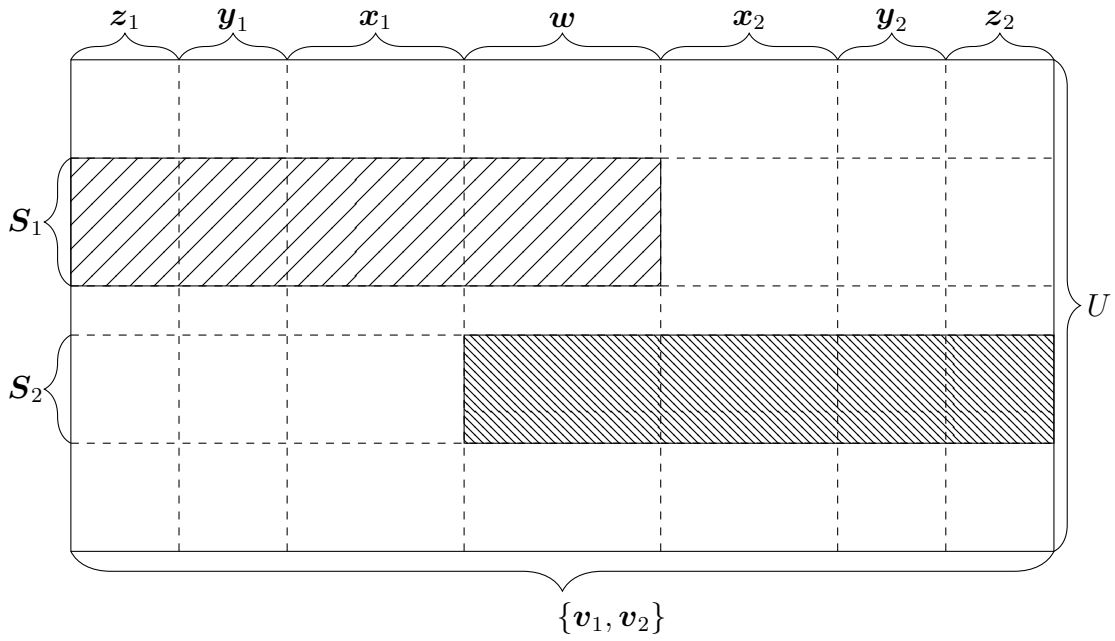


Figure 2.1: *Sample data and parameters of the samples $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$. The horizontal axis corresponds to the variables: $\boldsymbol{z}_t$, $\boldsymbol{y}_t$, $\boldsymbol{x}_t$ and $\boldsymbol{w}$. The vertical axis represents the labels of the units in population U. The area* ▨ *represents the data sampled in $\boldsymbol{S}_1$. The area* ▧ *represents the data sampled in $\boldsymbol{S}_2$.*

We adopt a design-based approach, were the $\boldsymbol{v}_{ti}$ are fixed quantities and sampling is the only source of randomness (Neyman, 1934). The distribution of the sample $\boldsymbol{S}_t$ is specified by the probability distribution of $\boldsymbol{S}_t$, which is denoted by $\mathcal{P}_t(\boldsymbol{S}_t)$. Note that the observations are not independent and identically distributed. We follow Hartley and Rao's (1968) framework under which the population labels are non-informative.

Let $\boldsymbol{\theta}_{tU}$ be a fixed, unknown population parameter of interest, a function of $\boldsymbol{v}_t$. Let parameter $\boldsymbol{\theta}_U$ be a concatenation of parameters of interest related to each of

the samples, that is, $\boldsymbol{\theta}_U = \left(\boldsymbol{\theta}_{1U}^\top, \boldsymbol{\theta}_{2U}^\top\right)^\top$. It follows that $\boldsymbol{\theta}_U$ is a function of the values of the variables $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$. The parameter of interest $\boldsymbol{\theta}_U$ is defined as the solution of the following system of population estimating equations.

$$\boldsymbol{G}(\boldsymbol{\theta}) := (\boldsymbol{G}_1(\boldsymbol{\theta}_1)^\top, \boldsymbol{G}_2(\boldsymbol{\theta}_2)^\top)^\top = \mathbf{0}_\nu, \tag{2.1}$$

where $\nu = \nu_1 + \nu_2$,

$$\boldsymbol{G}_t(\boldsymbol{\theta}_t) := \sum_{i \in U} \boldsymbol{g}_{ti}(\boldsymbol{v}_{ti}, \boldsymbol{\theta}_t) = \mathbf{0}_{\nu_t}, \quad t = 1, 2, \tag{2.2}$$

and $\mathbf{0}_\nu$ denotes an ν-vector of zeros. The vector $\boldsymbol{g}_{ti}(\boldsymbol{v}_{ti}, \boldsymbol{\theta}_t)$ is a ν_t-vector function of a subset of variables $\boldsymbol{v}_{ti}$. We assume that the $\boldsymbol{g}_{ti}(\boldsymbol{v}_{ti}, \boldsymbol{\theta}_t)$ are such that the solution of (2.2) is unique and that $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ denotes the parameter space of $\boldsymbol{\theta}_U$. Various definitions of the function $\boldsymbol{g}_{ti}(\boldsymbol{v}_{ti}, \boldsymbol{\theta}_t)$ are possible. For example, when $\boldsymbol{g}_{ti}(\boldsymbol{v}_{ti}, \boldsymbol{\theta}_t) = \boldsymbol{y}_{t;i} - \boldsymbol{\theta}_t$, the unique solution of (2.1) is a vector of Hájek - type estimates of the population means of $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$. When $\boldsymbol{g}_{ti}(\boldsymbol{v}_{ti}, \boldsymbol{\theta}_t) = \boldsymbol{y}_{t;i} - Nn_t^{-1}\pi_{t;i}\boldsymbol{\theta}_t$, we get the Horvitz-Thompson estimates of the population means. When $\boldsymbol{g}_{ti}(\boldsymbol{v}_{ti}, \boldsymbol{\theta}_t) = \boldsymbol{y}_{t;i} - n_t^{-1}\pi_{t;i}\boldsymbol{\theta}_t$, we obtain the estimates of the population totals. Note that the parameters and estimating equations based on each sample can be different, e.g. because different variables are measured in each sample, or because different functions are of interest. To simplify the notation, we replace $\boldsymbol{g}_{ti}(\boldsymbol{v}_{ti}, \boldsymbol{\theta}_t)$ by $\boldsymbol{g}_{ti}(\boldsymbol{\theta}_t)$, in the following text.

Let δ_{ti} be the sample inclusion indicator

$$\delta_{ti} := \begin{cases} 1 & \text{if } i \in \boldsymbol{S}_t \\ 0 & \text{otherwise}\cdot \end{cases} \tag{2.3}$$

Let

$$\boldsymbol{g}_i(\boldsymbol{\theta}) \;\; := \;\; \left(\; \delta_{1i} \; \boldsymbol{g}_{1i}(\boldsymbol{\theta}_1)^\top \; , \; \delta_{2i} \; \boldsymbol{g}_{2i}(\boldsymbol{\theta}_2)^\top \; \right)^\top \tag{2.4}$$

and

$$\pi_i := \delta_{1i}\pi_{1i} + \delta_{2i}\pi_{2i}. \tag{2.5}$$

Note that (2.5) means that $\pi_i := \pi_{1i}$ if $i \in \boldsymbol{S}_1$ and $\pi_i := \pi_{2i}$ if $i \in \boldsymbol{S}_2$.

Very often some population level parameters, such as totals, means, ratios or counts are known for the whole population or for a specific domain, e.g. from census or administrative records. Suppose that a q_t-vector of population parameters $\boldsymbol{\varphi}_{tU}$ is known. The known parameter $\boldsymbol{\varphi}_{tU}$ is defined as the unique solution of

$$\sum_{i \in U} \mathbf{f}_{ti}(\boldsymbol{x}_{ti}, \boldsymbol{\varphi}_{tU}) = \mathbf{0}_{q_t}, \tag{2.6}$$

where the vector $\mathbf{f}_{ti}(\boldsymbol{x}_{ti}, \boldsymbol{\varphi}_{tU})$ is a q_t-vector function of $\boldsymbol{x}_{ti}$ and $\boldsymbol{x}_{ti}$ are selected components of $\boldsymbol{v}_{ti}$. To simplify the notation, we replace $\mathbf{f}_{ti}(\boldsymbol{x}_{ti}, \boldsymbol{\varphi}_{tU})$ by $\mathbf{f}_{ti}(\boldsymbol{\varphi}_{tU})$, in the following text. Let

$$\boldsymbol{\varphi}_{U} := \left(\boldsymbol{\varphi}_{1U}^\top, \boldsymbol{\varphi}_{2U}^\top\right)^\top \tag{2.7}$$

denote the overall q-vector of known parameters, with $q = q_1 + q_2$.

Consider the following sample level estimating equation:

$$\widehat{\mathbf{F}}_\pi(\boldsymbol{\varphi}) = \sum_{i \in \boldsymbol{S}} d_i \mathbf{f}_i(\boldsymbol{\varphi}) = \mathbf{0}_q \tag{2.8}$$

where $d_i = \pi_i^{-1}$ are the design weights, π_i are defined by (2.5) and $\mathbf{f}_i(\boldsymbol{\varphi})$ is defined in an analogous way to $\boldsymbol{g}_i(\boldsymbol{\theta})$. The estimate $\widehat{\boldsymbol{\varphi}}$ of $\boldsymbol{\varphi}_U$ is obtained as the value which solves (2.8). In a general case, there is no guarantee that $\widehat{\boldsymbol{\varphi}}$ is equal to $\boldsymbol{\varphi}_U$. Adjusted weights p_i are said to possess the generalized calibration property (e.g. Owen, 1991; Chaudhuri et al., 2008) if the solution to the equation

$$\widehat{\mathbf{F}}(\boldsymbol{\varphi}) = \sum_{i \in \boldsymbol{S}} p_i \mathbf{f}_i(\boldsymbol{\varphi}) = \mathbf{0}_q \tag{2.9}$$

is equal to $\boldsymbol{\varphi}_U$. The constraint

$$\widehat{\boldsymbol{\varphi}} = \boldsymbol{\varphi}_U \tag{2.10}$$

is called a *benchmark* or *calibration* constraint (Deville and Särndal, 1992*a*). Note that (2.10) simplifies to (1.23) if $\boldsymbol{\varphi}_U$ is a population total of variable $\boldsymbol{x}$. Benchmark constraints are often used in survey practice, especially in official statistics. If the variable of interest and the auxiliary variables are highly correlated, benchmark constraints on $\boldsymbol{\varphi}$ might improve the precision of the estimator $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}_U$, defined as the solution to the sample estimating equation

$$\widehat{\boldsymbol{G}}(\boldsymbol{\theta}) = \sum_{i \in \boldsymbol{S}} p_i \boldsymbol{g}_i(\boldsymbol{\theta}) = \mathbf{0}_\nu \cdot \tag{2.11}$$

Sometimes the benchmark constraints are also used for practical reasons (e.g. in order to obtain numerical consistency with values published from a census or administrative sources). Note that the $\mathbf{f}_i(\boldsymbol{\varphi})$ cannot be a function of $\boldsymbol{\theta}$, i.e., all its components need to be known.

Let $\boldsymbol{\xi}_U$ be the unknown population mean $N^{-1} \sum_{i \in U} \boldsymbol{\xi}(\boldsymbol{w}_i)$ of a known function $\boldsymbol{\xi}$ of the common variable $\boldsymbol{w}$. For example, we may have $\boldsymbol{\xi}(\boldsymbol{w}_i) = \boldsymbol{w}_i$, or $\boldsymbol{\xi}(\boldsymbol{w}_i) = \boldsymbol{w}_i^2$, or $\boldsymbol{\xi}(\boldsymbol{w}_i) = \delta(\boldsymbol{w}_i \leq \alpha)$, where $\delta(\cdot)$ is an indicator function equal to 1 if the argument is true and to 0 otherwise and α is a known constant, e.g. a quantile of distribution. Suppose for now that we have $\boldsymbol{\xi}(\boldsymbol{w}_i) = \boldsymbol{w}_i$, that is, $\boldsymbol{\xi}_U$ is the population mean of the variable $\boldsymbol{w}$. A more general case will be discussed in chapters 2.4 and 2.5.

We can estimate $\boldsymbol{\xi}_U$ from an estimating equation based on either of the two samples. Let $\widehat{\boldsymbol{\xi}}_t$ be the solution of the estimating equation based on sample t:

$$\widehat{\mathbf{H}}_{t\pi}(\boldsymbol{\xi}) = \sum_{i \in \boldsymbol{S}_t} d_{t;i} \boldsymbol{h}_{t;i}(\boldsymbol{w}_{t;i}, \boldsymbol{\xi}) = \mathbf{0}_{r_t}, \tag{2.12}$$

where $d_{t;i} = \pi_{t;i}^{-1}$ are the design weights, r_t is the dimension of vector $\boldsymbol{h}_{t;i}(\boldsymbol{w}_{t;i}, \boldsymbol{\xi})$ and $\boldsymbol{h}_i(\boldsymbol{w}_i, \boldsymbol{\xi}) := \boldsymbol{w}_i - N n_t^{-1} \pi_{ti}\, \boldsymbol{\xi}$, for $t = 1$ and 2. A similar estimating equation

was proposed by Berger and De La Riva Torres (2016). Solving (2.12) with respect to $\boldsymbol{\xi}$ gives a Horvitz-Thompson - type estimate of the population mean (Berger and De La Riva Torres, 2016)

$$\widehat{\boldsymbol{\xi}}_t = n_t N^{-1} \sum_{i \in \boldsymbol{S}_t} d_{t;i} \boldsymbol{w}_i \left(\sum_{i \in \boldsymbol{S}_t} d_{t;i} \pi_{ti} \right)^{-1}. \tag{2.13}$$

The estimate $\widehat{\boldsymbol{\xi}}_t$ can be obtained either from $\boldsymbol{S}_1$ or $\boldsymbol{S}_2$. The estimates $\widehat{\boldsymbol{\xi}}_1$ and $\widehat{\boldsymbol{\xi}}_2$ obtained from each of the samples are not guaranteed to be equal.

We define the alignment property as the requirement that

$$\widehat{\boldsymbol{\xi}}_1 = \widehat{\boldsymbol{\xi}}_2. \tag{2.14}$$

In the next chapters we propose an empirical likelihood method for obtaining the adjusted design weights p_{ti} such that the resulting estimators have the calibration property, as defined by equation (2.10), and the alignment property (2.14). While the estimators which possess the calibration property are commonly used and a design-based empirical likelihood estimator with this property has already been proposed (see chapter 1 and Berger and De La Riva Torres (2016)), the estimators with the alignment property are not so common. Below we discuss some existing estimators which have the alignment property, including the generalized regression type estimators (Zieschang, 1990; Renssen and Nieuwenbroek, 1997; Merkouris, 2004), as well as the pseudoempirical likelihood (Wu, 2004*a*) and population empirical likelihood (Chen and Kim, 2014) estimators.

2.3 Some existing approaches

There are two main types of design-based estimators that ensure alignment of estimates from two or more surveys: the generalized regression family estimators,

including the methods of Zieschang (1990); Renssen and Nieuwenbroek (1997) and Merkouris (2004) and the empirical likelihood type methods, namely the pseudo empirical likelihood estimator (Wu, 2004*a*) and the population empirical likelihood estimator (Chen and Kim, 2014). Historically the first method was proposed by Zieschang (1990) for alignment of totals of the common variables. The method consists of two steps. First, a composite estimate of the total of the common variable is calculated as a linear combination of regression estimates obtained from each of the samples. That is, for a scalar common variable w, the composite estimator of the population total $W = \sum_{i \in U} w_i$ takes the following form:

$$\hat{W}_{CR} = \phi \hat{W}_1^R + (1 - \phi)\hat{W}_2^R, \tag{2.15}$$

where $\hat{W}_t^R$ is a regression estimator of W calculated from the t-th sample and ϕ is a scaling factor between 0 and 1. In the second step an additional calibration type constraint is included in the extended regression system for estimation of any non-common parameters. In other words, each of the samples is calibrated on the same composite estimate of the total of the common variable (Merkouris, 2004).

Renssen and Nieuwenbroek (1997) proposed an optimal weighting coefficient ϕ for the linear combination (2.15), based on the approximate variances of estimators obtained from both surveys:

$$\phi = \frac{\hat{V}(\hat{W}_2^R)}{\hat{V}(\hat{W}_1^R) + \hat{V}(\hat{W}_2^R)}. \tag{2.16}$$

These results were further extended by Merkouris (2004), who proposed a method which does not require the intermediate step of estimating the total of the common variable and does not require estimating the variances as in (2.16). Merkouris (2004) also provided a generalisation of the available approaches to estimation of population totals of the non-common variables. The composite regression estimators for the totals of variables y_1 and y_2 take the following form Merkouris

(2004):

$$\hat{Y}_1^{CR} = \hat{Y}_1^R + \hat{\boldsymbol{B}}_{y_1}(\boldsymbol{I} - \boldsymbol{\Phi})(\hat{\boldsymbol{W}}_2^R - \hat{\boldsymbol{W}}_1^R) \tag{2.17}$$

$$\hat{Y}_2^{CR} = \hat{Y}_2^R + \hat{\boldsymbol{B}}_{y_2}\boldsymbol{\Phi}(\hat{\boldsymbol{W}}_2^R - \hat{\boldsymbol{W}}_1^R), \tag{2.18}$$

where

$$\hat{Y}_t^R = \hat{Y}_t^{HT} + \boldsymbol{y}_t^\top \boldsymbol{\Lambda}_t \boldsymbol{x}_t (\boldsymbol{x}_t^\top \boldsymbol{\Lambda}_t \boldsymbol{x}_t)^{-1}(\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t), \tag{2.19}$$

$$\hat{\boldsymbol{W}}_t^R = \hat{\boldsymbol{W}}_t^{HT} + \boldsymbol{w}_t^\top \boldsymbol{\Lambda}_t \boldsymbol{x}_t (\boldsymbol{x}_t^\top \boldsymbol{\Lambda}_t \boldsymbol{x}_t)^{-1}(\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t) \tag{2.20}$$

$$\hat{\boldsymbol{B}}_{y_s} = \boldsymbol{y}_t^\top \boldsymbol{L}_{s_t} \boldsymbol{w}_t (\boldsymbol{w}_t^\top \boldsymbol{L}_{s_t} \boldsymbol{w}_t)^{-1} \tag{2.21}$$

$$\boldsymbol{L}_t = \boldsymbol{\Lambda}_t (\boldsymbol{I} - \boldsymbol{x}_t (\boldsymbol{x}_t^\top \boldsymbol{\Lambda}_t \boldsymbol{x}_t)^{-1} \boldsymbol{x}_t^\top \boldsymbol{\Lambda}_t) \tag{2.22}$$

and $\boldsymbol{\Lambda}_t$ is a diagonal weighting matrix with the ii-th entry equal to $\pi_{t;i}^{-1}$. The $\hat{Y}_t^R$ is a generalized regression estimator of the total Y_t and $*_t$ denotes values of variable $*$ observed in sample $\boldsymbol{S}_t$. The coefficient $\boldsymbol{\Phi}$ is an adjustment factor which can take different forms.

When $\boldsymbol{\Phi} = \gamma \boldsymbol{I}$, with γ being a scaling coefficient ranging from 0 to 1, we obtain Renssen and Nieuwenbroek's (1997) estimator (Merkouris, 2004). In particular, the coefficient

$$\boldsymbol{\Phi} = \hat{V}(\hat{\boldsymbol{W}}_2^R)\{\hat{V}(\hat{\boldsymbol{W}}_1^R) + \hat{V}(\hat{\boldsymbol{W}}_2^R)\}^{-1}, \tag{2.23}$$

minimises the estimated variance of the composite estimate of the total of the common variable (Renssen and Nieuwenbroek, 1997; Merkouris, 2004).

Using

$$\boldsymbol{\Phi} = \boldsymbol{w}_2^\top \boldsymbol{\Lambda}_2 \boldsymbol{w}_2 (\boldsymbol{w}_1^\top \boldsymbol{\Lambda}_1 \boldsymbol{w}_1 + \boldsymbol{w}_2^\top \boldsymbol{\Lambda}_2 \boldsymbol{w}_2)^{-1} \tag{2.24}$$

yields Zieschang's (1990) estimator (Merkouris, 2004).

The choice of

$$\boldsymbol{\Phi} = \boldsymbol{w}_2^\top \boldsymbol{L}_2 \boldsymbol{w}_2 (\boldsymbol{w}_1^\top \boldsymbol{L}_1 \boldsymbol{w}_1 + \boldsymbol{w}_2^\top \boldsymbol{L}_2 \boldsymbol{w}_2)^{-1} \tag{2.25}$$

gives Merkouris's (2004) composite regression estimator, where $\boldsymbol{L}_t$ is defined by (2.22).

The estimator can be further modified to include a correction factor which accounts for the differences in the sample sizes and design effects between the surveys: (Merkouris, 2004)

$$\phi = \frac{n_1\{DEFF(\boldsymbol{S}_1)\}^{-1}}{n_1\{DEFF(\boldsymbol{S}_1)\}^{-1} + n_2\{DEFF(\boldsymbol{S}_2)\}^{-1}}, \tag{2.26}$$

where $DEFF(\boldsymbol{S}_t)$ is the design effect associated with the t-th sample. The factor ϕ is incorporated into (2.25) in the following way:

$$\boldsymbol{\Phi} = \phi \boldsymbol{w}_2^\top \boldsymbol{L}_2 \boldsymbol{w}_2 \{(1-\phi)\boldsymbol{w}_1^\top \boldsymbol{L}_1 \boldsymbol{w}_1 + \phi \boldsymbol{w}_2^\top \boldsymbol{L}_2 \boldsymbol{w}_2\}^{-1}. \tag{2.27}$$

This is equivalent to dividing the ij-th element in the matrix $\boldsymbol{\Lambda}_t$ by a factor $q_{t;i} = n_t\{DEFF(\boldsymbol{S}_t)\}^{-1}$ (Merkouris, 2010*a*).

Merkouris's (2004) estimator accounts for both the difference in variances of $\hat{\boldsymbol{W}}_1^R$ and $\hat{\boldsymbol{W}}_2^R$, and the different levels of regression fit in the $\hat{\boldsymbol{W}}_1^R$ and $\hat{\boldsymbol{W}}_2^R$ in (2.20) (Merkouris, 2004).

The approximate design variance of the estimator $\hat{Y}_1^{CR}$ is given by: (Merkouris, 2004)

$$\begin{aligned}\widehat{Var}(\hat{Y}_1^{CR}) = \widehat{Var}(\hat{Y}_1^R) + \boldsymbol{B}_{y_1}(\boldsymbol{I}-\boldsymbol{\Phi})\{\widehat{Var}(\hat{\boldsymbol{W}}_1^R) + \widehat{Var}(\hat{\boldsymbol{W}}_2^R)\}(\boldsymbol{I}-\boldsymbol{\Phi})^\top \boldsymbol{B}_{y_1}^\top \\ -2\boldsymbol{B}_{y_1}(\boldsymbol{I}-\boldsymbol{\Phi})\{\widehat{Cov}(\hat{Y}_1^R, \hat{\boldsymbol{W}}_1^R)\}\,.\end{aligned} \tag{2.28}$$

The relative efficiency of the estimators listed above has been discussed in detail by Merkouris (2004, 2010*a*,*b*, 2015). The main difference between Zieschang's (1990) estimator and Renssen and Nieuwenbroek's (1997) and Merkouris's (2004) estimators is that the latter two are corrected for the difference in efficiency of the two surveys. Renssen and Nieuwenbroek's (1997) estimator includes the correction factor in the coefficient $\boldsymbol{\Phi}$, which is proportional to the estimated relative variances of the regression estimators for the total of the common variable calculated from each of the samples. The practical implication of this is the necessity to estimate these variances before the final weights are obtained. Merkouris's (2004) point estimator does not rely on variance estimation. Instead, sample sizes (and design effects if different designs are used in the surveys), are included directly in the extended regression coefficient $\boldsymbol{B}$. In some cases, Merkouris's (2004) estimator is design optimal (Merkouris, 2004).

Merkouris (2015) discusses composite regression estimators that are minimum-variance linear unbiased combinations of estimators obtained from each sample. These estimators are called best linear unbiased estimators (BLUE) (see Chipperfield and Steel, 2009). If the samples $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are independent, the optimal weighting matrix $\boldsymbol{\Lambda}_t^0$ has the ij-th element equal to $(\pi ij - \pi_i\pi_j)(\pi_i\pi_j\pi_{ij})^{-1}$, where π_{ij} are second order sampling probabilities for units i and j (Merkouris, 2015). For dependent samples, estimation of variances and covariances of the estimators is necessary to obtain a BLUE.

Wu (2004*a*) proposed a pseudoempirical likelihood approach to aligning estimates of means from two surveys. The maximum pseudoempirical likelihood estimator of the mean $\bar{y}_t = N^{-1}\sum_{i\in\boldsymbol{S}_t} y_{t;i}$ is equal to:

$$\hat{\bar{y}}_t = \sum_{i\in\boldsymbol{S}_t} p_{t;i}y_{t;i}, \tag{2.29}$$

where the weights $p_{t;i}$ are estimated by the values which maximise the pseudoem-

pirical likelihood function:

$$\ell(\boldsymbol{p}) = \sum_{i \in \boldsymbol{S}_1} \pi_{1;i}^{-1} \log(p_{1;i}) + \sum_{i \in \boldsymbol{S}_2} \pi_{2;i}^{-1} \log(p_{2;i}) \tag{2.30}$$

under the following constraints:

$$\sum_{i \in \boldsymbol{S}_1} p_{1;i} = 1, \sum_{i \in \boldsymbol{S}_2} p_{2;i} = 1, \tag{2.31}$$

$$\sum_{i \in \boldsymbol{S}_1} p_{1;i} x_{1;i} = \bar{x}_1, \sum_{i \in \boldsymbol{S}_2} p_{2;i} x_{2;i} = \bar{x}_2, \sum_{i \in \boldsymbol{S}_1} p_{1;i} w_{1;i} = \sum_{i \in \boldsymbol{S}_2} p_{2;i} w_{2;i}, \tag{2.32}$$

where $\bar{x}_t$ is the known population mean of variable x_t. Wu (2004*a*) proposes two methods to compute the pseudoempirical likelihood weights $\hat{p}_{t;i}$. The first approach consists of using an iterative algorithm, where in each iteration first the maximum likelihood estimate of the mean of the common variable is calculated and then the estimated mean is used to construct a benchmark constraint. The second approach circumvents the necessity of estimating the unknown mean of the common variable and imposes constraint (2.32) directly.

The maximum pseudoempirical likelihood estimator is asymptotically equivalent to a regression estimator similar to the estimator proposed by Zieschang (1990), but creates weights which are positive by definition (Wu, 2004*a*). A version of the pseudoempirical likelihood estimator similar to Renssen and Nieuwenbroek's (1997) estimator with the optimal weighting coefficient, where the unknown mean of the common variable is estimated by a linear combination of the estimators obtained from separate samples and then used to construct a constraint on the estimator for the parameter of interest was also proposed (Wu, 2004*a*).

Chen and Kim (2014) proposed a population empirical likelihood method to combine information from non-nested two-phase sampling. The method involves finding the weights which maximise the population empirical likelihood function:

$$\ell(\boldsymbol{p}) = \sum_{i=1}^{N} \log(p_i), \tag{2.33}$$

where N is the population size, subject to the following population level constraints:

$$\sum_{i=1}^{N} p_i = 1 \tag{2.34}$$

$$\sum_{i=1}^{N} p_i x_{1;i}(\delta_{1;i}\pi_{1;i}^{-1} - 1) = 0, \sum_{i=1}^{N} p_i x_{2;i}(\delta_{2;i}\pi_{2;i}^{-1} - 1) = 0 \tag{2.35}$$

$$\sum_{i=1}^{N} p_i f_i(x_i, \varphi_U)(\delta_{1;i}\pi_{1;i}^{-1} - \delta_{2;i}\pi_{2;i}^{-1}) = 0 \tag{2.36}$$

$$\sum_{i=1}^{N} p_i \delta_{2;i}\pi_{2;i}^{-1} g_i(y_i, \theta_U) = 0, \tag{2.37}$$

where δ_{ti} is a sample membership indicator and equals 1 if unit i was selected in $\boldsymbol{S}_t$ and 0 otherwise. The $f_i(x_i, \varphi_U)$ and $g_i(y_i, \theta_U)$ are estimating functions for the known population parameter φ_U and the parameter of interest θ_U respectively.

Under Poisson sampling and rejective Poisson sampling, when the sampling fraction is negligible (i.e., $n_1 N^{-1} \to 0$ and $n_2 N^{-1} \to 0$), the maximum population empirical likelihood estimator of the parameter θ_U is asymptotically equivalent to the optimal Generalized Method of Moments estimator (Hansen, 1982). Under the above conditions and some regularity conditions (see Chen and Kim, 2014), the population empirical likelihood ratio function is pivotal and follows a χ^2 distribution asymptotically, which can be used to construct Wilks (1938) type confidence intervals.

Methods outside of the design-based paradigm have also been proposed. While it is beyond the scope of this work to characterise them all, examples include a model-assited approach of Kim and Rao (2012), a model based small area application by Kim et al. (2015) and a bayesian bootstrap approach by Dong et al. (2014).

The empirical likelihood approach proposed in the following paragraphs fills in some gaps in the currently available methods. It gives point estimators and confidence intervals for a wide class of parameters expressed as solutions to estimating

equations. It also allows to use various functions of the common variable in the alignment constraints. The regression-type estimators and the pseudoempirical likelihood estimator are restricted to estimation of means and totals. The empirical likelihood approach allows to construct asymmetric, range-preserving Wilks (1938) type confidence intervals based directly on the χ^2 approximation of the empirical likelihood ratio function, without any corrections. This is not the case for the regression type estimators and the pseudoempirical likelihood estimator, as they require variance estimates.

Estimation of parameters through the use of estimating equations as well as construction of asymmetric confidence intervals based on the asymptotic χ^2 distribution of a likelihood ratio function is possible with the population empirical likelihood approach. The proposed empirical likelihood approach, however, only considers sample data, does not require knowledge of the population size and can be used for estimation from stratified samples selected with large sampling fractions. It is also closer to the original formulation of empirical likelihood in that it is defined at the sample level.

2.4 Empirical likelihood approach proposed

In this chapter we develop an empirical likelihood method to obtain estimates for the parameter of interest $\boldsymbol{\theta}_U$, such that the benchmark constraints based on the known parameters $\boldsymbol{\varphi}_{1U}$ and $\boldsymbol{\varphi}_{2U}$, as well as the alignment constraint on the mean of a function of the common variable $\boldsymbol{w}$, are respected.

Consider the following two samples joint empirical log-likelihood function:

$$\ell(\boldsymbol{m}) := \sum_{i \in \boldsymbol{S}_1} \log(m_{1i}) + \sum_{i \in \boldsymbol{S}_2} \log(m_{2i}), \tag{2.38}$$

where $\log(\cdot)$ denotes the natural logarithm. The m_{ti} are unknown positive scale

loads (e.g. Hartley and Rao, 1969) associated with unit $i \in \boldsymbol{S}_t$. A similar joint empirical log-likelihood function was proposed by Owen (2001). Note that function (2.38) is convex.

The joint empirical log-likelihood function (2.38) can be written as

$$\ell(\boldsymbol{m}) = \sum_{i \in \boldsymbol{S}} \log(m_i), \tag{2.39}$$

where $m_i := m_{1i}$ if $i \in \boldsymbol{S}_1$ and $m_i := m_{2i}$ if $i \in \boldsymbol{S}_2$, that is,

$$m_i := \delta_{1i}\ m_{1i} + \delta_{2i}\ m_{2i}, \tag{2.40}$$

$\boldsymbol{S}$ is the pooled sample as defined in chapter 2.2 and δ_{ti} is the sampling indicator as defined in 2.3.

Let $\boldsymbol{\theta}_{U} = \left(\boldsymbol{\theta}_{1U}^{\top}, \boldsymbol{\theta}_{2U}^{\top}\right)^{\top}$ be the fixed unknown parameter of interest, defined as the solution to equation 2.1. Let $\boldsymbol{\varphi}_{U}$ be the known population parameter defined as the solution to equation 2.6. Let $\boldsymbol{\theta}$ be a vector in the parameter space $\boldsymbol{\Theta}$ of the parameter of interest $\boldsymbol{\theta}_{U}$. Let the $\widehat{m}_i^{\star}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{U})$ be the values which maximise the expression (2.39), for a given vector $\boldsymbol{\theta}$, subject to $m_i > 0$ and the following constraints:

1. *Unknown parameter constraints*

$$\sum_{i \in \boldsymbol{S}_1} m_{1i}\ \boldsymbol{g}_{1i}(\boldsymbol{\theta}_1) = \mathbf{0}_{\nu_1} \quad \text{and} \quad \sum_{i \in \boldsymbol{S}_2} m_{2i}\ \boldsymbol{g}_{2i}(\boldsymbol{\theta}_2) = \mathbf{0}_{\nu_2}, \tag{2.41}$$

2. *Design constraints*

$$\sum_{i \in \boldsymbol{S}_1} m_{1i}\ \pi_{1i} = n_1 \quad \text{and} \quad \sum_{i \in \boldsymbol{S}_2} m_{2i}\ \pi_{2i} = n_2, \tag{2.42}$$

3. *Known parameter constraints*, requiring that the known population parameters are reproduced, i.e., that the generalized calibration property (2.10)

holds

$$\sum_{i \in \boldsymbol{S}_1} m_{1i}\ \mathbf{f}_{1i}(\boldsymbol{\varphi}_{1U}) = \mathbf{0}_{q_1} \quad \text{and} \quad \sum_{i \in \boldsymbol{S}_2} m_{2i}\ \mathbf{f}_{2i}(\boldsymbol{\varphi}_{2U}) = \mathbf{0}_{q_2}, \tag{2.43}$$

4. *Alignment constraints*, requiring that both samples give the same point estimates for the mean of a known function $\boldsymbol{\xi}$ of the common variable $\boldsymbol{w}$

$$\sum_{i \in \boldsymbol{S}_1} m_{1i}\ \boldsymbol{\xi}(\boldsymbol{w}_i) = \sum_{i \in \boldsymbol{S}_2} m_{2i}\ \boldsymbol{\xi}(\boldsymbol{w}_i) \cdot \tag{2.44}$$

Constraint (2.44) ensures alignment of the estimates for the function of the common variable $\boldsymbol{w}$. In chapter 2.5.1, we discuss some possible choices for the alignment constraint. Constraint (2.43) is the optional benchmark constraint, which ensures that the generalised calibration property (see chapter 2.2) holds. The design constraint (2.42) plays a key role in derivation of the asymptotic properties of the maximum empirical likelihood point estimator proposed in section 2.5. Constraint (2.41) will be used to obtain point estimates and confidence intervals for the unknown parameter $\boldsymbol{\theta}_U$. This is explained in detail in chapters 2.5 and 2.9. For now we should just note that the constraint (2.41) can only be imposed for a specified value of $\boldsymbol{\theta}$, i.e., specific values of parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. When $\widehat{m}_i^{\star}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)$ are used as arguments of function (2.39), values of (2.39) depend on the value of the parameter $\boldsymbol{\theta}$ used in constraint (2.41). Using different candidate values of $\boldsymbol{\theta}$ and evaluating the resulting function (2.39) allows to find the point estimate $\widehat{\boldsymbol{\theta}}$ and the bounds of confidence intervals or regions. Note that because of constraints (2.41), (2.43) and (2.44), values $\widehat{m}_i^{\star}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)$ depend on $\boldsymbol{\varphi}_U$ and $\boldsymbol{\theta}$, as well as on the values $\boldsymbol{\xi}(\boldsymbol{w}_i)$.

The system of constraints (2.41)-(2.44) can be written as

$$\sum_{i \in \boldsymbol{S}} m_i\ \boldsymbol{c}_i^{\star}(\boldsymbol{\theta}) = \boldsymbol{C}^{\star}, \tag{2.45}$$

where m_i is defined by expression (2.40) and

$$\begin{aligned}
\boldsymbol{c}_i^{\star}(\boldsymbol{\theta}) &:= \left(\boldsymbol{c}_i^{\top} \ , \ \boldsymbol{g}_i(\boldsymbol{\theta})^{\top} \right)^{\top}, && (2.46)\\
\boldsymbol{C}^{\star} &:= \left(\boldsymbol{C}^{\top}, \mathbf{0}_{\nu}^{\top}\right)^{\top}, && (2.47)\\
\boldsymbol{c}_i &:= \left(\boldsymbol{p}_i^{\top} \ , \ \mathbf{f}_i(\boldsymbol{\varphi}_{U})^{\top} \ , \ \boldsymbol{\xi}_i^{\circ\top}\right)^{\top}, && (2.48)\\
\boldsymbol{C} &:= \left(n_1, n_2, \ \mathbf{0}_q^{\top} \ , \ \mathbf{0}_r^{\top}\right)^{\top}, && (2.49)
\end{aligned}$$

with

$$\begin{aligned}
\boldsymbol{g}_i(\boldsymbol{\theta}) &:= \left(\delta_{1i} \ \boldsymbol{g}_{1i}(\boldsymbol{\theta}_1)^{\top} \ , \ \delta_{2i} \ \boldsymbol{g}_{2i}(\boldsymbol{\theta}_2)^{\top} \right)^{\top}, && (2.50)\\
\boldsymbol{p}_i &:= \left(\delta_{1i}\pi_{1i} \ , \ \delta_{2i}\pi_{2i} \right)^{\top}, && (2.51)\\
\mathbf{f}_i(\boldsymbol{\varphi}_{U}) &:= \left(\delta_{1i}\mathbf{f}_{1i}(\boldsymbol{\varphi}_{1U})^{\top} \ , \ \delta_{2i}\mathbf{f}_{2i}(\boldsymbol{\varphi}_{2U})^{\top} \right)^{\top}, && (2.52)\\
\boldsymbol{\xi}_i^{\circ} &:= (-1)^{\delta_{2i}} \ \boldsymbol{\xi}(\boldsymbol{w}_i), && (2.53)
\end{aligned}$$

ν being the dimension of vector $\boldsymbol{g}_i(\boldsymbol{\theta})$, q being the dimension of the vector $\mathbf{f}_i(\boldsymbol{\varphi}_{U})$ and r denoting the dimension of vector $\boldsymbol{\xi}(\boldsymbol{w}_i)$.

We assume that $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}_{U}$ in constraints (2.41) and (2.43) are such that $\boldsymbol{C}^{\star}$ is an inner point of the convex hull formed by the sample observations $\{\boldsymbol{c}_i^{\star}(\boldsymbol{\theta}) : i \in \boldsymbol{S}\}$. This implies that the solution $\{\widehat{m}_i^{\star}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{U}) : i \in \boldsymbol{S}\}$ exists.

Berger and De La Riva Torres (2016) showed that, by using the method of Lagrange's multipliers, $\widehat{m}_i^{\star}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{U})$ can be derived as

$$\widehat{m}_i^{\star}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{U}) = \left\{\pi_i + \boldsymbol{\eta}^{\star\top}\boldsymbol{c}_i^{\star}(\boldsymbol{\theta})\right\}^{-1}, \tag{2.54}$$

where $\boldsymbol{\eta}^{\star}$ is a vector of Lagrange's multipliers such that constraint (2.45) is met. This result holds in the two samples context with $\pi_i := \pi_{1i}$ if $i \in \boldsymbol{S}_1$ and $\pi_i := \pi_{2i}$ if $i \in \boldsymbol{S}_2$, or equivalently

$$\pi_i = \delta_{1i}\pi_{1i} + \delta_{2i}\pi_{2i}. \tag{2.55}$$

2.5 Maximum empirical likelihood point estimator

Let $\ell(\boldsymbol{\theta}|\boldsymbol{\varphi}_{U})$ denote the maximum value of the function (2.39) for a given vector $\boldsymbol{\theta}$, under $m_i > 0$ and the constraint (2.45); that is,

$$\ell(\boldsymbol{\theta}|\boldsymbol{\varphi}_{U}) := \sum_{i\in\boldsymbol{S}} \log\{\widehat{m}_i^{\star}(\boldsymbol{\theta},\boldsymbol{\varphi}_{U})\}. \tag{2.56}$$

The maximum empirical likelihood point estimator of $\boldsymbol{\theta}_{U}$ is defined as the vector $\widehat{\boldsymbol{\theta}}$ which maximises the function (2.56); that is,

$$\widehat{\boldsymbol{\theta}} := \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \ell(\boldsymbol{\theta}|\boldsymbol{\varphi}_{U}). \tag{2.57}$$

We will call $\widehat{\boldsymbol{\theta}}$ *aligned empirical likelihood estimator*.

Berger and Kabzinska (2017) showed that $\widehat{\boldsymbol{\theta}}$ is given by the solution of a sample estimating equation

$$\widehat{\boldsymbol{G}}(\boldsymbol{\theta}) := \sum_{i\in\boldsymbol{S}} \widehat{m}_i(\boldsymbol{\varphi}_{U})\ \boldsymbol{g}_i(\boldsymbol{\theta}) = \mathbf{0}_{\nu}, \tag{2.58}$$

where $\widehat{m}_i(\boldsymbol{\varphi}_{U})$ are the values m_i that maximise function (2.39) under $m_i > 0$ and

$$\sum_{i\in\boldsymbol{S}} m_i\ \boldsymbol{c}_i = \boldsymbol{C}, \tag{2.59}$$

where $\boldsymbol{c}_i$, $\boldsymbol{C}$ and $\boldsymbol{g}_i(\boldsymbol{\theta})$ are defined by (2.48), (2.49) and (2.50) respectively. The proof is based on the observation that

$$\ell(\boldsymbol{\theta}|\boldsymbol{\varphi}_{U}) \leqslant \sum_{i\in\boldsymbol{S}} \log\{\widehat{m}_i(\boldsymbol{\varphi}_{U})\} \tag{2.60}$$

for any value of $\boldsymbol{\theta}$ such that $\boldsymbol{C}^{\star}$ is an inner point of the convex hull formed by $\{\boldsymbol{c}_i^{\star}(\boldsymbol{\theta}) : i \in \boldsymbol{S}\}$. Then, considering that when $\mathring{\boldsymbol{\theta}} = (\mathring{\boldsymbol{\theta}}_1^{\top}, \mathring{\boldsymbol{\theta}}_2^{\top})^{\top}$ is the unique solution to (2.58), $\widehat{m}_i^{\star}(\mathring{\boldsymbol{\theta}},\boldsymbol{\varphi}_{U}) = \widehat{m}_i(\boldsymbol{\varphi}_{U})$, where $\widehat{m}_i^{\star}(\mathring{\boldsymbol{\theta}},\boldsymbol{\varphi}_{U})$ is defined by (2.54), we have that

$\ell(\mathring{\boldsymbol{\theta}}|\boldsymbol{\varphi}_{U}) = \sum_{i\in\boldsymbol{S}} \log\{\widehat{m}_i(\boldsymbol{\varphi}_{U})\}$. This implies that $\mathring{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}$, where $\widehat{\boldsymbol{\theta}}$ is the maximum empirical likelihood point estimator, defined by (2.57) (Berger and Kabzinska, 2017).

We can express $\widehat{m}_i(\boldsymbol{\varphi}_{U})$ in an analogous way to $\widehat{m}_i^{\star}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{U})$:

$$\widehat{m}_i(\boldsymbol{\varphi}_{U}) = \left(\pi_i + \boldsymbol{\eta}^{\top}\boldsymbol{c}_i\right)^{-1}. \tag{2.61}$$

The values $\widehat{m}_i(\boldsymbol{\varphi}_{U})$ are the adjusted design weights produced by the proposed empirical likelihood procedure. They can be used in an analogous way as Deville and Särndal's (1992*b*) calibration weights in order to obtain point estimates for the parameters of interest. For example, $\sum_{i\in\boldsymbol{S}} \widehat{m}_i(\boldsymbol{\varphi}_{U}) y_i$ will give an estimate of the population total of variable y. The empirical likelihood adjusted weights $\widehat{m}_i(\boldsymbol{\varphi}_{U})$ possess the calibration property, i.e., solving equations (2.43) with $m_{1i} = \widehat{m}_{1i}(\boldsymbol{\varphi}_{U})$ and $m_{2i} = \widehat{m}_{2i}(\boldsymbol{\varphi}_{U})$ with respect to $\boldsymbol{\varphi}_1$ and $\boldsymbol{\varphi}_2$ gives the known values $\boldsymbol{\varphi}_1$ and $\boldsymbol{\varphi}_2$ used in constraint (2.43).

Practical aspects of calculating the adjusted weights $\widehat{m}_i(\boldsymbol{\varphi}_{U})$ and the point estimate $\widehat{\boldsymbol{\theta}}$ are discussed in chapter 5.

2.5.1 Estimation of the mean of the function of the common variable

In the previous paragraphs we treated the unknown common variable $\boldsymbol{w}$ as auxiliary information. Suppose that we want to estimate the population mean $\boldsymbol{\xi}_{U} = N^{-1}\sum_{i\in U}\boldsymbol{\xi}(\boldsymbol{w}_i)$ of the known function $\boldsymbol{\xi}$ of the common variable $\boldsymbol{w}$. The most likely application is when we simply wish to estimate the population mean of the common variable $\boldsymbol{w}$, in which case the function $\boldsymbol{\xi}(\boldsymbol{w}_i)$ is equal to $\boldsymbol{w}_i$. This is the most practically applicable formulation. However, other functions $\boldsymbol{\xi}(\boldsymbol{w}_i)$ can be used. For example, when the parameter of interest $\boldsymbol{\theta}_t$ is the variance of $y_{t;i}$, we

might use the function $\boldsymbol{\xi}(\boldsymbol{w}_i) = \boldsymbol{w}_i^2$ in an alignment constraint in order to increase correlation between $\boldsymbol{\xi}(\boldsymbol{w}_i)$ and $\boldsymbol{g}_{ti}(\boldsymbol{\theta}_t)$. The mean of this function also can be estimated.

In a general case, the population mean $\boldsymbol{\xi}_{U}$ can be defined as the solution of the population level estimating equation:

$$\sum_{i \in U} \boldsymbol{h}_i(\boldsymbol{w}_i, \boldsymbol{\xi}) = \boldsymbol{0}, \tag{2.62}$$

where $\boldsymbol{h}_i(\boldsymbol{w}_i, \boldsymbol{\xi}) := \boldsymbol{\xi}(\boldsymbol{w}_i) - N n_t^{-1} \pi_{ti}\ \boldsymbol{\xi}$, for $t = 1$ and 2. We will use $\boldsymbol{h}_i(\boldsymbol{\xi})$ to denote $\boldsymbol{h}_i(\boldsymbol{w}_i, \boldsymbol{\xi})$ for simplicity henceforth.

The maximum empirical likelihood point estimator $\widehat{\boldsymbol{\xi}}$ of $\boldsymbol{\xi}_{U}$ is obtained as the value which maximises

$$\ell(\boldsymbol{\xi}|\boldsymbol{\varphi}_{U}) := \sum_{i \in \boldsymbol{S}} \log\{\widehat{m}_i^{\star}(\boldsymbol{\xi}|\boldsymbol{\varphi}_{U})\}, \tag{2.63}$$

where $\widehat{m}_i^{\star}(\boldsymbol{\xi}|\boldsymbol{\varphi}_{U})$ are the values which maximise (2.63) for a given value $\boldsymbol{\xi}$, under $m_i > 0$ and

$$\sum_{i \in \boldsymbol{S}} m_i\ \boldsymbol{c}_{\boldsymbol{i}}^{**} = \boldsymbol{C}^{**}, \tag{2.64}$$

with

$$\boldsymbol{c}_{\boldsymbol{i}}^{**} = \left(\boldsymbol{c_i}^{\top}, \boldsymbol{h}_i(\boldsymbol{\xi})^{\top}\right)^{\top}, \tag{2.65}$$

$$\boldsymbol{C}^{**} = \left(\boldsymbol{C}^{\top}, \boldsymbol{0}_{2r}^{\top}\right)^{\top}, \tag{2.66}$$

where

$$\boldsymbol{h}_i(\boldsymbol{\xi}) \ := \ \left(\ \delta_{1i}\ \boldsymbol{h}_{1i}(\boldsymbol{\xi})^{\top}\ ,\ \delta_{2i}\ \boldsymbol{h}_{2i}(\boldsymbol{\xi})^{\top}\ \right)^{\top}. \tag{2.67}$$

Note that this is equivalent to including $\boldsymbol{\xi}$ within $\boldsymbol{\theta}$.

Based on an argument similar to the one presented in chapter 2.5, finding the value $\widehat{\boldsymbol{\xi}}$ which maximises (2.63) is equivalent to solving the following estimating equation for $\boldsymbol{\xi}$:

$$\sum_{i \in \boldsymbol{S}} \widehat{m}_i(\boldsymbol{\varphi}_U)\ \boldsymbol{h}_i(\boldsymbol{\xi}) = \mathbf{0}_{2r}. \tag{2.68}$$

Equation (2.68) can also be written as

$$\sum_{i \in \boldsymbol{S}_1} \widehat{m}_{1i}(\boldsymbol{\varphi}_U)\ \boldsymbol{h}_{1i}(\boldsymbol{\xi}) = \mathbf{0}_r \quad \text{and} \quad \sum_{i \in \boldsymbol{S}_2} \widehat{m}_{2i}(\boldsymbol{\varphi}_U)\ \boldsymbol{h}_{2i}(\boldsymbol{\xi}) = \mathbf{0}_r. \tag{2.69}$$

The solutions of the equations (2.69) are $\widehat{\boldsymbol{\xi}}_1$ and $\widehat{\boldsymbol{\xi}}_2$, where $\widehat{\boldsymbol{\xi}}_t = N^{-1} \sum_{i \in \boldsymbol{S}_t} \widehat{m}_{ti}(\boldsymbol{\varphi}_U)\, \boldsymbol{\xi}(\boldsymbol{w}_i)$, $t = 1, 2$. Constraint (2.44), which is imposed on the adjusted weights $\widehat{m}_{1i}(\boldsymbol{\varphi}_U)$ and $\widehat{m}_{2i}(\boldsymbol{\varphi}_U)$, implies that both equations in (2.69) give the same estimate, that is, $\widehat{\boldsymbol{\xi}}_1 = \widehat{\boldsymbol{\xi}}_2$.

Note that various functions of the variable $\boldsymbol{w}_i$ can be used to define the alignment constraint (2.44). In particular, these functions can be chosen to maximise the correlation between the $\boldsymbol{\xi}(\boldsymbol{w}_i)$ and $\boldsymbol{g}_i(\boldsymbol{\theta})$ (see chapter 2.6.2 for a discussion).

2.6 Asymptotic properties of the maximum empirical likelihood point estimator

In this chapter the asymptotic properties of the aligned empirical likelihood estimator (2.58) are established. We start by specifying the assumed regularity conditions. We then derive the generalized regression type estimator asymptotically equivalent to the aligned empirical likelihood estimator and discuss its properties. We also show that the aligned empirical likelihood estimator is asymptotically $\sqrt{n}$ design-consistent.

2.6.1 Regularity conditions

Conditions on the sampling design

Suppose that the the sampling design is such that the following regularity conditions hold for $t = 1, 2$:

$$\max_{i \in \boldsymbol{S}_t} \left\{ \frac{N}{n_t} \pi_{ti} \right\} = O_{\mathcal{P}}(1) \quad \text{and} \quad \max_{i \in \boldsymbol{S}_t} \left\{ \frac{n_t}{N} \pi_{ti}^{-1} \right\} = O_{\mathcal{P}}(1) \tag{2.70}$$

$$N^{-1} \| \widehat{\boldsymbol{C}}_\pi - \boldsymbol{C} \| = \boldsymbol{\mathcal{O}}_{\mathcal{P}}(n^{-1/2}), \tag{2.71}$$

$$\max\{ \| \boldsymbol{c}_i \| : i \in \boldsymbol{S} \} = o_{\mathcal{P}}(n^{1/2}), \tag{2.72}$$

$$\| \widehat{\boldsymbol{S}} \| = \boldsymbol{\mathcal{O}}_{\mathcal{P}}(1), \tag{2.73}$$

$$\| \widehat{\boldsymbol{S}}^{-1} \| = \boldsymbol{\mathcal{O}}_{\mathcal{P}}(1), \tag{2.74}$$

$$\frac{n^{\tau-1}}{N^\tau} \sum_{i \in \boldsymbol{S}} \frac{\| \boldsymbol{c}_i \|^\tau}{\pi_i^\tau} = O_{\mathcal{P}}(1) \quad (\tau = 2,\ 3,\ 4), \tag{2.75}$$

with $n = n_1 + n_2$ and

$$\widehat{\boldsymbol{S}} := -\frac{n}{N^2} \sum_{i \in \boldsymbol{S}} \frac{\boldsymbol{c}_i \boldsymbol{c}_i^\top}{\pi_i^2}, \tag{2.76}$$

$$\widehat{\boldsymbol{C}}_\pi := \sum_{i \in \boldsymbol{S}} \frac{\boldsymbol{c}_i}{\pi_i}, \tag{2.77}$$

where $\boldsymbol{c}_i$, $\boldsymbol{C}$ and π_i are respectively defined by (2.48), (2.49) and (2.55). The orders $\boldsymbol{\mathcal{O}}_{\mathcal{P}}(\cdot)$ and $\boldsymbol{o}_{\mathcal{P}}(\cdot)$ denote matrices which are such that $\| \boldsymbol{\mathcal{O}}_{\mathcal{P}}(a) \| = O_{\mathcal{P}}(a)$ and $\| \boldsymbol{o}_{\mathcal{P}}(a) \| = o_{\mathcal{P}}(a)$, where $\| \boldsymbol{A} \|$ is the Euclidean norm, i.e., $\| \boldsymbol{A} \| := \text{trace}(\boldsymbol{A}^\top \boldsymbol{A})^{1/2}$.

A thorough discussion of conditions (2.70)-(2.75) can be found in Berger and De La Riva Torres (2016). Condition (2.70) is the key condition which ensures that the π_{ti} are of the same order as n_t/N (Krewski and Rao, 1981). Condition (2.71) is a standard law of large numbers condition (e.g. Isaki and Fuller, 1982; Krewski

and Rao, 1981). Condition (2.72) holds for common unequal probability sampling designs (Chen and Sitter, 1999, Appendix 2). Conditions (2.73) and (2.74) hold when $-\widehat{\boldsymbol{S}}$ is positive definite and when there exists a positive definite matrix $-\boldsymbol{S}$ such that $\widehat{\boldsymbol{S}} - \boldsymbol{S} = \boldsymbol{o}_{\mathcal{P}}(1)$ and $\|\boldsymbol{S}\| = O(1)$ (Berger, 2015, Lemma B.4). Condition (2.75) is a Lyapunov-type condition for the existence of moments (e.g. Krewski and Rao, 1981, p. 1014, Deville and Särndal, 1992*a*, p. 381). For conditions (2.71)-(2.75) to hold, we assume that $\boldsymbol{\xi}(\boldsymbol{w}_i) = \boldsymbol{\mathcal{O}}_{\mathcal{P}}(1)$ for all $i \in \boldsymbol{S}$, which is achieved when the components of $\boldsymbol{\xi}(\boldsymbol{w}_i)$ are bounded (Berger and De La Riva Torres, 2016). For condition (2.72) to hold, the components of $\boldsymbol{c}_i$ have to be bounded in probability. This can be justified by substituting constraint (2.42) by the following (Berger and De La Riva Torres, 2016):

$$\sum_{i \in \boldsymbol{S}_1} m_{1i}\ N n_1^{-1} \pi_{1i} = N n_1^{-1} n_1 \quad \text{and} \quad \sum_{i \in \boldsymbol{S}_2} m_{2i}\ N n_2^{-1} \pi_{2i} = N n_2^{-1} n_2. \tag{2.78}$$

Note that this can be done without a loss of generality and does not have any implications for practical applications, as the quantity $N n_t^{-1}$ appears at both sides of the equation. In particular, the population size N does not have to be known.

Conditions on the parameter of interest

Suppose also that $\boldsymbol{\theta}_U$ is such that the following conditions hold:

$$\widehat{\boldsymbol{G}}_{\pi}(\boldsymbol{\theta}_U) = \boldsymbol{\mathcal{O}}_{\mathcal{P}}(N n^{-1/2}), \tag{2.79}$$

$$\frac{n^{\tau-1}}{N^{\tau}} \sum_{i \in \boldsymbol{S}} \frac{\|\boldsymbol{g}_i(\boldsymbol{\theta}_U)\|^{\tau}}{\pi_i^{\tau}} = O_{\mathcal{P}}(1) \quad (\tau = 2,\ 3,\ 4), \tag{2.80}$$

$$\widehat{\boldsymbol{\nabla}}(\boldsymbol{\theta}) := \frac{1}{N} \frac{\partial \widehat{\boldsymbol{G}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \quad \text{is continuous in } \boldsymbol{\theta} \in \boldsymbol{\Theta}_U, \tag{2.81}$$

$$\frac{1}{N} \| \frac{\partial \widehat{\boldsymbol{\nabla}}(\boldsymbol{\theta})_k}{\partial \boldsymbol{\theta}} \| = \boldsymbol{\mathcal{O}}_{\mathcal{P}}(1) \quad \text{uniformly for all } \boldsymbol{\theta} \in \boldsymbol{\Theta}_U, \tag{2.82}$$

$$\|\widehat{\boldsymbol{\nabla}}(\boldsymbol{\theta}_U)\| \asymp_p 1, \tag{2.83}$$

$$|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_U| = o_{\mathcal{P}}(1), \tag{2.84}$$

where $\widehat{\boldsymbol{\nabla}}(\boldsymbol{\theta})_k$ is the $k-th$ row of matrix $\widehat{\boldsymbol{\nabla}}(\boldsymbol{\theta})$, $k-1,2,...,K$; K is the number of rows in matrix $\widehat{\boldsymbol{\nabla}}(\boldsymbol{\theta})$,

$$\widehat{\boldsymbol{G}}_{\pi}(\boldsymbol{\theta}_U) \;:=\; \sum_{i\in\boldsymbol{S}} \frac{1}{\pi_i}\boldsymbol{g}_i(\boldsymbol{\theta}_U) \tag{2.85}$$

and $\boldsymbol{\Theta}_U$ is a compact neighbourhood containing $\boldsymbol{\theta}_U$. Similar conditions can be found in (Berger and De La Riva Torres, 2016).

Condition (2.79) is a law of large numbers condition, because $\widehat{\boldsymbol{G}}_{\pi}(\boldsymbol{\theta}_U)$ is the unbiased Horvitz and Thompson (1952) estimator of $\boldsymbol{G}(\boldsymbol{\theta}_U) = \mathbf{0}_{\nu}$. Conditions under which condition (2.79) holds can be found in Isaki and Fuller (1982). Condition (2.80) is a Lyapunov-type condition for the existence of moments of $\boldsymbol{g}_i(\boldsymbol{\theta}_U)$. Conditions (2.81), (2.82) and (2.83) ensure that Taylor series expansion of $\widehat{\boldsymbol{G}}(\boldsymbol{\theta})$ exists (Berger and De La Riva Torres, 2016). Condition (2.83) means that the derivative $\widehat{\boldsymbol{\nabla}}(\boldsymbol{\theta}_U)$ is finite and that $\widehat{\boldsymbol{G}}(\boldsymbol{\theta})$ is not flat in the neighbourhood of $\boldsymbol{\theta}_U$ (Berger and De La Riva Torres, 2016). Condition (2.84) ensures consistency of $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}_U$. This condition can be justified through a reasoning similar to that presented by Qin and Lawless (1994, Lemma 1). An analogous assumption is made e.g. by Godambe and Thompson (2009). Note that constraints (2.79), (2.80) and (2.83) need to hold for $\boldsymbol{\theta}_U$ only, that is, when $\boldsymbol{\theta}$ is equal to the true population value $\boldsymbol{\theta}_U$.

2.6.2 Asymptotic equivalence of the maximum empirical likelihood point estimator to a generalized regression type estimator

Let

$$\widehat{\boldsymbol{G}}_{\pi}(\boldsymbol{\theta}) = \sum_{i\in S} \frac{\boldsymbol{g}_i(\boldsymbol{\theta})}{\pi_i} \tag{2.86}$$

be Horvitz and Thompson's (1952) estimator of $\boldsymbol{\theta}_U$.

The following theorem establishes asymptotic equivalence between the proposed empirical likelihood estimator $\widehat{\boldsymbol{\theta}}$, defined as the solution of (2.58), and a generalized regression type estimator.

Theorem 1. *Under conditions* (2.70)-(2.75)*, for all* $\boldsymbol{\theta}$ *which are such that:*

$$\frac{1}{nN^2}\sum_{i\in\boldsymbol{S}}\frac{||\boldsymbol{g}_i(\boldsymbol{\theta})||^2}{\pi_i^2} = O_{\mathcal{P}}(n^{-2}), \tag{2.87}$$

we have that

$$\widehat{\boldsymbol{G}}(\boldsymbol{\theta}) = \widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}) + \boldsymbol{o}_{\mathcal{P}}(Nn^{-1/2}), \tag{2.88}$$

where $\widehat{\boldsymbol{G}}(\boldsymbol{\theta})$ *is defined by* (2.58) *and*

$$\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}) = \widehat{\boldsymbol{G}}_\pi(\boldsymbol{\theta}) + \widehat{\boldsymbol{B}}(\boldsymbol{\theta},\boldsymbol{\varphi}_U)^\top(\boldsymbol{C} - \widehat{\boldsymbol{C}}_\pi), \tag{2.89}$$

$$\widehat{\boldsymbol{B}}(\boldsymbol{\theta},\boldsymbol{\varphi}_U) := \Big(\sum_{i\in\boldsymbol{S}}\frac{1}{\pi_i^2}\boldsymbol{c}_i\boldsymbol{c}_i^\top\Big)^{-1}\sum_{i\in\boldsymbol{S}}\frac{1}{\pi_i^2}\boldsymbol{c}_i\boldsymbol{g}_i(\boldsymbol{\theta})^\top, \tag{2.90}$$

with $\widehat{\boldsymbol{C}}_\pi$ *defined by* (2.77) *and* $\widehat{\boldsymbol{G}}_\pi(\boldsymbol{\theta})$ *defined by* (2.86)*.*

The proof can be found in the Appendix. The first step in the proof is showing that the Lagrange multipliers $\boldsymbol{\eta}^\top$ in (2.61) are bounded by $O_{\mathcal{P}}(n^{1/2}N^{-1})$. Then the estimating equation (2.58) is expressed in terms of a regression type estimator plus an error term, for which an asymptotic order is established. Theorem 1 holds for all $\boldsymbol{\theta}$ which satisfy (2.87), that is, not only when $\boldsymbol{\theta}$ equals $\boldsymbol{\theta}_U$.

The estimator (2.89) can also be written as:

$$\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}) = \{\widehat{\boldsymbol{G}}_{r1}(\boldsymbol{\theta}_1)^\top, \widehat{\boldsymbol{G}}_{r2}(\boldsymbol{\theta}_2)^\top\}^\top, \tag{2.91}$$

where

$$\widehat{\boldsymbol{G}}_{r1}(\boldsymbol{\theta}_1) \;:=\; \widehat{\boldsymbol{G}}_{1\pi}(\boldsymbol{\theta}_1) - \widehat{\boldsymbol{B}}_{1\mathrm{f}1}(\boldsymbol{\theta},\boldsymbol{\varphi}_U)^\top\,\widehat{\mathbf{f}}_{1\pi}(\boldsymbol{\varphi}_1) - \widehat{\boldsymbol{B}}_{1\mathrm{f}2}(\boldsymbol{\theta},\boldsymbol{\varphi}_U)^\top\,\widehat{\mathbf{f}}_{2\pi}(\boldsymbol{\varphi}_2)$$

$$+\widehat{\boldsymbol{B}}_{1\boldsymbol{\xi}}(\boldsymbol{\theta},\boldsymbol{\varphi}_U)^\top\big(\widehat{\boldsymbol{\xi}}_{2\pi}-\widehat{\boldsymbol{\xi}}_{1\pi}\big), \tag{2.92}$$

$$\widehat{\boldsymbol{G}}_{r2}(\boldsymbol{\theta}_2) := \widehat{\boldsymbol{G}}_{2\pi}(\boldsymbol{\theta}_2)-\widehat{\boldsymbol{B}}_{2\mathbf{f}1}(\boldsymbol{\theta},\boldsymbol{\varphi}_U)^\top\,\widehat{\mathbf{f}}_{1\pi}(\boldsymbol{\varphi}_1)-\widehat{\boldsymbol{B}}_{2\mathbf{f}2}(\boldsymbol{\theta},\boldsymbol{\varphi}_U)^\top\,\widehat{\mathbf{f}}_{2\pi}(\boldsymbol{\varphi}_2)$$

$$+\widehat{\boldsymbol{B}}_{2\boldsymbol{\xi}}(\boldsymbol{\theta},\boldsymbol{\varphi}_U)^\top\big(\widehat{\boldsymbol{\xi}}_{1\pi}-\widehat{\boldsymbol{\xi}}_{2\pi}\big) \tag{2.93}$$

and

$$\widehat{\boldsymbol{G}}_{t\pi}(\boldsymbol{\theta}_t)=\sum_{i\in\boldsymbol{S}_t}\pi_{ti}^{-1}\boldsymbol{g}_{ti}(\boldsymbol{\theta}_{tU}), \tag{2.94}$$

$$\widehat{\mathbf{f}}_{t\pi}=\sum_{i\in\boldsymbol{S}_t}\pi_{ti}^{-1}\mathbf{f}_{ti}(\boldsymbol{\varphi}_{tU}), \tag{2.95}$$

$$\widehat{\boldsymbol{\xi}}_{t\pi}=\sum_{i\in\boldsymbol{S}_t}\pi_{ti}^{-1}\boldsymbol{\xi}(\boldsymbol{w}_i)\cdot \tag{2.96}$$

The terms $\widehat{\boldsymbol{B}}_{1\mathbf{f}1}(\boldsymbol{\theta},\boldsymbol{\varphi}_U)^\top\,\widehat{\mathbf{f}}_{1\pi}(\boldsymbol{\varphi}_1)$, $\widehat{\boldsymbol{B}}_{1\mathbf{f}2}(\boldsymbol{\theta},\boldsymbol{\varphi}_U)^\top\,\widehat{\mathbf{f}}_{2\pi}(\boldsymbol{\varphi}_2)$, $\widehat{\boldsymbol{B}}_{2\mathbf{f}1}(\boldsymbol{\theta},\boldsymbol{\varphi}_U)^\top\,\widehat{\mathbf{f}}_{1\pi}(\boldsymbol{\varphi}_1)$ and $\widehat{\boldsymbol{B}}_{2\mathbf{f}2}(\boldsymbol{\theta},\boldsymbol{\varphi}_U)^\top\,\widehat{\mathbf{f}}_{2\pi}(\boldsymbol{\varphi}_2)$ are regression terms based on the known population parameters $\boldsymbol{\varphi}_1$ and $\boldsymbol{\varphi}_2$. The terms $\widehat{\boldsymbol{B}}_{1\boldsymbol{\xi}}(\boldsymbol{\theta},\boldsymbol{\varphi}_U)^\top\big(\widehat{\boldsymbol{\xi}}_{2\pi}-\widehat{\boldsymbol{\xi}}_{1\pi}\big)$ and $\widehat{\boldsymbol{B}}_{2\boldsymbol{\xi}}(\boldsymbol{\theta},\boldsymbol{\varphi}_U)^\top\big(\widehat{\boldsymbol{\xi}}_{1\pi}-\widehat{\boldsymbol{\xi}}_{2\pi}\big)$ are the extended regression terms, where the estimates $\widehat{\boldsymbol{\xi}}_{1\pi}$ and $\widehat{\boldsymbol{\xi}}_{2\pi}$ are used in place of the known population parameters.

Berger and Kabzinska (2017) proved that if $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are independent, a design-consistent estimator of the variance-covariance matrix of (2.89), under the stratified maximum entropy sampling design (Hájek, 1981, Ch. 14), is given by:

$$\widehat{\boldsymbol{V}}_{\mathcal{P}}\big\{\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta})\big\}=\sum_{i\in\boldsymbol{S}}\frac{1}{\pi_i^2}\widetilde{\boldsymbol{g}}_i(\boldsymbol{\theta})\widetilde{\boldsymbol{g}}_i(\boldsymbol{\theta})^\top, \tag{2.97}$$

where

$$\widetilde{\boldsymbol{g}}_i(\boldsymbol{\theta})=\boldsymbol{g}_i(\boldsymbol{\theta})-\widehat{\boldsymbol{B}}(\boldsymbol{\theta},\boldsymbol{\varphi}_U)^\top\boldsymbol{c}_i\cdot \tag{2.98}$$

Following an argument presented by Berger (2011) and Berger and Kabzinska (2017), this result holds for high entropy sampling designs, such as Rao's (1965)

& Sampford's (1967) design or the randomised systematic design.

Berger et al. (2003) showed that in high entropy sampling designs, if the design constraints such as (2.42) are included as the first components of $\boldsymbol{c}_i$ and $\boldsymbol{C}$, the regression estimators (2.92) and (2.93) are asymptotically equal to the Montanari's (1987) optimal regression estimators.

Based on the regression estimator theory, we can expect a reduction in the variance of (2.89) when $\mathbf{f}_i(\boldsymbol{\varphi})$ are highly correlated with $\boldsymbol{g}_i(\boldsymbol{\theta})$. The effect of terms $\widehat{\boldsymbol{B}}_{1\boldsymbol{\xi}}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)^\top \big(\widehat{\boldsymbol{\xi}}_{2\pi} - \widehat{\boldsymbol{\xi}}_{1\pi}\big)$ and $\widehat{\boldsymbol{B}}_{2\boldsymbol{\xi}}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)^\top \big(\widehat{\boldsymbol{\xi}}_{1\pi} - \widehat{\boldsymbol{\xi}}_{2\pi}\big)$ on the variance (2.97) is twofold. First, there is an increase in variance due to the fact that the parameters $\widehat{\boldsymbol{\xi}}_{1\pi}$ and $\widehat{\boldsymbol{\xi}}_{2\pi}$ are estimated. However, there is also a decrease in variance if there is a high correlation between $\boldsymbol{\xi}(\boldsymbol{w}_i)$ and $\boldsymbol{g}_i(\boldsymbol{\theta})$. When the decrease in variance is larger than the increase, the overall effect of alignment on precision of (2.89) is positive. The advantage of the proposed approach is that the function $\boldsymbol{\xi}(\boldsymbol{w}_i)$ can be chosen to improve this correlation. For example, suppose that a variable w_i is correlated with a variable y_i. When $\boldsymbol{g}_i(\boldsymbol{\theta})$ is the estimating function for an α-quantile of the distribution of y_i, it is recommended to use $\boldsymbol{\xi}(\boldsymbol{w}_i) = \delta(w_i \leqslant \alpha)$, where $\delta(\cdot)$ is an indicator function equal to 1 if its argument is true and to 0 otherwise. If $\boldsymbol{g}_i(\boldsymbol{\theta})$ is the estimating function for the variance, $\boldsymbol{\xi}(\boldsymbol{w}_i) = (w_i, w_i^2)^\top$ should be used.

2.6.3 Asymptotic design-consistency of the maximum empirical likelihood point estimator

Consider a sequence of nested populations $U^{(\nu)}$ of size $N^{(\nu)}$, where $\nu = 1, 2, \ldots, \infty$ (Isaki and Fuller, 1982). Consider a sequence of samples $\boldsymbol{S}_t^{(\nu)}$ of size $n_t^{(\nu)} < N^{(\nu)}$ selected from $U^{(\nu)}$ according to a sampling design $\mathcal{P}_t^{(\nu)}(\boldsymbol{S}_t)$. We assume that $n_1^{(\nu)} \to \infty$ and $n_2^{(\nu)} \to \infty$, as $\nu \to \infty$. We also assume that $n_1^{(\nu)}/N \to 0$ and $n_2^{(\nu)}/N \to 0$, i.e., we assume that the sampling fraction is negligible. Extension to

non-negligible sampling fractions is discussed in chapter 2.11. Let $o_{\mathcal{P}}(\cdot)$ and $O_{\mathcal{P}}(\cdot)$ be the orders of convergence in probability with respect to the sampling design $\mathcal{P}_t^{(\nu)}(\boldsymbol{S}_t)$ (e.g. Isaki and Fuller, 1982), as $\nu \to \infty$. To simplify the notation, we drop the index ν in the following text.

The following theorem shows that the maximum empirical likelihood point estimator $\widehat{\boldsymbol{\theta}}$ is asymptotically $\sqrt{n}$ design-consistent.

Theorem 2. *Let $n := n_1 + n_2$. Under the regularity conditions* (2.70) (2.70)-(2.75), (2.79), (2.80) *(with $\tau = 2$),* (2.81)-(2.84)*, we have that $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_U\| = \boldsymbol{\mathcal{O}}_{\mathcal{P}}(n^{-1/2})$.*

The $\sqrt{n}$ design-consistency is achieved because of the design constraint (2.42). The proof can be found in the appendix. The proof shows that $\|\widehat{\boldsymbol{B}}(\boldsymbol{\theta}_U, \boldsymbol{\varphi}_U)\| \leqslant O_{\mathcal{P}}(1)$ and $N^{-1}\widehat{\boldsymbol{G}}(\boldsymbol{\theta}_U) = \boldsymbol{\mathcal{O}}_{\mathcal{P}}(n^{-1/2})$, where $\widehat{\boldsymbol{G}}(\boldsymbol{\theta})$ and $\widehat{\boldsymbol{B}}(\boldsymbol{\theta}_U, \boldsymbol{\varphi}_U)$ are respectively defined by (2.58) and (2.90). The asymptotic $\sqrt{n}$ design-consistency for $\widehat{\boldsymbol{\theta}}$ is then based on taking a Taylor expansion of $\widehat{\boldsymbol{G}}(\widehat{\boldsymbol{\theta}})$ in the neighbourhood of $\boldsymbol{\theta}_U$. Theorem 2 holds whether or not the common parameter $\boldsymbol{\xi}_U$ is included within $\boldsymbol{\theta}_U$ and whether or not $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are independent. Theorem 2 is an improved result compared to condition (2.84) in that a rate of convergence is established.

2.7 Effect of a difference in sample sizes on the maximum empirical likelihood point estimator

In practical applications samples $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ might have different sizes and utilise different designs. Following (2.92) and (2.93), the alignment constraint can be intuitively interpreted as calibration on a zero function defined by the difference between the estimates of the common parameter obtained from the two samples. Efficient ways of introducing an alignment constraint when samples considerably differ in size have been investigated by both Renssen and Nieuwenbroek (1997) and Merkouris (2004, 2010*a*, 2015). Renssen and Nieuwenbroek (1997) accounted for

differences in sample sizes by introducing a weighting coefficient which depends on relative variances. In their method, the total of the common variable is estimated by a weighted average of two separate sample estimates. This total is then used in the regression estimators of the totals of the variables of interest. In the composite regression estimator, the entries in the weighting matrix included in the regression coefficient can be adjusted by relative sample sizes and design effects (Merkouris, 2004, 2010*a*, 2015). The aligned empirical likelihood estimator does not include any adjustment factors. However, because of the design constraint (2.42), an implicit adjustment for the relative sample size is made. This can be seen in the coefficient $\widehat{\boldsymbol{B}}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)$ of the asymptotically equivalent generalized regression estimator (2.89). We discuss this below.

Consider a simple situation when there is no stratification, no benchmark constraints and there is a single common variable w and two equal scalar parameters of interest $\theta_1 = \theta_2$. Suppose that $n_1 \gg n_2$ and that $g_{1i}(\theta_1)$ and $g_{2i}(\theta_2)$ are the same estimating functions. In such a case, we would like the adjustment applied to $\widehat{G}_{1\pi}(\theta_1)$ to be smaller than the adjustment applied to $\widehat{G}_{2\pi}(\theta_2)$. Below we show that this is indeed the case.

We can express the coefficient $\widehat{\boldsymbol{B}}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)$ in equation (2.88) as

$$\widehat{\boldsymbol{B}}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U) = \{\boldsymbol{\Sigma}^{\circ}_{cc}\}^{-1} \boldsymbol{\Sigma}^{\circ}_{cg}. \tag{2.99}$$

It can be shown that

$$\boldsymbol{\Sigma}^{\circ}_{cg} = \begin{bmatrix} \mathcal{G}_1 & 0 \\ 0 & \mathcal{G}_2 \\ \mathcal{H}_1 & \mathcal{H}_2 \end{bmatrix}, \tag{2.100}$$

with

$$\mathcal{G}_1 = \sum_{i \in \boldsymbol{S}_1} \pi_i^{-1} g_{1i}(\theta_1), \tag{2.101}$$

$$\mathcal{G}_2 = \sum_{i \in \boldsymbol{S}_2} \pi_i^{-1} g_{2i}(\theta_2), \tag{2.102}$$

$$\mathcal{H}_1 = \sum_{i \in \boldsymbol{S}_1} \pi_i^{-2} g_{1i}(\theta_1) \xi_i^{\circ}, \tag{2.103}$$

$$\mathcal{H}_2 = \sum_{i \in \boldsymbol{S}_2} \pi_i^{-2} g_{2i}(\theta_2) \xi_i^{\circ} \tag{2.104}$$

and

$$\boldsymbol{\Sigma}_{cc}^{\circ} = \begin{bmatrix} n_1 & 0 & \mathcal{S}_{H1} \\ 0 & n_2 & \mathcal{S}_{H2} \\ \mathcal{S}_{H1} & \mathcal{S}_{H2} & \mathcal{S}_{H12} \end{bmatrix}, \tag{2.105}$$

with $\mathcal{S}_{H1} = \sum_{i \in \boldsymbol{S}_1} \pi_i^{-1} \xi_i^{\circ}$, $\mathcal{S}_{H2} = \sum_{i \in \boldsymbol{S}_2} \pi_i^{-1} \xi_i^{\circ}$, $\mathcal{S}_{H12} = \sum_{i \in \boldsymbol{S}} \pi_i^{-2} \xi_i^{\circ 2}$ and ξ_i° defined by 2.53.

Hence, (2.99) becomes

$$\widehat{\boldsymbol{B}}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{U}) = \{det(\boldsymbol{\Sigma}_{cc}^{\circ})\}^{-1} \widehat{\boldsymbol{\mathcal{B}}}, \tag{2.106}$$

with

$$\widehat{\boldsymbol{\mathcal{B}}} = \begin{bmatrix} \widehat{\mathcal{B}}_{1,1} & \widehat{\mathcal{B}}_{1,2} \\ \widehat{\mathcal{B}}_{2,1} & \widehat{\mathcal{B}}_{2,2} \\ \widehat{\mathcal{B}}_{3,1} & \widehat{\mathcal{B}}_{3,2} \end{bmatrix}, \tag{2.107}$$

$$\widehat{\mathcal{B}}_{1,1} = (n_2 \mathcal{S}_{H12} - \mathcal{S}_{H1} \mathcal{S}_{H2}) \mathcal{G}_1 - n_2 \mathcal{S}_{H1} \mathcal{H}_1, \tag{2.108}$$

$$\widehat{\mathcal{B}}_{2,1} = \mathcal{S}_{H1} \mathcal{S}_{H2} \mathcal{G}_1 - n_1 \mathcal{S}_{H2} \mathcal{H}_1, \tag{2.109}$$

$$\widehat{\mathcal{B}}_{3,1} = n_1 n_2 \mathcal{H}_1 - n_2 \mathcal{S}_{H1} \mathcal{G}_1, \tag{2.110}$$

$$\widehat{\mathcal{B}}_{1,2} = \mathcal{S}_{H1} \mathcal{S}_{H2} \mathcal{G}_2 - n_2 \mathcal{S}_{H1} \mathcal{H}_2, \tag{2.111}$$

$$\widehat{\mathcal{B}}_{2,2} = (n_1\mathcal{S}_{H12} - \mathcal{S}_{H1}\mathcal{S}_{H2})\mathcal{G}_2 - n_1\mathcal{S}_{H2}\mathcal{H}_2, \tag{2.112}$$

$$\widehat{\mathcal{B}}_{3,2} = n_1 n_2 \mathcal{H}_2 - n_1\mathcal{S}_{H2}\mathcal{G}_2. \tag{2.113}$$

In the special case considered, the vector $(\boldsymbol{C} - \widehat{\boldsymbol{C}}_\pi)$ in (2.89) has only three elements. The first two elements are equal to zero and the third element equals $(\widehat{\xi}_{1\pi} - \widehat{\xi}_{2\pi})$. Therefore, (2.92) and (2.93) simplify to

$$\widehat{G}_{r1}(\theta_1) = \widehat{G}_{1\pi}(\theta_1) - \hat{\mathcal{B}}_{1\xi}\big(\widehat{\xi}_{2\pi} - \widehat{\xi}_{1\pi}\big), \tag{2.114}$$

$$\widehat{G}_{r2}(\theta_2) = \widehat{G}_{2\pi}(\theta_2) - \hat{\mathcal{B}}_{2\xi}\big(\widehat{\xi}_{1\pi} - \widehat{\xi}_{2\pi}\big), \tag{2.115}$$

with

$$\begin{aligned}\hat{\mathcal{B}}_{1\xi} &= \{det(\boldsymbol{\Sigma}^\circ_{cc})\}^{-1}\widehat{\mathcal{B}}_{3,1} \qquad (2.116)\\ &= n_2\{det(\boldsymbol{\Sigma}^\circ_{cc})\}^{-1}\\ &\quad\times\left\{n_1\sum_{i\in\boldsymbol{S}_1}\pi_i^{-2}\xi(w_i)g_{1i}(\theta_1) - \sum_{i\in\boldsymbol{S}_1}\pi_i^{-1}\xi(w_i)\sum_{i\in\boldsymbol{S}_1}\pi_i^{-1}g_{1i}(\theta_1)\right\},\end{aligned}$$

$$\begin{aligned}\hat{\mathcal{B}}_{2\xi} &= \{det(\boldsymbol{\Sigma}^\circ_{cc})\}^{-1}\widehat{\mathcal{B}}_{3,2} \qquad (2.117)\\ &= n_1\{det(\boldsymbol{\Sigma}^\circ_{cc})\}^{-1}\\ &\quad\times\left\{n_2\sum_{i\in\boldsymbol{S}_2}\pi_i^{-2}\xi(w_i)g_{2i}(\theta_2) - \sum_{i\in\boldsymbol{S}_2}\pi_i^{-1}\xi(w_i)\sum_{i\in\boldsymbol{S}_2}\pi_i^{-1}g_{2i}(\theta_2)\right\}.\end{aligned}$$

Consider a situation when units within $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are selected with equal probabilities $\pi_{ti} = n_t/N$. Substituting π_{ti} by n_t/N in (2.116) and (2.117) gives

$$\hat{\mathcal{B}}_{1\xi} = n_2N^2\{det(\boldsymbol{\Sigma}^\circ_{cc})\}^{-1}Cov_1\left\{\xi(w_i), g_{1i}(\theta_1)\right\}, \tag{2.118}$$

$$\hat{\mathcal{B}}_{2\xi} = n_1N^2\{det(\boldsymbol{\Sigma}^\circ_{cc})\}^{-1}Cov_2\left\{\xi(w_i), g_{2i}(\theta_2)\right\}, \tag{2.119}$$

where

$$Cov_t\left\{\xi(w_i), g_{ti}(\theta_t)\right\}\left\{n_t^{-1}\sum_{i\in\boldsymbol{S}_t}\xi(w_i)g_{ti}(\theta_t) - n_t^{-1}\sum_{i\in\boldsymbol{S}_t}\xi(w_i)\ n_t^{-1}\sum_{i\in\boldsymbol{S}_t}g_{ti}(\theta_t)\right\}.$$

(2.120)

The expected values of the coefficients $\hat{\mathcal{B}}_{1\xi}$ and $\hat{\mathcal{B}}_{2\xi}$ are given by

$$E(\hat{\mathcal{B}}_{1\xi}) = n_2 N^2 E\left[\{det(\mathbf{\Sigma}_{cc}^{\circ})\}^{-1}\right] E\left[Cov_1\left\{\xi(w_i), g_{1i}(\theta_1)\right\}\right] \quad (2.121)$$

$$E(\hat{\mathcal{B}}_{2\xi}) = n_1 N^2 E\left[\{det(\mathbf{\Sigma}_{cc}^{\circ})\}^{-1}\right] E\left[Cov_2\left\{\xi(w_i), g_{2i}(\theta_2)\right\}\right]. \quad (2.122)$$

The factor $N^2 E\left[\{det(\mathbf{\Sigma}_{cc}^{\circ})\}^{-1}\right]$ appears in both coefficients. The expected values of the covariances, $E\left[Cov_1\left\{\xi(w_i), g_{1i}(\theta_1)\right\}\right]$ and $E\left[Cov_2\left\{\xi(w_i), g_{2i}(\theta_2)\right\}\right]$ can be assumed to be of the same order, as $g_{1i}(\theta_1)$ and $g_{2i}(\theta_2)$ are values of the same function of the same variable. Therefore, when $n_1 \gg n_2$, we have that $E(\hat{\mathcal{B}}_{2\xi}) \gg E(\hat{\mathcal{B}}_{1\xi})$. When the coefficient $\hat{\mathcal{B}}_{1\xi}$ is very small, it has a negligible effect on the variance of $\widehat{G}_{r1}(\theta_1)$ (see (2.97)). The variance of $\widehat{G}_{r2}(\theta_2)$, however, would be highly influenced by the large term $\hat{\mathcal{B}}_{2\xi}$. A simulation study demonstrating performance of the aligned empirical likelihood estimator when two samples of very different sizes are aligned is presented in chapter 2.12.2.

Note that when the common parameter $\boldsymbol{\xi}$ is the only parameter of interest, the aligned empirical likelihood estimator is not the most efficient way of combining information from two samples. Chapter 3 introduces a general form of an empirical likelihood estimator that can be more precise in this situation. The relationship between that estimator and the aligned empirical likelihood estimator is discussed in chapter 3.6.

2.8 The empirical likelihood ratio statistic

In this chapter we show that the empirical likelihood ratio statistic defined for the two samples joint empirical log-likelihood function follows a χ^2 distribution asymptotically.

Consider the following empirical likelihood ratio statistic:

$$\widehat{r}(\boldsymbol{\theta}|\boldsymbol{\varphi}_{U}) = 2\left\{\ell(\widehat{\boldsymbol{m}}) - \ell(\boldsymbol{\theta}|\boldsymbol{\varphi}_{U})\right\}, \tag{2.123}$$

where $\ell(\widehat{\boldsymbol{m}})$ is defined by (2.39) with $\widehat{m}_i(\boldsymbol{\varphi}_{U})$ given by (2.61) and $\ell(\boldsymbol{\theta}|\boldsymbol{\varphi}_{U})$ is defined by (2.56).

The following theorem establishes the asymptotic distribution of (2.123).

Theorem 3. *Under conditions* (2.70)-(2.75)*,* (2.79) *and* (2.80)*, and assuming that the central limit theorem holds for the vector* (2.89)*, that is, (e.g. Scott and Wu, 1981)*

$$\widehat{\boldsymbol{V}}_{\mathcal{P}}\big\{\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{U})\big\}^{-1/2}\ \widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{U}) \xrightarrow{d} \boldsymbol{\mathcal{N}}(\mathbf{0}_{\nu}, \boldsymbol{I}_p), \tag{2.124}$$

where $\boldsymbol{\mathcal{N}}(\mathbf{0}_{\nu}, \boldsymbol{I}_p)$ *denotes the standardized multivariate normal distribution and* $\boldsymbol{I}_p$ *denotes the* $p \times p$ *identity matrix, we have that:*

$$\widehat{r}(\boldsymbol{\theta}_{U}|\boldsymbol{\varphi}_{U}) = \widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{U})^{\top}\ \widehat{\boldsymbol{V}}_{\mathcal{P}}\big\{\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{U})\big\}^{-1}\ \widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{U}) + O_{\mathcal{P}}(n^{-1/2}), \tag{2.125}$$

where $\widehat{\boldsymbol{V}}_{\mathcal{P}}\big\{\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{U})\big\}$ *is given by* (2.97) *with* $\boldsymbol{\theta} = \boldsymbol{\theta}_{U}$*.*

The assumption (2.124) is plausible as the random vector (2.89) is a smooth function of Horvitz and Thompson (1952) estimators.

Under high entropy designs and if $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are independent, $\widehat{\boldsymbol{V}}_{\mathcal{P}}\big\{\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{U})\big\}$ is a design-consistent estimator of the variance-covariance matrix of $\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{U})$ (see chapter 2.6.2 and (Berger and Kabzinska, 2017)). Therefore, the assumption (2.124) and Theorem 3 imply that

$$\widehat{r}(\boldsymbol{\theta}_{U}|\boldsymbol{\varphi}_{U}) \xrightarrow{d} \chi^2_{df=p}, \tag{2.126}$$

where $\chi^2_{df=p}$ denotes a χ^2-distribution with p degrees of freedom and p is the number of equations in (2.58). Thus, $\widehat{r}(\boldsymbol{\theta}_{U}|\boldsymbol{\varphi}_{U})$ is a pivotal statistic, i.e., its asymptotic

distribution does not depend on $\boldsymbol{\theta}_U$.

2.9 Tests and empirical likelihood confidence regions

Property (2.126) allows us to use the empirical likelihood ratio statistic (2.123) to construct confidence regions or confidence intervals for the parameter $\boldsymbol{\theta}_U$ and to test hypotheses about $\boldsymbol{\theta}_U$.

Suppose that we wish to test $H_0 : \boldsymbol{\theta}_U = \boldsymbol{\theta}_U^0$ against $H_A : \boldsymbol{\theta}_U \neq \boldsymbol{\theta}_U^0$. Under H_0, we have $\widehat{r}(\boldsymbol{\theta}_U^0|\boldsymbol{\varphi}_U) \xrightarrow{d} \chi^2_{df=p}$. The p-value is given by $\int_{\widehat{r}(\boldsymbol{\theta}_U^0|\boldsymbol{\varphi}_U)}^{\infty} f(x)dx$, where $f(x)$ is the density of the χ^2-distribution with p degrees of freedom.

An α-level confidence region for $\boldsymbol{\theta}_U$ is defined as the set of $\boldsymbol{\theta}_U^0$ such that $H_0 : \boldsymbol{\theta}_U = \boldsymbol{\theta}_U^0$ is not rejected at the $1-\alpha$ level (p-value $\geqslant 1-\alpha$); that is,

$$\alpha\text{-level confidence region of } \boldsymbol{\theta}_U := \left\{\boldsymbol{\theta} : \widehat{r}(\boldsymbol{\theta}|\boldsymbol{\varphi}_U) \leqslant \chi^2_{df=p}(\alpha)\right\}, \qquad (2.127)$$

where $\chi^2_{df=p}(\alpha)$ denotes the upper α-quantile of the $\chi^2_{df=p}$-distribution. Based on (2.126), this confidence region is asymptotically consistent. This means that the nominal coverage α is asymptotically achieved.

It is also possible to construct a confidence interval for a single scalar parameter θ_U when it is entirely defined by a single estimating equation which does not involve any unknown parameters. For example, if we wish to construct a confidence interval for the mean $\theta_U = N^{-1}\sum_{i\in U} y_{ti}$, the single estimating function can be defined as $g_{ti}(\theta_t) = y_{ti} - \theta$. The same principle can be used for totals, quantiles or ratios of any parameters defined by an estimating equation. With a single scalar estimating function ($p = 1$), the confidence region (2.127) reduces to a confidence interval. Practical aspects of finding the empirical likelihood confidence regions and confidence intervals are discussed in chapter 5.

2.10 Stratification

In the previous chapters we considered single stratum samples for brevity. In this chapter we show how stratification can be included. This requires adjusting the design constraint (2.42) using a method proposed by Berger and De La Riva Torres (2016), with a small change to account for the two samples setup. The adjustment for stratification is not necessary to calculate point estimates, but is important for construction of confidence intervals and regions.

Suppose that the population U is split into H_t groups $\{U_{t1}, \ldots, U_{th}, \ldots, U_{tH_t}\}$, which are disjoint and such that $\cup_{h=1}^{H_t} U_{th} = U$ $(t = 1, 2)$. A sample $\boldsymbol{S}_{th}$ of fixed size n_{th} is selected with unequal probabilities from U_{th}. We have $\boldsymbol{S}_1 = \cup_{h=1}^{H_1} \boldsymbol{S}_{1h}$, $\boldsymbol{S}_2 = \cup_{h=1}^{H_1} \boldsymbol{S}_{2h}$ and $n_t = \sum_{h=1}^{H_t} n_{th}$. Note that each of the samples can be stratified in a different way. We assume that the number of strata H_t is bounded.

Information about the stratification is included in the design (or stratification) variables: $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$. The values of $\boldsymbol{z}_t$ for unit i are given by the H_t-vector

$$\boldsymbol{z}_{ti} := (z_{t1i}, \ldots, z_{thi}, \ldots, z_{tH_ti})^\top, \tag{2.128}$$

with $z_{thi} = \pi_{ti}$ when $i \in U_{th}$ and $z_{thi} = 0$ otherwise.

When samples are selected using a stratified design, the constraint (2.42) takes the following form:

$$\sum_{i \in \boldsymbol{S}_1} m_{1i}\ \boldsymbol{z}_{1i} = \boldsymbol{n}_1^{(H)} \quad \text{and} \quad \sum_{i \in \boldsymbol{S}_2} m_{2i}\ \boldsymbol{z}_{2i} = \boldsymbol{n}_2^{(H)}, \tag{2.129}$$

where $\boldsymbol{n}_t^{(H)} = (n_{t1}, n_{t2}, ..., n_{tH})^\top$ is the vector of strata sample sizes. Theorems (1), (2) and (3) hold under stratified sampling designs (Berger and Kabzinska, 2017).

2.11 Without replacement sampling and large sampling fractions

In this chapter we show how the proposed approach can be adjusted to accommodate large sampling fractions.

When samples are selected without replacement, but the sampling fraction is negligible, the proposed approach can be used without changes. For large sampling fraction designs, the method proposed by Berger and De La Riva Torres (2016) and extended to the alignment case by Berger and Kabzinska (2017) can be used. This approach is based on using the so-called penalised empirical likelihood function and including finite population correction factors in the constraints system.

Theorems (1) and (2) hold under large sampling fractions, i.e., when the assumptions $n_1^{(\nu)}/N \to 0$ and $n_2^{(\nu)}/N \to 0$ are substituted by an assumption that $n_t^{(\nu)}/N \leqslant \gamma_t$, where γ_t is a constant such that $\gamma_t < 1$ (Berger and Kabzinska, 2017). For theorem (3) to hold, the empirical likelihood function (2.38) has to be replaced by the penalised empirical likelihood function

$$\widetilde{\ell}(\boldsymbol{m}) := \sum_{i \in \boldsymbol{S}} \{\log(m_i) + 1 - \pi_i m_i\} \cdot \tag{2.130}$$

The empirical likelihood ratio function (2.123) is replaced by

$$\widetilde{r}(\boldsymbol{\theta}|\boldsymbol{\varphi}_U) = 2\left\{\widetilde{\ell}(\widetilde{\boldsymbol{m}}) - \widetilde{\ell}(\boldsymbol{\theta}|\boldsymbol{\varphi}_U)\right\}, \tag{2.131}$$

with $\widetilde{\ell}(\widetilde{\boldsymbol{m}}) = \sum_{i \in \boldsymbol{S}} \log \widetilde{m}_i(\boldsymbol{\varphi}_U)$ and $\widetilde{\ell}(\boldsymbol{\theta}|\boldsymbol{\varphi}_U) = \sum_{i \in \boldsymbol{S}} \log \widetilde{m}_i^{\star}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)$. The adjusted weights $\widetilde{m}_i(\boldsymbol{\varphi}_U)$ are defined as values which maximise (2.130) under $m_i > 0$ and the constraint

$$\sum_{i \in \boldsymbol{S}} m_i \, \widetilde{\boldsymbol{c}}_i = \widetilde{\boldsymbol{C}}, \tag{2.132}$$

while $\widetilde{m}_i^{\star}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)$ maximise (2.130), for a given vector $\boldsymbol{\theta}$, under $m_i > 0$ and the constraint

$$\sum_{i \in \boldsymbol{S}} m_i \; \widetilde{\boldsymbol{c}}_i^{\star}(\boldsymbol{\theta}) = \widetilde{\boldsymbol{C}}^{\star}. \tag{2.133}$$

The constraint matrices take the following form:

$$\begin{aligned}
\widetilde{\boldsymbol{c}}_i^{\star}(\boldsymbol{\theta}) &:= \{ \; \widetilde{\boldsymbol{c}}_i^{\top} \; , \; \widetilde{\boldsymbol{g}}_i(\boldsymbol{\theta})^{\top} \; \}^{\top}, \\
\widetilde{\boldsymbol{C}}^{\star} &:= \{ \widetilde{\boldsymbol{C}}^{\top} \; , \; \widetilde{\boldsymbol{g}}(\boldsymbol{\theta})^{\top} \}^{\top}, \\
\widetilde{\boldsymbol{c}}_i &:= q_i \{ \boldsymbol{z}_i^{\top} \; , \; \mathbf{f}_i(\boldsymbol{\varphi}_U)^{\top} \; , \; \boldsymbol{\xi}_i^{\circ \top} \}^{\top}, \\
\widetilde{\boldsymbol{C}} &:= \{ \widetilde{\boldsymbol{z}}^{\top} \; , \; \widetilde{\mathbf{f}}(\boldsymbol{\varphi}_U)^{\top} \; , \; \widetilde{\boldsymbol{\xi}}^{\circ \top} \}^{\top},
\end{aligned}$$

$$\begin{aligned}
q_i &:= \delta_{1i} \; q_{1i} + \delta_{2i} \; q_{2i}, \\
q_{ti} &:= (1 - \pi_{ti})^{1/2}, \\
\boldsymbol{z}_i &:= \left(\; \delta_{1i} \boldsymbol{z}_{1i}^{\top} \; , \; \delta_{2i} \boldsymbol{z}_{2i}^{\top} \; \right)^{\top}, \\
\widetilde{\boldsymbol{g}}_i(\boldsymbol{\theta}) &:= q_i \; \boldsymbol{g}_i(\boldsymbol{\theta}), \\
\widetilde{\boldsymbol{z}} &:= \sum_{i \in \boldsymbol{S}} q_i \; \breve{\boldsymbol{z}}_i, \\
\widetilde{\boldsymbol{g}}(\boldsymbol{\theta}) &:= \sum_{i \in \boldsymbol{S}} (q_i - 1) \; \breve{\boldsymbol{g}}_i(\boldsymbol{\theta}), \\
\widetilde{\mathbf{f}}(\boldsymbol{\varphi}_U) &:= \sum_{i \in \boldsymbol{S}} (q_i - 1) \; \breve{\mathbf{f}}_i(\boldsymbol{\varphi}_U), \\
\widetilde{\boldsymbol{\xi}}^{\circ} &:= \sum_{i \in \boldsymbol{S}} (q_i - 1) \; \breve{\boldsymbol{\xi}}_i^{\circ},
\end{aligned}$$

where $\breve{\boldsymbol{g}}_i(\boldsymbol{\theta}) := \boldsymbol{g}_i(\boldsymbol{\theta}) \pi_i^{-1}$, $\breve{\boldsymbol{z}}_i := \boldsymbol{z}_i \pi_i^{-1}$, $\breve{\mathbf{f}}_i(\boldsymbol{\varphi}_U) := \mathbf{f}_i(\boldsymbol{\varphi}_U) \pi_i^{-1}$,$\breve{\boldsymbol{\xi}}_i^{\circ} := \boldsymbol{\xi}_i^{\circ} \pi_i^{-1}$ and $\boldsymbol{\xi}_i^{\circ}$ are defined by (2.53). The $\boldsymbol{z}_{ti}$ are the design variables (2.128). If there is a single stratum, $\boldsymbol{z}_i = \boldsymbol{p}_i$, where $\boldsymbol{p}_i$ is defined by expression (2.51).

The q_{ti} are finite population correction factors proposed by Hájek (1964). They reduce the effect of units with large sampling probabilities (Berger, 2005; Berger and De La Riva Torres, 2016). For large sampling fractions and moderate sample

sizes, Berger and Kabzinska (2017) recommend to substitute q_{ti} by $(1-\lambda_{ti})^{1/2}$, where the λ_{ti} are defined by the recursive formula (3.25) in Hájek (1981). The correction factors q_i are close to 1 when the sampling fraction is negligible.

2.12 Simulation studies

This chapter summarises the results of simulation studies carried out to assess the asymptotic properties of the aligned empirical likelihood estimator. We consider four different populations and samples of varying sizes. In chapters 2.12.1 and 2.12.2, we evaluate the precision of the aligned empirical likelihood point estimator. In chapter 2.12.3, we consider confidence intervals. The pseudoempirical likelihood estimator (Wu, 2004*a*) and the composite regression estimator (Merkouris, 2004) are calculated for reference. In Tables 2.1-2.7, 'AEL' refers to the aligned empirical likelihood approach proposed. 'PEL' refers to the pseudoempirical likelihood approach (Wu, 2004*a*). 'Com.' refers to the composite regression estimator (Merkouris, 2004). 'Reg.' refers to the single sample calibration estimator (Deville and Särndal, 1992*a*). The simulations were performed in the statistical environment R (R Core Team, 2015). For calculation of the pseudoempirical likelihood estimator, a revised version of Wu's (2005) code was used.

2.12.1 Point estimation

First, we consider an artificial population. N values y_{ti} are generated from the following models: $y_{1i} := 3+a_{1i}+x_{1i}+w_i-0.3e_{1i}$ and $y_{2i} := 12-a_{2i}-x_{2i}-0.5w_i+0.3e_{2i}$, where a_{ti}, x_{ti} and w_i are generated independently from an exponential distribution with rate 1 and $e_{ti} \sim \chi^2_{df=1}-1$. The generated values are treated as fixed. We consider $N = 100,000$, $N = 10,000$ and $N = 2,500$. This gives the following correlations: $cor(y_1, x_1) \approx 0.6$, $cor(y_1, w_1) \approx 0.6$, $cor(y_2, x_2) \approx -0.7$, $cor(y_2, w2) \approx -0.3$,

$cor(x_1, w_1) \approx 0$, $cor(x_2, w_2) \approx 0$. The selection probabilities π_{ti} are proportional to $a_{ti} + 2$, with extremely large and extremely small probabilities increased or reduced so that $0.8 < \pi_{ti}\ N/n_t < 1.2$. The auxiliary variable is x_{ti}. The common variable is w. A similar artificial population was proposed by Wu and Rao (2006). The totals of y_1 and y_2 are the parameters of interest in samples $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ respectively. Additionally, in each of the samples, the auxiliary variable x_t is measured. The common variable w is measured in both samples.

Second, we consider a similar artificial population with different correlation settings. The y_1 and y_2 are generated from the following models: $y_{1i} := 3 + a_{1i} + 3x_{1i} + 4w_i - 0.3e_{1i}$ and $y_{2i} := 12 - a_{2i} - 3x_{2i} - 2w_i + 0.3e_{2i}$. This gives the following correlations: $cor(y_1, x_1) \approx 0.6$, $cor(y_1, w_1) \approx 0.8$, $cor(y_2, x_2) \approx -0.8$, $cor(y_2, w2) \approx -0.5$, $cor(x_1, w_1) \approx 0$, $cor(x_2, w_2) \approx 0$. The other parameters are defined as in the first population.

The third population is the 2006 British Expenditure and Food Survey (Office for National Statistics and Department for Environment, Food and Rural Affairs, 2009) household dataset. The population size is $N = 6,645$. The number of people living in the household is used as auxiliary information with a known population total. Gross weekly income is the common variable. The total expenditure on clothing is estimated from $\boldsymbol{S}_1$. The total expenditure on food is estimated from $\boldsymbol{S}_2$.The correlations are equal to: $cor(y_1, x_1) \approx 0.3$, $cor(y_1, w_1) \approx 0.3$, $cor(y_2, x_2) \approx 0.4$, $cor(y_2, w2) \approx 0.4$, $cor(x_1, w_1) \approx 0$, $cor(x_2, w_2) \approx 0.4$. The selection probabilities are proportional to the total household expenditure, with extremely large and extremely small probabilities increased or reduced so that $0.8 < \pi_{ti}\ N/n_t < 1.2$. Please note that the values that are reported in the Tables 2.1 and 2.8 do not reflect the official estimates from the British Expenditure and Food Survey.

The fourth population is the synthetic dataset AMELIA (Alfons et al., 2011). This dataset represents an artificial population of $N = 3,781,289$ households, with variables simulated from the 'European Union statistics on income and living

conditions survey' (Eurostat, 2012). The selection probabilities are proportional to the tax on income and social insurance contributions. The number of households in each region is used as an auxiliary variable in $\boldsymbol{S}_2$. Total gross household income for each of four domains defined by the variable 'districts' is estimated from $\boldsymbol{S}_2$. The sizes of the domains are 26%, 28%, 22% and 24% of the population size. In $\boldsymbol{S}_1$, no information on domains is recorded and the population total of gross household income is the parameter of interest. There are also no auxiliary variables in $\boldsymbol{S}_1$. The correlation between the variables are: $cor(y_1, w1) \approx 0.5$ (approximately equal for all domains), $cor(y_2, x_2) \approx 0.7$ (approximately equal for all domains), $cor(y_2, w_2)^{(D_1)} \approx 0.1$, $cor(y_2, w_2)^{(D_2)} \approx 0.2$, $cor(y_2, w_2)^{(D_3)} \approx 0.2$, $cor(y_2, w_2)^{(D_4)} \approx 0.3$, $cor(x_1, w_1) \approx 0$. Notation $cor(y_2, w_2)^{(D_k)}$ corresponds to the correlation in the k-th domain.

The simulations based on the first, second and third population consist of $10,000$ iterations. For each iteration, two samples of the same size, $n = n_1 = n_2$, are drawn using a randomised systematic sampling design (e.g. Tillé, 2006, §7.2). The empirical likelihood estimators of the totals of y_1 and y_2 are calculated by solving equation (2.58) with respect to $\boldsymbol{\theta}$, with $\boldsymbol{\theta} = (\theta_1, \theta_2)^\top$,

$$\boldsymbol{g}_i(\boldsymbol{\theta}) = \left(\delta_{1i}\ g_{1i}(\theta_1)^\top\ ,\ \delta_{2i}\ g_{2i}(\theta_2)^\top \right)^\top, \tag{2.134}$$

δ_{1i} and δ_{2i} are defined by (2.3) and

$$g_{ti}(\theta) = y_{t;i} - (\pi_{t;i} n_t^{-1} \theta_t)\cdot \tag{2.135}$$

Because of constraint (2.42), the solution simplifies to

$$\hat{\theta}_t = \sum_{i \in \boldsymbol{S}_t} \widehat{m}_i(\boldsymbol{\varphi}_U) y_{t;i}\cdot \tag{2.136}$$

The pseudoempirical likelihood estimator has been defined for means. Therefore, the population mean is estimated and multiplied by the known population size. The composite regression estimator can be applied to totals directly.

Table 2.1: Relative absolute root mean square errors (%) for estimators of totals of the non-common variables in three populations of interest, with both samples of equal sizes. Randomised systematic sampling design. 'AEL': proposed aligned empirical likelihood estimator. 'PEL': pseudoempirical likelihood approach (Wu, 2004*a*). 'Com.': composite regression estimator (Merkouris, 2004). 'Reg.': single sample calibration estimator (Deville and Särndal, 1992*a*).

		Sample 1				Sample 2			
Populations	n	AEL	PEL	Com.	Reg.	AEL	PEL	Com.	Reg.
Artificial 1 ($N = 100\,000$)	1000	0.5	0.9	2.3	1.5	1.3	1.4	2.8	1.9
	200	1.1	2.1	5.0	3.3	2.8	3.1	6.4	4.6
Artificial 1 ($N = 10\,000$)	500	0.7	1.3	3.1	2.1	1.7	1.9	4.0	2.8
Artificial 1 ($N = 2500$)	250	1.0	1.8	4.5	2.7	2.4	2.6	5.5	3.9
	160	1.2	2.3	5.4	3.7	3.1	3.3	6.7	4.8
	80	1.9	3.5	8.0	5.6	4.6	5.0	10.0	7.5
Artificial 2 ($N = 100\,000$)	1000	0.9	0.9	2.0	1.7	2.0	2.3	3.7	2.8
	200	2.0	2.1	4.3	3.7	4.7	5.2	8.8	6.8
Artificial 2 ($N = 10\,000$)	500	1.3	1.4	2.7	2.4	2.7	3.1	5.5	4.1
Artificial 2 ($N = 2500$)	250	1.6	1.7	3.6	3.0	3.7	4.1	6.9	5.5
	160	2.2	2.3	4.7	4.2	4.4	5.4	8.9	6.9
	80	3.2	3.4	6.8	5.8	7.0	8.3	13.1	10.8
Expenditure & Food	500	6.4	6.5	6.5	5.9	3.0	3.1	3.4	2.6
	1000	4.3	4.4	4.4	4.0	2.0	2.1	2.3	1.7

Table 2.1 shows the observed relative root mean square errors (RRMSE) of three estimators of totals of the non-common variables: the maximum empirical likelihood estimator (2.57), the pseudoempirical likelihood estimator and the composite regression estimator. The RRMSE of the estimators understandably vary between the populations and parameters of interest. In the first population, the estimators of θ_2 (i.e., estimators of the total of y_2 from sample $\boldsymbol{S}_2$), are less precise than the estimators of θ_1. This is likely to be because the correlation between y_{2i} and the common variable w_i is weaker than the relevant correlations between y_{1i} and w_i. Variable y_{2i} also follows a more skewed distribution than y_{1i}.

Unsurprisingly, we see that the RRMSE increases when the sample size decreases. Note that this loss of precision is influenced by the absolute sample size rather than by the sampling fraction. For example, in the second line of the table, we have $n = 200$ and $N = 100,000$, which corresponds to a sampling fraction of 0.2%. The RRMSE is lower than that in the sixth line, where the sampling fraction is higher (3.2%), but the sample is very small (80 units). With the very small sample sizes, all estimators seem to have a high RRMSE. The aligned empirical likelihood estimator performs relatively well. In this case the empirical likelihood - based estimators are slightly more precise than the composite regression estimator.

With the second population data, even though the absolute value of the strength of the correlations between the variables and the common and non-common auxiliary variables is higher, all the estimators have higher RRMSE than when the first population is used with the corresponding sample sizes. This is in line with the increase in the RRMSE of the single sample calibration estimators for totals of y_1 and y_2 and can be explained by an increased skewness of y_1 and y_2. The proposed estimator shows relatively low RRMSE compared to the calibration, pseudoempirical likelihood and composite regression estimators.

We observe large RRMSE with the third population. This is the case for all the estimators considered. This is likely to be caused by the very high skewness of the variables of interest. Note that in the case of the third population data, where the

variables follow skewed distributions and the correlations between the variables of interest and the common variable are weak, the aligned estimators of the non-common variables are less precise than the single sample calibration estimators.

Table 2.2 presents the relative absolute root mean square errors of the estimators of totals of the common variable. Single sample calibration estimator calculated from $\boldsymbol{S}_1$ is included for reference. The results from population *Artificial 2* are not included as they are the same as those from population *Artificial 1*. This is because the common variable w follows the same distribution in both populations and has equal, negligible, correlation with the auxiliary variables. The three estimators have comparable precision. Note that in all cases the aligned estimators are more precise than the single sample calibration estimators.

Table 2.2: Relative absolute root mean square errors (%) for estimators of totals of the common variable in two populations of interest, with both samples of equal sizes. Randomised systematic sampling design. 'AEL': proposed aligned empirical likelihood estimator. 'PEL': pseudoempirical likelihood approach (Wu, 2004*a*). 'Com.': composite regression estimator (Merkouris, 2004). 'Reg.': single sample calibration estimator (Deville and Särndal, 1992*a*).

Populations	n	AEL	PEL	Com.	Reg.
Artificial 1 ($N = 100\,000$)	1000	2.3	2.3	2.5	3.7
Artificial 1 ($N = 10\,000$)	500	3.4	3.3	3.6	5.2
Artificial 1 ($N = 2500$)	250	4.3	4.1	4.8	6.6
	160	5.9	5.5	6.3	9.0
	80	8.5	8.1	9.3	13.3
Expenditure & Food	500	1.9	2.5	2.3	3.2
	1000	1.2	1.7	1.6	2.2

Table 2.3 shows RRMSE of the aligned empirical likelihood , pseudoempirical likelihood and composite regression estimators applied to the fourth population data. Four separate simulations are caried out in each of the domains. Each simulation consists of $3,000$ iterations and the totals of y_1 and y_2 are the target parameters. The RRMSE are of course larger for the estimates of the domain totals than for the overall population totals. The three estimators considered have comparable precision.

Table 2.3: Relative absolute root mean square errors (%) for estimators of totals of the non-common variables in the artificial population *AMELIA*, with both samples of equal sizes. Randomised systematic sampling design. 'AEL': proposed aligned empirical likelihood estimator. 'PEL': pseudoempirical likelihood approach (Wu, 2004*a*). 'Com.': composite regression estimator (Merkouris, 2004).

		Sample 1			Sample 2		
	n	AEL	PEL	Com.	AEL	PEL	Com.
Domain 1 (26%)	3000	3.2	3.3	3.3	6.5	6.5	6.8
Domain 2 (28%)	3000	2.8	2.8	2.8	5.3	5.3	5.5
Domain 3 (22%)	3000	3.3	3.3	3.3	6.1	6.1	6.3
Domain 4 (24%)	3000	3.1	3.1	3.1	5.7	5.8	5.8

2.12.2 Samples of different sizes

In Tables 2.1 - 2.3, we considered $n_1 = n_2$. For the next series of simulations, we investigate the effect of small n_1 compared to n_2. The estimates of $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are dependent because of the alignment constraint (2.44), which can intuitively be explained as calibrating estimates of one sample towards the estimates of the other sample. In chapter 2.7 we show that the coefficient $\widehat{\boldsymbol{B}}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)$ of the generalized regression estimator asymptotically equivalent to the aligned empirical likelihood estimator is weighted by the sample sizes, which causes the adjustment applied to the weights of the larger sample to be smaller than the adjustment applied to the weights of the smaller sample. Here we show results of a simulation which aims to asses if and, if so, how much, the estimates of the non-common parameters from the large sample $\boldsymbol{S}_2$ are deteriorated by alignment with the small sample $\boldsymbol{S}_1$.

We use an artificial population of size $N = 100,000$ generated according to the following model:

$$y_t = 3 + a + 2x_t + 2w + 0.3e, \tag{2.137}$$

with a and e defined as in chapter 2.12.1 and $\pi_{ti} \sim a + 2$. This gives the following correlation settings: $cor(y_1, x_1) \approx 0.7$, $cor(y_1, w_1) \approx 0.7$, $cor(y_2, x_2) \approx 0.7$,

$cor(y_2, w2) \approx 0.7$, $cor(x_1, w_1) \approx 0$, $cor(x_2, w_2) \approx 0$.

Note that y_1 and y_2 are generated from the same model, so that the effect of differences in sample sizes is not confused with the effect of different distributions of the two parameters of interest. The parameter of interest in each of the samples is the total of y_t. The common variable is w and x_1 and x_2 are used as auxiliary variables in the first and the second sample respectively, with the population totals known. The variables x_1, x_2 and the common variable w follow either exponential distribution with the rate parameter equal to one, normal distribution with mean equal to zero and standard deviation equal to one, or normal distribution with mean equal to five and standard deviation equal to one. The distributions are given in the results tables. The simulation consists of $10,000$ iterations. In each iteration, the size of $\boldsymbol{S}_2$ is $1,000$. We let the size of $\boldsymbol{S}_1$ vary between 100 and $1,000$ units. Samples are selected using random systematic sampling.

In Table 2.4, we have the RRMSE of the proposed estimator of the totals of y_1 and y_2 obtained from $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$, with the distributions of y_1 and y_2 defined by (2.137). We also have the RRMSE of the pseudoempirical likelihood approach (Wu, 2004*a*), the composite regression estimator (Merkouris, 2004) and the single sample calibration estimator (Deville and Särndal, 1992*a*). We include two versions of the composite regression estimator, with the adjustment factor equal respectively to $q_{ti} = n_t(1 - n_t N^{-1})^{-1}$ and $q_{ti} = \pi_{t;i}(1 - \pi_{t;i})^{-1}$ (Merkouris, 2010*a*, section 3.1).

We notice that the proposed estimator is always at least as precise as the single sample calibration estimator. The difference in precision of these two estimators is low or none in the large sample $\boldsymbol{S}_2$, especially when n_1 is small. However, the proposed estimator is always more precise than the calibration estimator in $\boldsymbol{S}_1$.

We observe a slight deterioration of the proposed estimator based on $\boldsymbol{S}_2$ as n_1 decreases, yet the relative root mean square error never exceeds the relative root mean square error of the single sample calibration estimator. As expected, the relative root mean square error of the proposed estimator based on the smaller

Table 2.4: Relative absolute root mean square errors (%) for estimators of totals of the non-common variables with alignment of samples of different sizes. Randomised systematic sampling design. $N = 100\,000$, $n_2 = 1000$, n_1 varies as described in the table. Artificial data generated according to the model $y_t = 3 + a + 2x_t + 2w + 0.3e$ with a and e defined as in chapter 2.12.1, $\pi_{ti} \sim a + 2$. 'AEL': proposed aligned empirical likelihood estimator. 'PEL': pseudoempirical likelihood approach (Wu, 2004*a*). 'Com.[a]': composite regression estimator (Merkouris, 2004) with no adjustment for different sample sizes. 'Com.[b]': 'adjusted' composite regression estimator (Merkouris, 2004, 2010*a*) with $q_{ti} = n_t(1 - n_t N^{-1})^{-1}$. 'Com.[c]': 'adjusted' composite regression estimator (Merkouris, 2004, 2010*a*) with $q_{ti} = \pi_{t;i}(1 - \pi_{t;i})^{-1}$. 'Reg.': single sample calibration estimator (Deville and Särndal, 1992*a*).

			Sample 1						Sample 2					
n_1	x_t	w_t	Com.[a]	Com.[b]	Com.[c]	Reg.	PEL	AEL	Com.[a]	Com.[b]	Com.[c]	Reg.	PEL	AEL
100	$N(0,1)$	$N(5,1)$	2.7	1.3	1.4	3.1	0.9	0.9	3.5	1.9	0.9	0.9	1.1	0.8
	exp(1)	$N(5,1)$	3.3	2.7	7.8	4.3	1.7	1.6	10.5	5.8	1.2	1.2	3.7	1.2
	$N(0,1)$	exp(1)	3.3	2.7	7.8	4.3	1.7	1.6	10.5	5.8	1.2	1.2	3.7	1.2
	exp(1)	exp(1)	4.1	3.8	8.8	4.9	1.6	1.2	10.7	5.8	1.5	1.5	2.4	0.8
300	$N(0,1)$	$N(5,1)$	1.2	0.8	1.1	1.7	0.6	0.8	1.7	1.1	0.8	0.9	0.7	0.7
	exp(1)	$N(5,1)$	1.2	2.0	4.5	2.2	1.3	1.1	4.9	3.1	1.1	1.2	2.0	1.1
	$N(0,1)$	exp(1)	1.2	2.0	4.5	2.2	1.3	1.1	4.9	3.1	1.1	1.2	2.0	1.1
	exp(1)	exp(1)	1.9	2.6	5.2	2.7	1.1	0.9	5.4	3.4	1.4	1.5	1.4	0.8
600	$N(0,1)$	$N(5,1)$	0.7	0.7	0.8	1.1	0.5	0.7	0.9	0.8	0.7	0.9	0.5	0.7
	exp(1)	$N(5,1)$	1.4	1.9	2.9	1.5	1.2	1.0	2.8	2.2	1.1	1.2	1.4	0.9
	$N(0,1)$	exp(1)	1.4	1.9	2.9	1.5	1.2	1.0	2.8	2.2	1.1	1.2	1.4	0.9
	exp(1)	exp(1)	1.8	2.3	3.4	2.0	0.9	0.7	3.3	2.6	1.5	1.5	1.0	0.7
1000	$N(0,1)$	$N(5,1)$	0.6	0.6	0.6	0.9	0.4	0.6	0.6	0.6	0.6	0.9	0.4	0.6
	exp(1)	$N(5,1)$	1.8	1.8	1.7	1.3	1.2	0.9	1.9	1.9	1.8	1.2	1.2	0.9
	$N(0,1)$	exp(1)	1.8	1.8	1.7	1.3	1.2	0.9	1.9	1.9	1.8	1.2	1.2	0.9
	exp(1)	exp(1)	2.1	2.1	2.0	1.6	0.8	0.6	2.1	2.1	2.0	1.5	0.9	0.6

sample $\boldsymbol{S}_1$ increases as n_1 decreases. This deterioration is more pronounced when the common variable is skewed than when it is normally distributed. In all the cases, we observe a large gain in precision of the estimator based on $\boldsymbol{S}_1$, compared to the single sample calibration estimator. We should also note that the proposed estimator deals well with skewed auxiliary variables and a skewed common variable, compared to the calibration estimator and the composite regression estimator.

The pseudoempirical likelihood estimator of the first sample shows comparable precision to the proposed estimator. When n_1 is small, however, the pseudoempirical likelihood estimator of the second (large) sample is slightly less precise than the proposed estimator. The pseudoempirical likelihood estimator has slightly higher RRMSE than the empirical likelihood estimator when the auxiliary variables are skewed, and slightly lower RRMSE than the empirical likelihood estimator when the auxiliary variables are normally distributed.

The composite regression estimator of $\boldsymbol{S}_1$ with no adjustment for different sample sizes is always more precise than the corresponding calibration estimators when $n_1 < n_2$. We observe a deterioration of the non-adjusted composite regression estimator of the second (large) sample, when n_1 decreases. This is more pronounced when x and w are skewed. When there is a large difference between n_1 and n_2, the composite regression estimator adjusted for different sample sizes has much lower RRMSE than the corresponding composite regression estimator with no adjustment. The adjusted composite regression estimators of $\boldsymbol{S}_1$ are in most cases more precise than the single sample regression estimator. The adjusted composite regression estimator of $\boldsymbol{S}_2$, however, may be less precise than the single sample regression estimator when $\boldsymbol{S}_1$ is small and x and w are skewed. Two sample size adjustment factors were tested in the composite regression estimator:an adjustment based directly on the sample size ('Com.$^{\text{b}}$' with $q_{ti} = n_t(1 - n_t N^{-1})^{-1}$) and one based on the sampling probabilities ('Com.$^{\text{c}}$': with $q_{ti} = \pi_{t;i}(1-\pi_{t;i})^{-1}$). When the sample sizes are equal, there is no difference between these two estimators.

When there is a large difference between n_1 and n_2, estimator 'Com.[b]' seems to produce more precise estimates in $\boldsymbol{S}_1$ than 'Com.[c]', especially when x, w and y are skewed. 'Com.[c]', however, is more precise in $\boldsymbol{S}_2$. We should note that simple adjustment factors were used in the adjusted composite regression estimators. A semi-optimal adjustment factor based on the estimated variance and considering the sampling designs could be derived (see Merkouris, 2010*a*), which could lead to a more precise estimator.

Table 2.5 shows RRMSE of the estimators of the common variable. For comparison, the calibration estimator based on $\boldsymbol{S}_2$ is also calculated. We can see that the larger $\boldsymbol{S}_1$ gets, the more precise the aligned estimators are compared to the calibration estimator. The composite regression estimator with no adjustment for different sample sizes has lower RRMSE than the single sample calibration estimator only when $\boldsymbol{S}_1$ is of size 600 or larger and can be considerably less precise when $\boldsymbol{S}_1$ is small. The composite regression estimator 'Com.[c]', including the correction factor $q_{ti} = \pi_{t;i}(1 - \pi_{t;i})^{-1}$, is at least as precise as the calibration estimator, even with small n_1, and can be slightly more precise than the calibration estimator already when $n_1 = 300$. When the common variable is estimated, 'Com.[c]' is more precise than 'Com.[b]' in all cases.

The aligned empirical likelihood estimator is more precise than the calibration estimator in all cases, although the difference is marginal with very small $\boldsymbol{S}_1$ and increases as $\boldsymbol{S}_1$ grows. The pseudoempirical likelihood estimator is slightly less precise than the empirical likelihood estimator when $\boldsymbol{S}_1$ is very small and slightly more precise than the empirical likelihood estimator when $\boldsymbol{S}_1$ grows, particularly with normally distributed x and w.

Table 2.5: Relative absolute root mean square errors (%) for estimators of totals of the common variable with alignment of samples of different sizes. Randomised systematic sampling design. $N = 100\,000$, $n_2 = 1000$, n_1 varies as described in the table. Artificial data generated according to the model $y_t = 3 + a + 2x_t + 2w + 0.3e$ with a and e defined as in chapter 2.12.1, $\pi_{ti} \sim a + 2$. 'AEL': proposed aligned empirical likelihood estimator. 'PEL': pseudoempirical likelihood approach (Wu, 2004*a*). 'Com.[a]': composite regression estimator (Merkouris, 2004) with no adjustment for different sample sizes. 'Com.[b]': 'adjusted' composite regression estimator (Merkouris, 2004, 2010*a*) with $q_{ti} = n_t(1 - n_t N^{-1})^{-1}$. 'Com.[c]': 'adjusted' composite regression estimator (Merkouris, 2004, 2010*a*) with $q_{ti} = \pi_{t;i}(1 - \pi_{t;i})^{-1}$. 'Reg.': single sample calibration estimator (Deville and Särndal, 1992*a*).

n_1	x_t	w_t	Com.[a]	Com.[b]	Com.[c]	Reg.	PEL	AEL
100	$N(0,1)$	$N(5,1)$	3.5	2.0	1.1	1.1	1.1	1.0
	exp(1)	$N(5,1)$	10.4	6.1	3.4	3.5	5.8	3.4
	$N(0,1)$	exp(1)	10.4	6.1	3.4	3.5	5.8	3.4
	exp(1)	exp(1)	10.4	5.9	3.7	3.7	5.3	3.2
300	$N(0,1)$	$N(5,1)$	1.7	1.2	1.1	1.1	0.7	1.0
	exp(1)	$N(5,1)$	4.8	3.6	3.2	3.5	3.4	3.0
	$N(0,1)$	exp(1)	4.8	3.6	3.2	3.5	3.4	3.0
	exp(1)	exp(1)	5.2	3.8	3.4	3.7	3.5	3.0
600	$N(0,1)$	$N(5,1)$	1.0	0.9	0.9	1.1	0.5	0.9
	exp(1)	$N(5,1)$	2.9	2.7	2.8	3.5	2.6	2.7
	$N(0,1)$	exp(1)	2.9	2.7	2.8	3.5	2.6	2.7
	exp(1)	exp(1)	3.2	3.0	2.9	3.7	2.5	2.6
1000	$N(0,1)$	$N(5,1)$	0.8	0.8	0.8	1.1	0.5	0.8
	exp(1)	$N(5,1)$	2.5	2.5	2.5	3.5	2.4	2.5
	$N(0,1)$	exp(1)	2.5	2.5	2.5	3.5	2.4	2.5
	exp(1)	exp(1)	2.5	2.5	2.5	3.7	2.2	2.3

2.12.3 Confidence intervals: British Labour Force Survey

In this chapter we check the coverage of the proposed empirical likelihood confidence intervals. We apply the proposed methodology to two relatively difficult datasets: one with skewed variables and one with skewed variables and outliers. We also produce estimates for domains of varying sizes.

The population considered is a subset of $N = 89,181$ individuals with a non-zero gross weakly income selected from the 2013 'British Quarterly Labour Force Survey' (October-December) (Office for National Statistics. Social Survey Division, 2015). The parameter of interest, estimated from $\boldsymbol{S}_1$, is the total number of hours worked per week broken down by domains defined by the following industry sectors:

(i) Public administration, education and health
(ii) Distribution, hotels and restaurants
(iii) Banking and finance
(iv) Manufacturing
(v) Transport and communication
(vi) Construction
(vii) Other services
(viii) Agriculture, forestry, fishing, energy and water

The domains differ in size, as can be seen in Table 2.7.

The domain membership information is only collected in $\boldsymbol{S}_1$. This sample is used to estimate the total number of hours worked per domain. We introduce an alignment constraint on gross weekly pay, which is measured in both $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$. The sizes of the domains defined by the sectors (i)-(viii) are used as auxiliary variables in $\boldsymbol{S}_1$. No auxiliary variables are measured in $\boldsymbol{S}_2$. The correlations are respectively equal to: $cor(y_1, x_1) \approx 0.8$, $cor(y_1, w_1) \approx 0.3$, $cor(y_2, w2) \approx 0.7$, $cor(x_1, w_1) \approx 0.1$. In Table 2.7, we report the overall coverages and the tail error rates for the confidence intervals for the total number of hours worked based on $\boldsymbol{S}_1$.

Two samples, each of size 3,000, are selected 10,000 times using the randomised

systematic sampling design. The selection probabilities are proportional to the net weekly income.

Confidence intervals for the proposed estimator are defined by (2.127). Confidence intervals for the composite regression estimator are based on variances of separate regression estimators, as in Merkouris (2004). These variances are estimated by Hartley and Rao's (1962) estimator. The variance of the pseudoempirical likelihood estimator is based on the equivalence between the pseudoempirical likelihood estimator and a generalised regression estimator, as was proposed by Wu (2004*a*).

Table 2.6 shows the coverages and tail error rates of the proposed estimator. For comparison, we also show the coverages and tail error rates of the composite regression estimator and the pseudoempirical likelihood estimator. Because we are no longer focusing on the precision gains, the results for the single sample calibration estimator were omitted. In Table 2.7, we test the effect of outlying values on the coverage of confidence intervals. We introduce into the distribution of the variable of interest 5% of artificial outliers generated independently from an uniform distribution $U(y_{\max}, 3 \times y_{\max})$, where $y_{\max}$ is the maximum value of the total number of hours worked per week observed in the sub-sample.

The coverages of the empirical likelihood confidence intervals are similar to the coverages of the pseudoempirical likelihood and the composite regression confidence intervals. In Table 2.6, where no outliers were introduced into the variable of interest, the confidence interval coverages are close to the nominal level for all but the two smallest domains. When outliers are present, the under-coverage of confidence intervals in the smallest domains increases. This is true for all the methods. The empirical likelihood confidence intervals seem to have marginally better coverage than other estimators with a moderate domain size (domain vii), whether or not the outliers are present.

The tail error rates of all the considered confidence intervals are unbalanced and significantly different from 2.5% in several cases. The left tail error rates are

Table 2.6: Coverages and tail error rates of confidence intervals for total number of hours worked per week per domains. British Quarterly Labour Force Survey. No outliers introduced. N_d = population domain size, $\bar{n}_d$ = average domain sample size. Confidence intervals based on the first sample. 'AEL': proposed aligned empirical likelihood estimator. 'PEL': pseudoempirical likelihood approach (Wu, 2004*a*). 'Com.': composite regression estimator (Merkouris, 2004). The values reported in this Table do not reflect the official estimates from the British Labour Force Survey. †: coverages (or tail error rates) significantly different from 95% (or 2.5%), with p-value $\leqslant 0.05$.

Domains	Coverages			Left tail err. rates			Right tail err. rates				
(Sectors)	AEL	PEL	Com.	AEL	PEL	Com.	AEL	PEL	Com.	N_d/N	$\bar{n}_d$
(i)	95.0	96.0†	94.8	2.8	1.9†	2.1†	2.2	2.1†	3.1†	0.36	1090
(ii)	95.1	95.1	95.4	2.2	1.3†	1.2†	2.7	3.6†	3.4 †	0.18	552
(iii)	95.1	93.7†	94.2†	2.1†	1.9†	1.7†	2.7	4.4†	4.1†	0.14	434
(iv)	95.3	93.9†	94.0†	2.3	2.0†	2.3	2.4	4.1†	3.7†	0.11	327
(v)	95.2	95.4	95.4	2.1†	1.3†	1.4†	2.6	3.3†	3.2†	0.08	249
(vi)	94.7	93.9†	94.1†	2.0†	1.3†	1.2†	3.2†	4.8†	4.7†	0.05	139
(vii)	93.8†	91.7†	92.6†	2.2	1.0†	0.8†	4.0†	7.3†	6.6†	0.04	131
(viii)	94.0†	93.3†	93.4†	3.0†	2.1†	2.1†	3.1†	4.6†	4.5†	0.03	79

Table 2.7: Coverages and tail error rates of confidence intervals for total number of hours worked per week per domains. British Quarterly Labour Force Survey. 5 % outliers introduced into the variable of interest. N_d = population domain size, $\bar{n}_d$ = average domain sample size. Confidence intervals based on the first sample. 'AEL': proposed aligned empirical likelihood estimator. 'PEL': pseudoempirical likelihood approach (Wu, 2004*a*). 'Com.': composite regression estimator (Merkouris, 2004). The values reported in this Table do not reflect the official estimates from the British Labour Force Survey. †: coverages (or tail error rates) significantly different from 95% (or 2.5%), with p-value $\leqslant 0.05$.

Domains	Coverages			Left tail err. rates			Right tail err. rates				
(Sectors)	AEL	PEL	Com.	AEL	PEL	Com.	AEL	PEL	Com.	N_d/N	$\bar{n}_d$
(i)	94.4†	94.2†	94.3†	2.6	1.3†	1.3†	3.0†	4.5†	4.4†	0.36	1090
(ii)	95.0	94.8	94.6	2.4	1.2†	1.3†	2.7	4.0†	4.1†	0.18	552
(iii)	95.3	95.0	94.9	2.0†	0.9†	0.9†	2.7	4.1†	4.2†	0.14	434
(iv)	94.8	94.7	94.3†	2.4	0.9†	1.0†	2.7	4.4†	4.6†	0.11	327
(v)	95.8†	94.6†	94.4†	2.2†	1.2†	0.9†	2.0†	4.2†	4.7†	0.08	249
(vi)	93.7†	93.1†	92.9†	2.7	0.5†	0.5†	3.6†	6.5†	6.6†	0.05	139
(vii)	94.9	92.6†	92.4†	2.2†	0.7†	0.7†	3.0†	6.7†	6.9†	0.04	131
(viii)	90.0†	89.4†	89.0†	3.0†	0.9†	0.9†	7.0†	9.8†	10.1†	0.03	79

lower than 2.5% and the right tail error rates are higher than 2.5%. This effect is explained by the positive skewness of the data. The tails are even more unbalanced when outliers are present. However, the tail error rates of the empirical likelihood confidence interval are more balanced than those of the symmetric confidence intervals and usually closer to 2.5%. We observe these slightly better tail error rates, because the confidence interval (2.127) is determined by the distribution of the data. This is a common feature of empirical likelihood (e.g. Owen, 2001).

2.12.4 Confidence intervals: quantiles

In this chapter we apply the proposed method to estimation of quantiles of distribution. We use the estimating function for α-quantile proposed by Berger and De La Riva Torres (2016):

$$g_i(\theta) = \zeta(y_{(i)}, \theta) - \alpha, \tag{2.138}$$

where

$$\zeta(y_{(i)}, \theta) = \delta(y_i \leq \theta) + \frac{\theta - y_{(i-1)}}{y_{(i)} - y_{(i-1)}} \delta(y_{(i-1)} \leq \theta)\{1 - \delta(y_{(i)} \leq \theta)\}, \tag{2.139}$$

$y_{(i)}$ is the value of the i-th unit when all the units in the sample are arranged in increasing order and $\delta(\cdot)$ is an indicator function equal to 1 if the argument is true and to 0 otherwise. The composite regression estimator and the pseudoempirical likelihood estimator are not considered in this section, because they were developed for means or totals. For this series of simulations, we consider populations *Artificial 1*, *Expenditure & Food* and *AMELIA* described in chapter 2.12.1. In each iteration two samples of the same size, $n_1 = n_2$, are drawn using a randomised systematic sampling design with the selection probabilities as described in chapter 2.12.1. Each simulation is based on $10,000$ iterations.

Coverages and tail error rates of the empirical likelihood confidence interval (2.127) are presented in Table 2.8. The overall coverages and tail error rates are of an acceptable order. We observe low coverages, going down to 92.5%, for the smallest sample size ($n = 80$). Some of the simulations use non-negligible sampling fractions. This is the case for the artificial population with $N = 2,500$ and $n = 240$ and the 'Expenditure and Food Survey data' with $n = 1,000$. Acceptable coverages are observed in these cases. However, the tail error rates are unbalanced. Further analyses, not presented here, showed that this effect is associated with the skewness of the selection probabilities. With normally distributed selection probabilities, both tails are approximately equal to 2.5%. The effect of the skewness of the selection probabilities was only observed for large sampling fractions.

Table 2.8: Coverages (Cov.) and tail error rates (Left & Right) of confidence intervals for 80% and 90% quantiles. Population data described in chapter 2.12.1. Confidence intervals for total expenditure on clothing within four domains, estimates based on the first sample. †: coverages (or tail error rates) significantly different from 95% (or 2.5%).

		80% Quantiles			90% Quantiles		
Population	n	Cov.	Left	Right	Cov.	Left	Right
Artificial 1 ($N = 100\,000$)	1000	94.7	2.7	2.6	95.1	2.4	2.6
	200	94.4†	2.2†	3.4†	94.5†	2.0†	3.5†
Artificial 1 ($N = 10\,000$)	500	95.3	2.9†	1.9†	94.7	2.9†	2.4
Artificial 1 ($N = 2\,500$)	240	94.5†	3.1†	2.4	94.7	2.8	2.5
	160	94.7	2.8	2.6	94.0†	2.5	3.5†
	80	93.9†	1.9†	4.3†	92.5†	1.9†	5.6†
Expenditure & Food	500	95.1	2.5	2.4	94.9	2.3	2.8
(Tot. exp. clothing)	1000	94.7	3.7†	1.7†	94.9	3.6†	1.6†
AMELIA							
Domain 1 (26%)	3000	94.8	2.5	2.7	95.2	2.5	2.4
Domain 2 (28%)	3000	95.0	2.2	2.8†	94.6	2.7	2.8†
Domain 3 (22%)	3000	94.4†	2.4	3.1†	94.1	2.9†	3.0†
Domain 4 (24%)	3000	95.4	1.9†	2.7	94.5	2.1†	3.4†

2.13 Conclusions

We propose a novel empirical likelihood method for aligning estimates from multiple surveys in the presence of population level auxiliary information. It can handle stratification and is applicable to large sampling fractions.

The proposed approach can be used to estimate a class of parameters expressed as solutions of estimating equations, to construct confidence intervals, and to test the statistical significance of parameters of interest. The proposed approach does not rely on linearisation, re-sampling or joint selection probabilities.

In our simulation studies, the proposed approach gives point estimates of an acceptable precision and confidence intervals with good coverage, as long as the sample is sufficiently large. When the variables of interest are skewed, contain outliers and when samples differ hugely in size, the empirical likelihood estimator may be more precise than the regression-based methods.

Aligning estimates of a small sample to estimates of a much larger sample was found to considerably increase the precision of the point estimator based on the smaller sample. This is in line with the results reported for the composite regression estimator (Merkouris, 2010*a*). The aligned empirical likelihood estimator performs well in these settings. The precision of the point estimator from the large sample is only very slightly deteriorated. Note that this is achieved without estimation of any adjustment factors. If a precise variance estimate can be easily obtained, then the adjusted composite regression estimator may be more efficient. However, when a complex sampling design is used and the variable of interest is skewed, this variance might be difficult to estimate and therefore the empirical likelihood estimator might be preferred, as it has acceptable performance.

The aligned empirical likelihood estimator works well in samples of large or moderate size. In very small samples, the empirical likelihood confidence intervals

tend to show some under-coverage. However, the symmetric confidence intervals are also not free from this problem. We can see that when the variables of interest and the auxiliary variables are skewed, the empirical likelihood tail error rates tend to be more balanced than the error rates of symmetric confidence intervals.

The simulation studies show that in some selected cases the aligned empirical likelihood estimator is more precise than the regression based estimators. The main purpose of applying empirical likelihood to aligning estimates, however, is not an increase in precision. In fact, if the second order selection probabilities, or precise variance estimates are available, an adjusted composite regression estimator is likely to be more precise. The aligned empirical likelihood estimator, however, is computationally simpler as no variance estimates are required. More importantly, it also accommodates more complex parameters than means and totals. This applies both to the parameters of interest and to the known population parameters. Moreover, the aligned empirical likelihood allows to choose a function of the common variable used in the alignment constraint which maximises the correlation with the variables of interest. This may bring gains in precision.

The main limitation of the aligned empirical likelihood estimator is in the assumption of independence between the samples. Note that the $\sqrt{n}$ design-consistency of the point estimator holds whether or not the samples are independent. However, the pivotal property of the log-likelihood ratio function relies on the independence of the samples. This makes the proposed approach suitable only for independent surveys, but unsuitable e.g. for longitudinal studies or nested two-phase sampling. In the case of dependent samples, the composite regression estimator can be used (see Merkouris, 2015).

Finally, while it is not the focus of this piece of work, it is worth noting that aligning estimates requires careful selection of variables and some survey design effort so that the variables used for alignment indeed measure the same characteristic. This includes e.g. harmonizing the question wording across surveys (see e.g. (Karlberg et al., 2015)). It is also necessary that the surveys are carried out close enough

in time so that we can reasonably assume that they are both carried out in the same population and that any in- and out-migration, as well as changes in the population characteristics, are negligible.

Chapter 3

Empirical likelihood multiplicity adjusted estimator for multiple frame surveys

3.1 Introduction

Using more than one sampling frame may improve coverage of the target population, increase precision of estimation of target parameters or reduce sampling cost, especially when a single frame containing all population units is not available or expensive to sample from. For instance, mobile phone frames are increasingly used together with landlines in CATI research (e.g. Barr et al., 2012) in order to increase coverage in surveys of the general population. Multiple frames are also used to oversample rare populations (Kalton, 2009). Inference from multiple frame surveys has attracted a lot of researchers' attention and several multiple frame estimators are available.

Recent papers by Singh and Mecatti (2011) and Mecatti and Singh (2014) showed how most of the existing multiple frame estimators can be expressed in the form of the Generalized Multiplicity Adjusted Horwitz-Thompson estimator. The idea of multiplicity estimation consists of pooling all the units selected from all the frames into one sample and finding a vector of adjustment factors which is applied to the design weights so that the increased selection probability of units which

appear in more than one frame is accounted for. This approach can be applied to inference from multiple frame (i.e., not only dual frame) surveys. It can also be applied to other sampling designs. For example, the Generalized Weight Share estimator used to make inference from indirect sampling surveys (Lavallée, 2009) can be expressed as a GMHT estimator (Singh and Mecatti, 2011).

We propose an Empirical Likelihood inference method which adopts the flexible multiplicity approach, allowing for selection of various multiplicity adjustments, and can easily handle additional calibration type constraints. The proposed multiplicity adjusted empirical likelihood estimator is derived from the design based empirical likelihood approach proposed for single frame surveys by Berger and De La Riva Torres (2016). It shares the benefits of Berger and De La Riva Torres's (2016) method, such as suitability for estimation of parameters with a skewed distribution or range preserving confidence intervals obtained directly from a log-likelihood ratio function.

Below we start with a brief summary of the problem and describe some existing estimators for dual frame surveys. We discuss how this approach is generalised to multiple frames through the multiplicity approach of Singh and Mecatti (2011) and include examples of some of the available multiplicity adjusted estimators.

We then show how Berger and De La Riva Torres's (2016) design based empirical likelihood can be extended to accommodate multiple frame surveys and discuss the properties of the resulting multiplicity adjusted empirical likelihood estimator. The logic of deriving the multiplicity adjusted estimator is similar to that shown in chapter 2 for the aligned empirical likelihood estimator, although there are some differences in the constraints used and the asymptotic framework assumed. We briefly show the key steps in extending the theoretical results of chapter 2 to the multiple frame case and reference chapter 2 whenever its results can be applied directly.

We also show how the constraints commonly used in estimation from multiple

frames, such as alignment constraints on the overlapping domain, or benchmark constraints on the size of the overlap between sampling frames, can be expressed in the empirical likelihood methodology. This is followed by some Monte Carlo simulations showing how the proposed estimator performs in various settings.

3.2 Some existing dual frame estimators

Dual and multiple frame surveys have been studied extensively and several estimators have been proposed. The standard notation used in the literature splits the population into domains defined by the sampling frames. Suppose that there are two sampling frames Q_A and Q_B, none of which has a perfect coverage. However, together both frames cover completely a population U. The following three domains can be identified in the population U: $\mathcal{D}_{A^-}$, $\mathcal{D}_{AB}$, $\mathcal{D}_{B^-}$, of sizes N_{A^-}, N_{AB} and N_{B^-} respectively, where $\mathcal{D}_{A^-} := Q_A - Q_B$, $\mathcal{D}_{B^-} := Q_B - Q_A$ and $\mathcal{D}_{AB} := Q_A \cap Q_B$.

Suppose that two samples, $\boldsymbol{S}_A$ and $\boldsymbol{S}_B$, are selected independently from frames Q_A and Q_B respectively. Suppose that we want to estimate a fixed finite population parameter $\boldsymbol{\theta}$, e.g. a total $Y = \sum_{i \in U} y_i$. Neither $\boldsymbol{S}_A$ nor $\boldsymbol{S}_B$ alone give a good estimate for $\boldsymbol{\theta}$, due to the under-coverage of the frames Q_A and Q_B. An unbiased estimator of $\boldsymbol{\theta}$ can be obtained if both samples are used together. However, any estimator has to account for the fact that the frames Q_A and Q_B overlap, that is, that some population units may be selected from more than one frame. Note that this problem occurs whether or not the samples $\boldsymbol{S}_A$ and $\boldsymbol{S}_B$ overlap, because the increased selection probabilities of the units in the overlap would cause bias if a non-adjusted estimator was used.

Figure 3.1 shows a 'spreadsheet' representation of the sampling frames. The columns represent the population domains. A separate sample is selected from each frame. The samples in the picture overlap. This overlap may or may not

occur in practice. The overlap between samples does not play any inferential role.

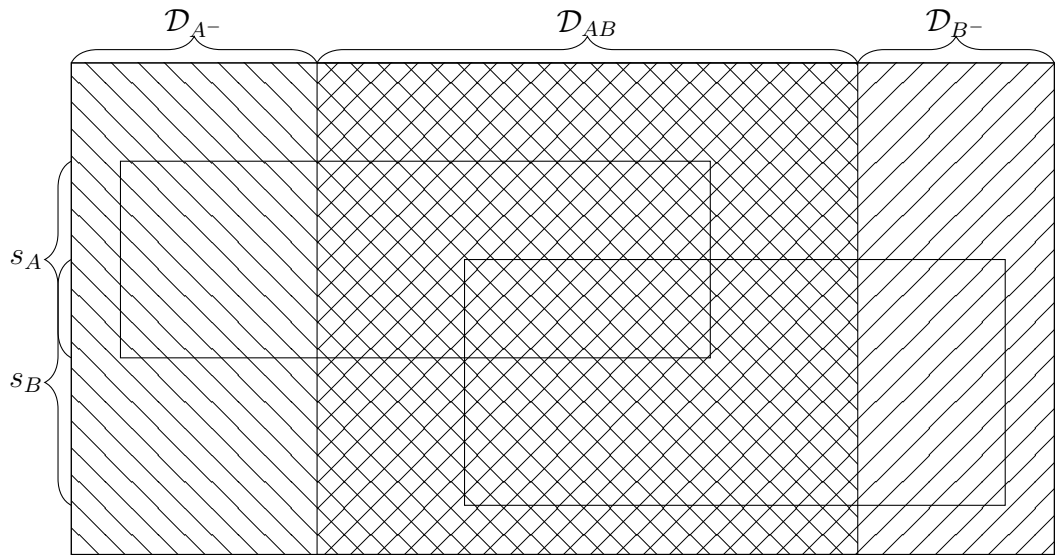


Figure 3.1: *Illustration of the sampling frames within population U and the samples selected. The horizontal axis corresponds to the population domains. The vertical axis represents the samples. The area* ▨ *represents sampling frame Q_A. The area* ▨ *represents sampling frame Q_B. The area* ▨ *represents the overlap between sampling frames Q_A and Q_B.*

The various estimators available for dual frame surveys are often divided into two groups: double (or separate) frame estimators and single (or combined) frame estimators. There are some differences in how these two terms are used in the literature. Lohr (2000) classifies all the estimators which are calculated by pooling together the units sampled from both frames and modifying weights for units belonging to the overlapping domain $\mathcal{D}_{AB}$, as single frame estimators. Singh and Mecatti (2011) use the terms 'separate' and 'combined' to reflect the level of information required for the sampled units. The separate frame estimators require that for every sampled unit the following is known: selection probability in the frame from which the unit was selected, the number of frames from which the unit could have been selected and, in some cases, identification of the frames from which the unit could have been selected. The combined frame estimators in this classification require also knowledge of the unit selection probabilities in all the relevant frames. In the review below we follow the approach that reflects the way in which estimators are calculated rather than the information required.

Therefore, we classify all the estimators that can be computed as a combination of separate domain estimators as separate frame estimators. Note that some of the estimators also require knowledge of the domain sizes N_{A^-}, N_{AB} and N_{B^-}.

3.2.1 Separate frame estimators

The first group contains estimators which can be calculated as combinations of estimators obtained separately from each of the frames for the domains $\mathcal{D}_{A^-}$, $\mathcal{D}_{AB}$, $\mathcal{D}_{B^-}$. Let $\hat{Y}_{A^-}$ denote the Horvitz and Thompson (1952) estimator of the total of the variable y for the domain $\mathcal{D}_{A^-}$, $\hat{Y}_{B^-}$ denote the Horwitz-Thompson estimator of the total of the variable y for the domain $\mathcal{D}_{B^-}$, $\hat{Y}_{AB}^{(A)}$ denote the Horvitz-Thompson estimator of the total of the variable y for the domain $\mathcal{D}_{AB}$ based on the sample selected from frame Q_A and $\hat{Y}_{AB}^{(B)}$ denote the Horvitz-Thompson estimator of the total of the variable y for the domain $\mathcal{D}_{AB}$ based on the sample selected from frame Q_B. The basic idea of separate frame estimators was formulated by Hartley (1962). Hartley's (1962) estimator of the population total of y takes the following form:

$$\hat{Y}^{HR} = \hat{Y}_{A^-} + \hat{Y}_{B^-} + \phi\hat{Y}_{AB}^{(A)} + (1-\phi)\hat{Y}_{AB}^{(B)}, \tag{3.1}$$

where the optimal coefficient ϕ is given by

$$\phi = \frac{V(\hat{Y}_{AB}^{(B)}) + Cov(\hat{Y}_{B^-}, \hat{Y}_{AB}^{(B)}) + Cov(\hat{Y}_{A^-}, \hat{Y}_{AB}^{(A)})}{V(\hat{Y}_{AB}^{(A)}) + V(\hat{Y}_{AB}^{(B)})}\,. \tag{3.2}$$

In practice the coefficient ϕ is estimated from the sample data.

A simple version of Hartley's (1962) estimator when estimation of these variances and covariances is not possible, called Simple Multiplicity (Mecatti, 2005), sets $\phi = 0.5$ for a dual frame case.

Fuller and Burmeister (1972) proposed to adjust the estimator by a regression on

the difference between the estimates of the size of the overlapping domain $\mathcal{D}_{AB}$ obtained from the two frames:

$$\hat{Y}^{FB} = \hat{Y}_{A^-} + \hat{Y}_{B^-} + \beta_1^{(FB)}\hat{Y}_{AB}^{(A)} + (1-\beta_1^{(FB)})\hat{Y}_{AB}^{(B)} + \beta_2^{(FB)}(\hat{N}_{AB}^{(A)} - \hat{N}_{AB}^{(B)}), \quad (3.3)$$

where $\hat{N}_{AB}^{(A)} = \sum_{i\in \boldsymbol{S}_A \cap \mathcal{D}_{AB}} \pi_i^{-1}$ is the estimate of N_{AB} calculated from the sample selected from frame Q_A and $\hat{N}_{AB}^{(B)} = \sum_{i\in \boldsymbol{S}_A \cap \mathcal{D}_{AB}} \pi_i^{-1}$ is the estimate of N_{AB} calculated from the sample selected from frame Q_B. The optimal coefficients are given by

$$(\beta_1^{(FB)}, \beta_2^{(FB)})^\top = \quad -\boldsymbol{\Sigma}^{-1} \left\{ \begin{array}{c} Cov(\hat{Y}_{A^-} + \hat{Y}_{B^-} + \hat{Y}_{AB}^{(B)}, \hat{Y}_{AB}^{(A)} - \hat{Y}_{AB}^{(B)}) \\ Cov(\hat{Y}_{A^-} + \hat{Y}_{B^-} + \hat{Y}_{AB}^{(B)}, \hat{N}_{AB}^{(A)} - \hat{N}_{AB}^{(B)}) \end{array} \right\}, \quad (3.4)$$

where

$$\boldsymbol{\Sigma} = \left\{ \begin{array}{cc} V(\hat{Y}_{AB}^{(A)} - \hat{Y}_{AB}^{(B)}) & Cov(\hat{Y}_{AB}^{(A)} - \hat{Y}_{AB}^{(B)}, \hat{N}_{AB}^{(A)} - \hat{N}_{AB}^{(B)}) \\ Cov(\hat{Y}_{AB}^{(A)} - \hat{Y}_{AB}^{(B)}, \hat{N}_{AB}^{(A)} - \hat{N}_{AB}^{(B)}) & V(\hat{N}_{AB}^{(A)} - \hat{N}_{AB}^{(B)}) \end{array} \right\}.$$

In practice, the coefficients $(\beta_1^{(FB)}, \beta_2^{(FB)})^\top$ are estimated from the sample data.

Estimator (3.3) was proposed for single stage simple random sampling without replacement (Arcos et al., 2015). Skinner and Rao (1996) proposed a Pseudo Maximum Likelihood estimator which can be used in complex sampling designs:

$$\begin{aligned} \hat{Y}^{SR} = \quad & \frac{N_A - \hat{N}_{AB}^{SR}(\phi^{(SR)})}{\hat{N}_A}\hat{Y}_{A^-} + \frac{N_B - \hat{N}_{AB}^{SR}(\phi^{(SR)})}{\hat{N}_B}\hat{Y}_{B^-} + \\ & + \frac{\hat{N}_{AB}^{SR}(\phi^{(SR)})}{\phi\hat{N}_{AB}^{(A)} + (1-\phi)\hat{N}_{AB}^{(B)}}\{\phi^{(SR)}\hat{Y}_{AB}^{(A)} + (1-\phi^{(SR)})\hat{Y}_{AB}^{(B)}\}, \end{aligned} \quad (3.5)$$

where $\hat{N}_{AB}^{SR}(\phi^{(SR)})$ is the smallest of the roots of the following equation:

$$\begin{aligned} 0 = \quad & \{\frac{\phi^{(SR)}}{N_B} + \frac{(1-\phi^{(SR)})}{N_A}\}^2 - \{1 + \frac{\phi N_{AB}^{(A)}}{N_B} + \frac{(1-\phi^{(SR)})N_{AB}^{(B)}}{N_A}\}x + \\ & + \phi^{(SR)}N_{AB}^{(A)} + (1-\phi^{(SR)})N_{AB}^{(B)} \end{aligned} \quad (3.6)$$

and $\phi^{(SR)} \in (0,1)$.

The optimal coefficient $\phi^{(SR)}$ is given by

$$\phi^{(SR)} = \frac{\hat{N}_A N_B V(\hat{N}_{AB}^{(B)})}{\hat{N}_A N_B V(\hat{N}_{AB}^{(B)}) + \hat{N}_B N_A V(\hat{N}_{AB}^{(A)})}. \tag{3.7}$$

3.2.2 Combined frame estimators

Kalton and Anderson (1986) and Bankier (1986) proposed estimators which operate on 'pooled' samples and adjust weights for the units belonging to the overlapping domain $\mathcal{D}_{AB}$. These estimators require that for every sampled unit, selection probabilities in all the sampling frames from which the unit could have been selected are known. Kalton and Anderson's (1986) estimator for the population total of y takes the following form:

$$\begin{aligned} \hat{Y}^{KA} = \hat{Y}_{A^-} + \hat{Y}_{B^-} &+ \sum_{i \in \boldsymbol{S}_{AB}^{(A)}} \left(\pi_i^{(A)} + \pi_i^{(B)}\right)^{-1} y_i \\ &+ \sum_{i \in \boldsymbol{S}_{AB}^{(B)}} \left(\pi_i^{(A)} + \pi_i^{(B)}\right)^{-1} y_i, \end{aligned} \tag{3.8}$$

where $\boldsymbol{S}_{AB}^{(A)}$ are the units sampled from frame Q_A which belong to domain $\mathcal{D}_{AB}$, $\boldsymbol{S}_{AB}^{(B)}$ are the units sampled from frame Q_B which belong to domain $\mathcal{D}_{AB}$ and $\pi_i^{(*)}$ is the probability that unit i is selected from frame Q_*.

Bankier's (1986) estimator is based on the observation that when the two samples are selected independently, the probability that unit i is selected in sample $\boldsymbol{S}_A$ drawn from Q_A or in sample $\boldsymbol{S}_B$ drawn from Q_B, is equal to the sum of the probabilities that unit i is selected from each of the frames minus the product of these probabilities (i.e., the probability of being selected from both frames). Thus,

the estimator takes the following form:

$$\hat{Y}^{BA} = \hat{Y}_{A^-} + \hat{Y}_{B^-} + \sum_{i \in \boldsymbol{S}_{AB}^{(A)} \cup \boldsymbol{S}_{AB}^{(B)}} \left\{ \pi_i^{(A)} + \pi_i^{(B)} - \pi_i^{(A)} \pi_i^{(B)} \right\}^{-1} y_i \cdot \tag{3.9}$$

The summation in the last term is done over all unique units selected from the domain $\mathcal{D}_{AB}$, i.e., second occurrences of the same unit are removed from estimation. Note that this requires that not only the selection probabilities from both frames are known for each unit in the sample, but also that records are linked between samples $\boldsymbol{S}_A$ and $\boldsymbol{S}_B$.

Skinner (1991) proposed the Raking Ratio estimator

$$\hat{Y}^{RR} = \frac{N_A - \hat{N}_{AB}^{Rake}}{\hat{N}_A} \hat{Y}_{A^-} + \frac{N_B - \hat{N}_{AB}^{Rake}}{\hat{N}_B} \hat{Y}_{B^-} + \frac{\hat{N}_{AB}^{Rake}}{\hat{N}_{AB}^{S}} \hat{Y}_{AB}^{S}, \tag{3.10}$$

where

$$\begin{aligned} \hat{Y}_{AB}^{S} &= \sum_{i \in \boldsymbol{S}_{AB}^{(A)}} \left(\pi_i^{(A)} + \pi_i^{(B)} \right)^{-1} y_i \\ &\quad + \sum_{i \in \boldsymbol{S}_{AB}^{(B)}} \left(\pi_i^{(A)} + \pi_i^{(B)} \right)^{-1} y_i, \end{aligned} \tag{3.11}$$

$$\hat{N}_{AB}^{S} = \sum_{i \in \boldsymbol{S}_{AB}^{(A)}} \left(\pi_i^{(A)} + \pi_i^{(B)} \right)^{-1} + \sum_{i \in \boldsymbol{S}_{AB}^{(B)}} \left(\pi_i^{(A)} + \pi_i^{(B)} \right)^{-1} \tag{3.12}$$

and $\hat{N}_{AB}^{Rake}$ is the smaller of the roots of the following equation:

$$\hat{N}_{AB}^{S} x^2 - \{ \hat{N}_{AB}^{S} (N_A + N_B) + \hat{N}_A \hat{N}_B \} x + \hat{N}_{AB}^{S} N_A N_B = 0 \cdot \tag{3.13}$$

Estimation from single frame samples is often based on calibration (Deville and Särndal, 1992*a*), where the design weights are adjusted to create calibration weights which reproduce known population level totals or means of auxiliary variables (see

chapter 2.2 for a definition of the generalized calibration property). Calibration may increase precision of the point estimates of the target variables and ensures numerical consistency between the sample estimates and the known population values. Ranalli et al. (2016) extended the calibration approach to the dual frame case. Suppose that a vector of population totals $\boldsymbol{X} = \sum_{i \in U} \boldsymbol{x}_i$ of p auxiliary variables $\boldsymbol{x}_i = (\boldsymbol{x}_{1i}, \boldsymbol{x}_{2i}, ..., \boldsymbol{x}_{pi})$, is known. The dual frame calibration estimator is defined as:

$$\hat{Y}^{(CalS)} = \sum_{i \in \boldsymbol{S}_A \cup \boldsymbol{S}_B} w_i^{cal} y_i, \tag{3.14}$$

where w_i^{cal} are chosen to minimise a defined distance function $G(w_i^{cal}, w_i^*)$ under the constraint $\sum_{i \in \boldsymbol{S}_A \cup \boldsymbol{S}_B} w_i^{cal} \boldsymbol{x}_i = \boldsymbol{X}$. The 'design' weights w_i^* differ between domains and so we have $w_i^* = (\pi_i^{(A)})^{-1}$ for $i \in \boldsymbol{S}_A \cap \mathcal{D}_{A^-}$, $w_i^* = (\pi_i^{(B)})^{-1}$ for $i \in \boldsymbol{S}_B \cap \mathcal{D}_{B^-}$ and $w_i^* = (\pi_i^{(A)} + \pi_i^{(B)})^{-1}$ for $i \in \boldsymbol{S}_{AB}^{(A)} \cup \boldsymbol{S}_{AB}^{(B)}$. When the sampling probabilities are only known for the frame from which a unit was sampled, the w_i^* are defined by $\phi(\pi_i^{(A)})^{-1}$ for $i \in \boldsymbol{S}_{AB}^{(A)}$ and by $(1-\phi)(\pi_i^{(B)})^{-1}$ for $i \in \boldsymbol{S}_{AB}^{(B)}$, with $\phi \in (0, 1)$.

Rao and Wu (2009*b*) proposed a pseudoempirical likelihood estimator for a mean in the dual frame case. The estimator for the population mean $N^{-1} \sum_{i \in U} y_i$ takes the following form:

$$\hat{\bar{y}}^{(PEL)} = \frac{N_A}{N} \hat{\bar{y}}_{A^-} + \frac{N_B}{N} \hat{\bar{y}}_{B^-} + \frac{\phi N_{AB}}{N} \hat{\bar{y}}_{AB}^{(A)} + \frac{(1-\phi) N_{AB}}{N} \hat{\bar{y}}_{AB}^{(B)}, \tag{3.15}$$

where ϕ is fixed and $\phi \in (1, 0)$, $\hat{\bar{y}}_{A^-} = \sum_{i \in \boldsymbol{S}_A \cap \mathcal{D}_{A^-}} \hat{p}_i^{(A)} y_i$, $\hat{\bar{y}}_{B^-} = \sum_{i \in \boldsymbol{S}_B \cap \mathcal{D}_{B^-}} \hat{p}_i^{(B)} y_i$, $\hat{\bar{y}}_{AB}^{(A)} = \sum_{i \in \boldsymbol{S}_{AB}^{(A)}} \hat{p}_i^{(AB;A)} y_i$, $\hat{\bar{y}}_{AB}^{(B)} = \sum_{i \in \boldsymbol{S}_{AB}^{(B)}} \hat{p}_i^{(AB;B)} y_i$ and $\hat{p}_i^{(A)}$, $\hat{p}_i^{(B)}$, $\hat{p}_i^{(AB;A)}$, $\hat{p}_i^{(AB;B)}$ are values which maximise the pseudoempirical likelihood function:

$$\begin{aligned} \ell(p_i^{(A)}, p_i^{(B)}, p_i^{(AB;A)}, p_i^{(AB;B)})^{PEL} = & \frac{N_A}{N} \sum_{i \in \boldsymbol{S}_A} \frac{(\pi_i^{(A)})^{-1}}{\sum_{i \in \boldsymbol{S}_A \cap \mathcal{D}_{A^-}} (\pi_i^{(A)})^{-1}} \log(p_i^{(A)}) \\ & + \frac{N_B}{N} \sum_{i \in \boldsymbol{S}_B} \frac{(\pi_i^{(B)})^{-1}}{\sum_{i \in \boldsymbol{S}_B \cap \mathcal{D}_{B^-}} (\pi_i^{(B)})^{-1}} \log(p_i^{(B)}) \end{aligned}$$

$$+\frac{\phi N_{AB}}{N}\sum_{i\in \boldsymbol{S}_{AB}^{(A)}}\frac{(\pi_i^{(A)})^{-1}}{\sum_{i\in \boldsymbol{S}_{AB}^{(A)}}(\pi_i^{(A)})^{-1}}\log(p_i^{(AB;A)})$$
$$+\frac{(1-\phi) N_{AB}}{N}\sum_{i\in \boldsymbol{S}_{AB}^{(B)}}\frac{(\pi_i^{(B)})^{-1}}{\sum_{i\in \boldsymbol{S}_{AB}^{(B)}}(\pi_i^{(B)})^{-1}}\log(p_i^{(AB;B)}) \tag{3.16}$$

under the constraints: $\sum_{i\in \boldsymbol{S}_A\cap \mathcal{D}_{A^-}} p_i^{(A)} = 1$, $\sum_{i\in \boldsymbol{S}_B\cap \mathcal{D}_{B^-}} p_i^{(B)} = 1$, $\sum_{i\in \boldsymbol{S}_{AB}^{(A)}} p_i^{(AB;A)} = 1$, $\sum_{i\in \boldsymbol{S}_{AB}^{(B)}} p_i^{(AB;B)} = 1$ and an alignment constraint on the estimate for the mean of a variable measured in the overlapping domain $\sum_{i\in \boldsymbol{S}_{AB}^{(A)}} p_i^{(AB;A)} y_i = \sum_{i\in \boldsymbol{S}_{AB}^{(B)}} p_i^{(AB;B)} y_i$. Additional benchmark constraints can also be defined.

3.3 Multiplicity estimation

The notation used in chapter (3.2) becomes complicated when more than two frames are used. Singh and Mecatti (2011) proposed the general form of the multiplicity adjusted estimator, which naturally accommodates inference from multiple frames. We briefly characterise this approach below.

Consider T sampling frames Q_t, $t \geq 2$, which together cover the entire population U. T samples $(\boldsymbol{S}_1, \boldsymbol{S}_2, ...\boldsymbol{S}_T)$ of sizes $(n_1, n_2, ...n_T)$ respectively, are selected independently, where $\boldsymbol{S}_t$ denotes the sample selected from frame Q_t. Let π_{ti} be the probability of selecting unit i from frame t. Note that the sampling frames usually overlap. The extent of the overlap may be unknown.

Let $\boldsymbol{S}$ of size $n = \sum_{t=1}^{T} n_t$ be the collection of labels of all the units selected in all the T samples. If a unit is selected k times, its label appears k times in $\boldsymbol{S}$. That is, the notation $\sum_{i\in \boldsymbol{S}}(\cdot)$ is equivalent to $\sum_{t=1}^{T}\sum_{i\in \boldsymbol{S}_t}(\cdot)$.

Suppose that the values of the set of variables $\boldsymbol{v} := \{\boldsymbol{y}, \boldsymbol{x}\}$ are collected for every unit in each of the samples $\boldsymbol{S}_1, \boldsymbol{S}_2, ..., \boldsymbol{S}_T$. The variable $\boldsymbol{y}$ is the variable of in-

terest.The vector $\boldsymbol{x}$ contains auxiliary variables for which a vector of population parameters $\boldsymbol{\varphi}_U$ is known. The parameter $\boldsymbol{\varphi}_U$ is defined as the vector of the unique solutions of the population estimating equation:

$$\sum_{i \in U} \boldsymbol{f}_i(\boldsymbol{x}_i, \boldsymbol{\varphi}_U) = \mathbf{0}_q, \tag{3.17}$$

where q is the dimension of vector $\boldsymbol{f}_i(\boldsymbol{x}_i, \boldsymbol{\varphi}_U)$.

Let $\kappa_{t;i}$ be the frame inclusion indicator, which is equal to 1 if the frame Q_t contains the i-th unit and to 0 otherwise. We assume that for every $i \in \boldsymbol{S}$, i.e., for every sampled unit i, the value of the multiplicity-adjusted selection probability ρ_i,

$$\rho_i = \pi_{t;i} \alpha_{t;i}^{-1} \tag{3.18}$$

is known, where $\alpha_{t;i}$ are the multiplicity adjustment factors such that, for all $i \in \boldsymbol{S}$, (Singh and Mecatti, 2011),

$$\sum_{t=1}^{T} \kappa_{t;i} \alpha_{t;i} = 1 \cdot \tag{3.19}$$

Condition (3.19) was first proposed by Birnbaum and Sirken (1965) for estimators based on samples obtained through network sampling.

The Generalized multiplicity-adjusted Horwitz-Thompson estimator (GMHT) (Singh and Mecatti, 2011; Rao and Wu, 2009*b*) of the population total $Y = \sum_{i \in U} y_i$, takes the following form:

$$\hat{Y}^{GMHT} = \sum_{t=1}^{T} \sum_{i \in \boldsymbol{S}_t} y_i \alpha_{t;i} (\pi_{t;i})^{-1} \cdot \tag{3.20}$$

In the simplest case, $\alpha_{t;i} = {M_i}^{-1}$, where $M_i = \sum_{t=1}^{T} \kappa_{t;i}$ is the multiplicity indicator equal to the number of sampling frames which contain unit i. This creates the Simple Multiplicity (Mecatti, 2005) estimator. Various choices of $\alpha_{t;i}$ are possible, depending on the information available on frame membership and sample inclusion

probabilities for the sampled units. In particular, all the separate frame and combined frame estimators discussed in Chapter 3.2, with the exception of Rao and Wu's (2009*b*) estimator, which was not discussed by the authors, can be obtained through re-formulation of $\alpha_{t;i}$ (Singh and Mecatti, 2011).

Singh and Mecatti (2011) proposed also the so-called Composite Multiplicity Adjusted estimator, which is a composite of the Simple Multiplicity (Mecatti, 2005) and Kalton and Anderson's (1986) estimators. The following adjustment is used:

$$\alpha_{t;i}^{(CM)} = \gamma_i M_i^{-1} + (1 - \gamma_i)\, \pi_{t;i} \left(\sum_{t=1}^{T} \pi_{t;i} \kappa_{t;i} \right)^{-1}, \tag{3.21}$$

where γ_i is chosen to minimize the variance of (3.20).

An estimator allowing for different adjustment factors $\alpha_{t;i}$ depending on the level of information available for each unit i, called Hybrid Multiplicity, was also proposed. For example, suppose that Kalton and Anderson's (1986) adjustment is used for units for which sampling probabilities from each of the frames are known and the Simple Multiplicity adjustment is used for the units for which only the frame count and the sampling probabilities for the frame from which the unit was actually selected are known. The adjustment factor takes the following form:

$$\alpha_{t;i}^{(HM)} = \alpha_{t;i}^{(SM)}(1 - \delta_{t;i}^{FULL}) + \delta_{t;i}^{FULL} \alpha_{t;i}^{(KA)}, \tag{3.22}$$

where $\delta_{t;i}^{FULL}$ equals 1 if sampling probabilities from each of the frames are known and 0 otherwise (Singh and Mecatti, 2014).

The idea of the multiplicity adjustment has also been applied to create calibration estimators for multiple frame samples. In particular, the Generalized Multiplicity-adjusted Regression Estimator (GMREG) for the total of y takes the following form (Ranalli et al., 2016; Singh and Mecatti, 2014):

$$\hat{Y}^{GMREG} = \hat{Y}^{GMHT} - \hat{\boldsymbol{\beta}}^{\top} \left(\hat{\boldsymbol{X}}^{GMHT} - \boldsymbol{X} \right) \tag{3.23}$$

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{x}^\top \boldsymbol{\Lambda} \boldsymbol{x}\right)^{-1} \boldsymbol{x}^\top \boldsymbol{\Lambda} \boldsymbol{y}, \tag{3.24}$$

where $\boldsymbol{\Lambda}$ is a p-by-p weighting matrix equal to $diag(\boldsymbol{d})$, where $\boldsymbol{d}$ is the vector of inverted multiplicity adjusted selection probabilities, $\boldsymbol{d} = (\rho_1^{-1}, \rho_2^{-1}, ..., \rho_n^{-1})$ and $\boldsymbol{X}$ is the known population total of the auxiliary variable x.

Singh and Mecatti (2014) propose a modified version of the calibration estimator, based on the Generalized Raking Estimator With Optimal Unbiased Modification (GROUM) (Singh and Wu, 2003; Singh et al., 2013). The GROUM estimator uses relative design adjustment factors to account for different sampling designs in strata. This results in increased precision in stratified designs compared to a simple calibration (Deville and Särndal, 1992a) estimator, due to inclusion of the stratum-specific relative design adjustment factors (see Singh et al. (2013) and Singh and Mecatti (2014) for details). Singh and Mecatti (2014) use the same method to account for different sampling designs applied in multiple frames.

The GROUM estimator for the population total of y takes the following form:

$$\hat{Y}^{GROUM} = \hat{Y}^{GMHT} - \hat{\boldsymbol{\beta}}_{GROUM}^\top \left(\hat{\boldsymbol{X}}_+^{GMHT} - \boldsymbol{X}_+\right) \tag{3.25}$$

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{x}^\top \boldsymbol{\gamma} \boldsymbol{\delta} \boldsymbol{x}\right)^{-1} \boldsymbol{x}^\top \boldsymbol{\gamma} \boldsymbol{\delta} \boldsymbol{\lambda} \boldsymbol{y}, \tag{3.26}$$

where $\boldsymbol{\delta}$ is a diagonal matrix of inverted sampling probabilities, $\boldsymbol{\gamma}$ is a matrix of inverse relative design adjustment factors, common for units selected in the same sample, and $\boldsymbol{\lambda}$ is a diagonal matrix of multiplicity adjustments. The vector of calibration totals $\hat{\boldsymbol{X}}_+^{GMHT}$ and $\boldsymbol{X}_+$ include, apart from the auxiliary variables $\boldsymbol{x}$, population counts and totals of y for the overlapping domain.

The Pseudo Empirical Likelihood estimator of Rao and Wu (2009b) has also been expressed in a multiplicity-inspired form. The pseudo empirical likelihood function for T sampling frames takes the following form:

$$\ell(\boldsymbol{p})^{PEL} = \frac{n}{\hat{N}} \sum_{t=1}^{T} \sum_{i \in \boldsymbol{S}_t} \pi_{t;i}^{-1} \alpha_{t;i} \log(p_i), \tag{3.27}$$

where $\alpha_{t;i} = \alpha_{t;i}^{SM}$, $n = \sum_{t=1}^{T} n_t$ and $\hat{N} = \sum_{t=1}^{T} \sum_{i \in \boldsymbol{S}_t} \pi_{t;i}^{-1} p_i$. The pseudo empirical likelihood adjusted weights $\hat{p}_i$ are obtained as the values which maximise (3.27) under $\sum_{t=1}^{T} \sum_{i \in \boldsymbol{S}_t} p_i = 1$. Additional constraints on known population totals of auxiliary variables may be added. Confidence intervals are obtained through a χ^2 approximation of the pseudo empirical likelihood ratio function corrected by the design effect (Rao and Wu, 2009*b*).

3.4 Empirical likelihood approach

The following paragraphs introduce a novel Empirical Likelihood approach to multiple frame estimation. This approach is based on the multiplicity adjustment method and is applicable to multiple frames. The frames may use different sampling designs. As in chapter 2, we follow the design based approach where the only source of randomness is in sampling and the parameters are fixed quantities (Neyman, 1934). We first introduce the method for a simple case when no stratification is used. In Chapter 3.5 we show how stratification and domain-specific auxiliary information, as well as constraints on the overlapping domain, can be handled.

Although the proposed estimator may seem similar to the Multiplicity Pseudo Empirical Likelihood estimator of Rao and Wu (2009*b*), it is in fact quite different. In the Multiplicity Pseudo Empirical Likelihood estimator, the adjustment factors are included in the likelihood function, while the proposed estimator handles them differently. The proposed approach to obtaining confidence intervals is also different and no estimation of design effects is involved. On the practical side, the proposed estimator can include various adjustment factors and can be used to estimate a wide range of parameters, while the Multiplicity Pseudo Empirical Likelihood estimator has been specifically defined for means and is based on the simple multiplicity adjustment factors $\alpha_{t;i}^{SM}$.

Suppose that we wish to estimate a fixed, unknown population parameter $\boldsymbol{\theta}_U$, a function of a subset of $\boldsymbol{v} = \{\boldsymbol{x}, \boldsymbol{y}\}$. The parameter $\boldsymbol{\theta}_U$ is defined by the unique solution of the population estimating equation

$$\sum_{i \in U} \boldsymbol{g}_i(\boldsymbol{v}_i, \boldsymbol{\theta}_U) = \mathbf{0}_\nu, \tag{3.28}$$

where p is the dimension of vector $\boldsymbol{g}_i(\boldsymbol{v}_i, \boldsymbol{\theta}_U)$. This general formulation allows for estimation of a wide range of parameters, e.g. means, quantiles, rates, ratios, regression parameters. In particular, if $\boldsymbol{g}_i(\boldsymbol{v}_i, \boldsymbol{\theta}_U) = \boldsymbol{v}_i - \boldsymbol{\theta}_U$, the parameter $\boldsymbol{\theta}_U$ is a vector of population means. We will use $\boldsymbol{g}_i(\boldsymbol{\theta}_U)$ to denote $\boldsymbol{g}_i(\boldsymbol{v}_i, \boldsymbol{\theta}_U)$ in the following text for brevity.

Consider the following joint empirical log-likelihood function of unknown scale loads m_i:

$$\ell(\boldsymbol{m}) = \sum_{t=1}^{T} \sum_{i \in \boldsymbol{S}_t} \log(m_i) \cdot \tag{3.29}$$

Note that the function (3.29) is defined at the sample level, for scale loads associated with units selected in each of the T samples. If a unit is selected p times, its scale load m_i is considered p times.

Consider the following system of constraints:

1. *Unknown parameters' constraint*:

$$\sum_{t=1}^{T} \sum_{i \in \boldsymbol{S}_t} m_i \boldsymbol{g}_i(\boldsymbol{\theta}) = \mathbf{0}_\nu, \tag{3.30}$$

   where $\boldsymbol{\theta}$ is a vector in the parameter space of the population parameter of interest $\boldsymbol{\theta}_U$ and p is the dimension of vector $\boldsymbol{g}_i(\boldsymbol{\theta})$;

2. *Sample size constraint*:

$$\sum_{i \in \boldsymbol{S}_t} m_i \rho_i = n_t, \;\; t = 1, 2, ..., T, \tag{3.31}$$

where ρ_i is the multiplicity adjusted selection probability defined by (3.18);

3. *Known parameters' constraint*: (Owen, 1991; Chaudhuri et al., 2008; Lesage, 2011);

$$\sum_{t=1}^{T} \sum_{i \in \boldsymbol{S}_t} m_i \boldsymbol{f}_i(\boldsymbol{x}_i, \boldsymbol{\varphi}_U) = \mathbf{0}_q, \tag{3.32}$$

where q is the dimension of vector $\boldsymbol{f}_i(\boldsymbol{x}_i, \boldsymbol{\varphi}_U)$, $\boldsymbol{\varphi}_U$ is a vector of known population (census) parameters $\varphi_1, \varphi_2, ..., \varphi_r$ and $\boldsymbol{f}_i(\boldsymbol{x}_i, \boldsymbol{\varphi}_U)$ is a function such that

$$\sum_{i \in U} \boldsymbol{f}_i(\boldsymbol{x}_i, \boldsymbol{\varphi}_U) = \mathbf{0}_q. \tag{3.33}$$

Constraint (3.30) involves the unknown population parameter of interest and is key in obtaining point estimates and confidence intervals. Constraint (3.31) is a multiple frames generalisation of a constraint defined for a single sample empirical likelihood estimator by Berger and De La Riva Torres (2016). This constraint ensures that the multiplicity adjusted empirical likelihood point estimator is design-consistent. Constraint (3.32) is optional. It is a generalisation of the customary calibration or benchmark type constraint (see e.g. Deville and Särndal (1992*a*)) defined on the known population parameters. This type of constraints is commonly used in survey inference, where the known population parameter vector $\boldsymbol{\varphi}_U$ consists of counts, means or totals known usually from censuses or administrative records. For example, $\boldsymbol{\varphi}_U$ may be a vector of known sizes of k age-sex groups. We can then define the function $\boldsymbol{f}_i(\boldsymbol{x}_i, \boldsymbol{\varphi}_U) = \boldsymbol{x}_i - \boldsymbol{\varphi}_U$ where the k-th element of vector $\boldsymbol{x}_i$, $x_{i;k}$, is a group membership indicator and equals 1 if the unit i belongs to the age-sex group k. The parameter $\boldsymbol{\varphi}_U$ is considered to be known, i.e., measured without error at the population level. Note that in order to apply constraint (3.32) we only need to know the unit level values $\boldsymbol{x}_i$ for the sampled units. Benchmark constraints are discussed in more detail in chapters 1 and 2. We will use $\mathbf{f}_i(\boldsymbol{\varphi}_U)$ in the following text to denote $\boldsymbol{f}_i(\boldsymbol{x}_i, \boldsymbol{\varphi}_U) = \boldsymbol{x}_i - \boldsymbol{\varphi}_U$ for brevity.

The system of constraints (3.30)-(3.32) can be written as

$$\sum_{i \in \boldsymbol{S}_t} m_i \boldsymbol{c}_i^\star(\boldsymbol{\theta}) = \boldsymbol{C}^\star, \tag{3.34}$$

where

$$\boldsymbol{c}_i^\star(\boldsymbol{\theta}) \;=\; \{\, \boldsymbol{c}_i^\top \,, \boldsymbol{g}_i(\boldsymbol{\theta})^\top \}^\top, \tag{3.35}$$

$$\boldsymbol{C}^\star \;=\; \left(\boldsymbol{C}^\top, \mathbf{0}_\nu^\top\right)^\top, \tag{3.36}$$

$$\boldsymbol{c}_i \;=\; \{\boldsymbol{d}_i^\top, \mathbf{f}_i(\boldsymbol{\varphi}_{U})^\top\}^\top, \tag{3.37}$$

$$\boldsymbol{C} \;=\; \left(\boldsymbol{D}^\top, \mathbf{0}_q^\top\right)^\top, \tag{3.38}$$

$\boldsymbol{d}_i$ and $\boldsymbol{D}$ are vectors of dimension T, with the t-th elements defined respectively by $d_{i;t} = \delta_{t;i}\, \rho_i$ and $D_t = n_t$, with

$$\delta_{ti} = \begin{cases} 1 & \text{if } i \in \boldsymbol{S}_t, \\ 0 & \text{otherwise.} \end{cases}$$

We assume that $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}_{U}$ are such that $\boldsymbol{C}^\star$ is an inner point of the convex hull formed by the sample observations $\{\boldsymbol{c}_i^\star(\boldsymbol{\theta}) : i \in \boldsymbol{S}\}$.

3.4.1 Maximum empirical likelihood point estimator

Let $\{\widehat{m}_i^\star(\boldsymbol{\theta}, \boldsymbol{\varphi}_{U}) : i \in \boldsymbol{S}\}$ be the vector of values which maximise the function (3.29), for a given vector $\boldsymbol{\theta}$, under $m_i > 0$ and (3.34). That is, let the maximum value of the joint empirical log-likelihood function (3.29) for a given vector $\boldsymbol{\theta}$, under

$m_i > 0$ and (3.34), be

$$\ell(\boldsymbol{\theta}|\boldsymbol{\varphi}_U) = \sum_{t=1}^{T} \sum_{i \in \boldsymbol{S}_t} \log\{\widehat{m}_i^{\star}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)\}\cdot \tag{3.39}$$

Following an argument presented by Berger and De La Riva Torres (2016), it can be shown that the vector $\{\widehat{m}_i^{\star}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U) : i \in \boldsymbol{S}\}$ is given by

$$\widehat{m}_i^{\star}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U) = \left\{\rho_i + \boldsymbol{\eta}^{\star\top} \boldsymbol{c}_i^{\star}(\boldsymbol{\theta})\right\}^{-1}, \tag{3.40}$$

where $\boldsymbol{\eta}^{\star}$ is a vector of Lagrange's multipliers, such that constraint (3.34) holds.

The maximum empirical likelihood point estimator of $\boldsymbol{\theta}_U$ is the vector $\widehat{\boldsymbol{\theta}}$ in the parameter space of $\boldsymbol{\theta}_U$ which maximises $\ell(\boldsymbol{\theta}|\boldsymbol{\varphi}_U)$ as defined by (3.39). We will call $\widehat{\boldsymbol{\theta}}$ *multiplicity adjusted empirical likelihood estimator*. Using an argument similar to that presented by Berger and De La Riva Torres (2016), recalled in chapter (2), it can be shown that if $\boldsymbol{\theta}_U$ is uniquely defined by the estimating equation (3.28), the estimator $\widehat{\boldsymbol{\theta}}$ is the unique solution of the sample estimating equation:

$$\hat{\boldsymbol{G}}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \sum_{i \in \boldsymbol{S}_t} \hat{m}_i \boldsymbol{g}_i(\boldsymbol{\theta}) = \mathbf{0}_{\nu}, \tag{3.41}$$

where the vector $\{\widehat{m}_i(\boldsymbol{\varphi}_U) : i \in \boldsymbol{S}\}$ maximises function (3.29) under the constraint

$$\sum_{t=1}^{T} \sum_{i \in \boldsymbol{S}_t} \hat{m}_i \boldsymbol{c}_i = \boldsymbol{C}, \;\; \hat{m}_i > 0 \tag{3.42}$$

and $\boldsymbol{c}_i$ and $\boldsymbol{C}$ are respectively defined by expressions (3.37) and (3.38).

In order to obtain the point estimate of $\boldsymbol{\theta}_U$ we need to find the values $\{\widehat{m}_i(\boldsymbol{\varphi}_U) : i \in \boldsymbol{S}\}$, which do not depend on $\widehat{\boldsymbol{\theta}}$, and solve the sample level estimating equation (3.41) for $\boldsymbol{\theta}$. By analogy with (3.40), the $\widehat{m}_i(\boldsymbol{\varphi}_U)$ are equal to

$$\widehat{m}_i(\boldsymbol{\varphi}_U) = \left\{\rho_i + \boldsymbol{\eta}^{\top} \boldsymbol{c}_i\right\}^{-1}, \tag{3.43}$$

where $\boldsymbol{\eta}$ is a vector of Lagrange's multipliers such that (3.42) holds.

The numerical aspects of estimating the $\widehat{m}_i(\boldsymbol{\varphi}_U)$ are discussed in chapter 5.

3.4.2 Relationship to a generalized regression type estimator

Below we establish the relationship between the multiplicty adjusted empirical likelihood estimator $\widehat{\boldsymbol{\theta}}$, defined as the solution of (3.41), and a generalized regression type estimator.

Let $\boldsymbol{S}$ be a collection of labels of units selected in all t samples $\boldsymbol{S}_t$, $t = 1, 2, ..., T$ and n be the size of the pooled sample, $n = \sum_{t=1}^{T} n_t$. Consider the following regularity conditions:

$$\max_{i\in\boldsymbol{S}}\left\{\frac{N}{n}\rho_i\right\} = O_{\mathcal{P}}(1) \quad \text{and} \quad \max_{i\in\boldsymbol{S}}\left\{\frac{n}{N}\rho_i^{-1}\right\} = O_{\mathcal{P}}(1), \tag{3.44}$$

$$N^{-1}\|\widehat{\boldsymbol{C}}_\pi - \boldsymbol{C}\| = \boldsymbol{\mathcal{O}}_{\mathcal{P}}(n^{-1/2}), \tag{3.45}$$

$$\max\{\|\boldsymbol{c}_i\| : i \in \boldsymbol{S}\} = o_{\mathcal{P}}(n^{1/2}), \tag{3.46}$$

$$\|\widehat{\boldsymbol{S}}\| = \boldsymbol{\mathcal{O}}_{\mathcal{P}}(1), \tag{3.47}$$

$$\|\widehat{\boldsymbol{S}}\|^{-1} = \boldsymbol{\mathcal{O}}_{\mathcal{P}}(1), \tag{3.48}$$

$$\frac{n^{\tau-1}}{N^\tau}\sum_{i\in\boldsymbol{S}}\frac{\|\boldsymbol{c}_i\|^\tau}{\rho_i^\tau} = O_{\mathcal{P}}(1) \quad (\tau = 2,\ 3,\ 4), \tag{3.49}$$

where

$$\hat{\boldsymbol{S}} = -\frac{n}{N^2}\sum_{t=1}^{T}\sum_{i\in\boldsymbol{S}_t}\frac{\boldsymbol{c_i}\boldsymbol{c_i}^\top}{\rho_i^2}. \tag{3.50}$$

Conditions (3.44) - (3.49) resemble the conditions presented in chapter 2.6.1. The only difference is that conditions (3.44) - (3.49) include the multiplicity adjusted

selection probabilities ρ_i rather than the design probabilities π_i. When $\alpha_{t;i}$ are fixed quantities, e.g. when the simple multiplicity (Mecatti, 2005) adjustment is used, conditions (3.44) - (3.49) and (2.70) - (2.75) are equivalent. When $\alpha_{t;i}$ are random quantities, the additional assumption

$$\max\{\alpha_{t;i}^{\tau} : i \in \boldsymbol{S}\} = O_{\mathcal{P}}(1), \quad (\tau = -1,\ 1,\ 2,\ 3,\ 4), \tag{3.51}$$

together with conditions (2.70) - (2.75) would usually imply conditions (3.44)-(3.49).

Theorem 4. *Under the assumptions* (3.44)-(3.49)*, for all* $\boldsymbol{\theta}$ *which are such that:*

$$\frac{1}{nN^2}\sum_{i\in\boldsymbol{S}}\frac{||\boldsymbol{g}_i(\boldsymbol{\theta})||^2}{\rho_i^2} = O_{\mathcal{P}}(n^{-2}), \tag{3.52}$$

the maximum empirical likelihood point estimator $\widehat{\boldsymbol{\theta}}$ *is asymptotically equivalent to a Generalized Regression Estimator* $\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta})$*:*

$$\hat{\boldsymbol{G}}(\boldsymbol{\theta}) = \widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}) + o_{\mathcal{P}}(Nn^{-\frac{1}{2}}), \tag{3.53}$$

where

$$\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}) = \widehat{\boldsymbol{G}}_{\pi}(\boldsymbol{\theta}) + \widehat{\boldsymbol{B}}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)^{\top}\left(\boldsymbol{C} - \hat{\boldsymbol{C}}_{\pi}\right), \tag{3.54}$$

$$\widehat{\boldsymbol{G}}_{\pi}(\boldsymbol{\theta}) = \sum_{t=1}^{T}\sum_{i\in\boldsymbol{S}_t}\frac{\boldsymbol{g}_i(\boldsymbol{\theta})}{\rho_i}, \tag{3.55}$$

$$\widehat{\boldsymbol{B}}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U) = \left(\sum_{t=1}^{T}\sum_{i\in\boldsymbol{S}_t}\frac{\boldsymbol{c}_i\boldsymbol{c}_i^{\top}}{\rho_i^2}\right)^{-1}\left(\sum_{t=1}^{T}\sum_{i\in\boldsymbol{S}_t}\frac{\boldsymbol{c}_i\boldsymbol{g}_i(\boldsymbol{\theta})^{\top}}{\rho_i^2}\right), \tag{3.56}$$

$$\hat{\boldsymbol{C}}_{\pi} = \sum_{t=1}^{T}\sum_{i\in\boldsymbol{S}_t}\frac{\boldsymbol{c}_i}{\rho_i}. \tag{3.57}$$

Note that (3.54) has the same structure as the estimator (2.89) in chapter 2. The difference is in the use of the multiplicity adjusted selection probabilities ρ_i rather than the design probabilities π_i.

The proof of theorem 4 is presented in the Appendix.

3.4.3 Asymptotic design-consistency of the empirical likelihood multiplicity adjusted estimator

In this chapter we define the asymptotic framework and show that $\widehat{\boldsymbol{\theta}}$ is asymptotically $\sqrt{n}$ design-consistent.

In order to accommodate the multiple frames settings we need to adjust the asymptotic framework discussed in chapter 2.6.3. We consider a sequence of nested populations $U^{(\nu)}$ of size $N^{(\nu)}$, where $\nu = 1, 2, \ldots, \infty$ (Isaki and Fuller, 1982). Each population $U^{(\nu)}$ consists of T sampling frames $U_t^{(\nu)}$ of sizes $N_t^{(\nu)}$ respectively, where T is a constant. A sequence of samples $\boldsymbol{S}_t^{(\nu)}$ of size $n_t^{(\nu)} \leqslant N_t^{(\nu)}$ is selected from $U_t^{(\nu)}$ according to a sampling design $\mathcal{P}_t^{(\nu)}(\boldsymbol{S}_t)$, respectively. We assume that $n_t^{(\nu)} \to \infty$ as $N^{(\nu)} \to \infty$. Note that this implies that also $N_t^{(\nu)} \to \infty$. A similar asymptotic framework is adopted for the multiple frames scenario e.g. by Singh and Mecatti (2011).

Suppose that $\boldsymbol{\theta}_{U}$ is such that the following conditions hold:

$$\widehat{\boldsymbol{G}}_{\pi}(\boldsymbol{\theta}_{U}) = \boldsymbol{\mathcal{O}}_{\mathcal{P}}(Nn^{-1/2}), \tag{3.58}$$

$$\frac{n^{\tau-1}}{N^{\tau}} \sum_{i \in \boldsymbol{S}} \frac{\|\boldsymbol{g}_i(\boldsymbol{\theta}_{U})\|^{\tau}}{\rho_i^{\tau}} = O_{\mathcal{P}}(1) \quad (\tau = 2,\ 3,\ 4), \tag{3.59}$$

$$\widehat{\boldsymbol{\nabla}}(\boldsymbol{\theta}) := \frac{1}{N}\frac{\partial \widehat{\boldsymbol{G}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \quad \text{is continuous in } \boldsymbol{\theta} \in \boldsymbol{\Theta}_{U}, \tag{3.60}$$

$$\frac{1}{N}\|\frac{\partial \widehat{\boldsymbol{\nabla}}(\boldsymbol{\theta})_k}{\partial \boldsymbol{\theta}}\| = \boldsymbol{\mathcal{O}}_{\mathcal{P}}(1) \quad \text{uniformly for all } \boldsymbol{\theta} \in \boldsymbol{\Theta}_{U}, \tag{3.61}$$

$$\|\widehat{\boldsymbol{\nabla}}(\boldsymbol{\theta}_{U})\| \asymp_p 1, \tag{3.62}$$

$$|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_{U}| = o_{\mathcal{P}}(1),, \tag{3.63}$$

where $\widehat{\boldsymbol{\nabla}}(\boldsymbol{\theta})_k$ is the $k-th$ row of matrix $\widehat{\boldsymbol{\nabla}}(\boldsymbol{\theta})$, $k = 1, 2, ..., K$; K is the number

of rows in matrix $\widehat{\boldsymbol{\nabla}}(\boldsymbol{\theta})$,

$$\widehat{\boldsymbol{G}}_{\pi}(\boldsymbol{\theta}_{U}) \;:=\; \sum_{i\in\boldsymbol{S}} \frac{1}{\rho_i}\boldsymbol{g}_i(\boldsymbol{\theta}_{U}), \tag{3.64}$$

$n = \sum_{t=1}^{T} n_t$ and $\boldsymbol{\Theta}_{U}$ is a compact neighbourhood containing $\boldsymbol{\theta}_{U}$.

Note that these assumptions are analogous to regularity conditions(2.79) - (2.84). A discussion of these conditions can be found in chapter 2.6.1.

Theorem 5 establishes the rate of convergence for the multiplicity adjusted empirical likelihood estimator $\widehat{\boldsymbol{\theta}}$.

Theorem 5. *Let* $n = \sum_{t=1}^{T} n_t$. *Under the regularity conditions* (3.44)-(3.49), (3.58), (3.59) *(with* $\tau = 2$*),* (3.60)-(3.63)*, we have* $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_{U} = \boldsymbol{\mathcal{O}}_{\mathcal{P}}(n^{-1/2})$.

The proof is presented in the Appendix.

3.4.4 Empirical Likelihood confidence intervals

In this chapter we define the empirical log-likelihood ratio function for the multiplicity adjusted empirical likelihood estimator. We also show that the asymptotic distribution of this function can be established in an analogous way to how it was done for the log-likelihood ratio function of the aligned empirical likelihood estimator discussed in chapter 2.

Let $\ell(\hat{\boldsymbol{m}})$ be defined by (3.29) with $\hat{\boldsymbol{m}}$ being the values which maximise (3.29) under the constraint (3.42). Let $\ell(\boldsymbol{\theta}|\boldsymbol{\varphi}_{U})$ be defined by (3.39) and let $\boldsymbol{\theta}$ be a vector in the parameter space of $\boldsymbol{\theta}_{U}$.

Consider the following empirical likelihood ratio function:

$$\widehat{r}(\boldsymbol{\theta}|\boldsymbol{\varphi}_{U}) = 2\{\ell(\hat{\boldsymbol{m}}) - \ell(\boldsymbol{\theta}|\boldsymbol{\varphi}_{U})\}\cdot \tag{3.65}$$

Suppose that the following regularity conditions hold:

$$\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta})^\top \left[\widehat{\boldsymbol{V}}_{\mathcal{P}}\{\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_U)\}\right]^{-\frac{1}{2}} \xrightarrow{d} \boldsymbol{\mathcal{N}}(\mathbf{0}_\nu, \boldsymbol{I}_p), \tag{3.66}$$

$$max(||\boldsymbol{g}_i(\boldsymbol{\theta}_U)|| : i \in \boldsymbol{S}) = o_{\mathcal{P}}(n^{\frac{1}{2}}), \tag{3.67}$$

$$\frac{n^{\tau-1}}{N^\tau} \sum_{i \in \boldsymbol{S}} \frac{\|\boldsymbol{g}_i(\boldsymbol{\theta}_U)\|^\tau}{\rho_i^\tau} = O_{\mathcal{P}}(1) \quad (\tau = 2,\ 3,\ 4)\cdot \tag{3.68}$$

Theorem 6. *Under the assumptions* (3.44)-(3.49) *and* (3.66)-(3.68),

$$\widehat{r}(\boldsymbol{\theta}_U|\boldsymbol{\varphi}_U) = \widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_U)^\top \left[\widehat{\boldsymbol{V}}_{\mathcal{P}}\{\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_U)\}\right]^{-1} \widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_U) + O_{\mathcal{P}}(n^{-1/2}) \tag{3.69}$$

where

$$\widehat{\boldsymbol{V}}_{\mathcal{P}}\{\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_U)\} = \sum_{t=1}^{T} \sum_{i \in \boldsymbol{S}_t} \frac{\{\boldsymbol{g}_i(\boldsymbol{\theta}_U) - \widehat{\boldsymbol{B}}(\boldsymbol{\theta}_U, \boldsymbol{\varphi}_U)^\top \boldsymbol{c}_i\}^2}{\rho_i^2}\cdot \tag{3.70}$$

The proof is presented in the Appendix. Following analogous reasoning to that presented in chapter 2.8, theorem 6 implies that (3.65) is a pivotal statistic and follows a $\chi^2_{df=d}$ distribution asymptotically, with d being the dimension of $\boldsymbol{\theta}_U$. This property can be used to construct Wilks (1938) type confidence intervals and to test hypothesis about the parameter $\boldsymbol{\theta}_U$ in an analogous way as presented in chapter 2.9.

3.5 Extensions

In this chapter we extend the proposed multiplicity adjusted empirical likelihood estimator to accommodate some more complex estimation scenarios. First, in chapter 3.5.1, we show how the estimator may be applied to inference from stratified samples. This is similar to how stratification is handled in chapter 2 and also follows the lines of the approach proposed by Berger and De La Riva Torres (2016).

In chapter 3.5.2 we show how benchmark constraints defined at a domain level can be included in the constraints system. Domain-level constraints are often used in multiple frame context. In particular, benchmarking on the known frame sizes, or known size of the overlap between frames, is used to increase precision of the estimates of the target parameters. A thorough discussion of the efficiency gains related to use of benchmark constraints on sizes of various domains was presented by Ranalli et al. (2016).

Finally, in chapter 3.5.3 we show how alignment constraints on the overlapping domain can be defined. Including an alignment constraint on the estimates obtained for the overlapping domain might improve the precision of the target parameter estimates. For example, it is often the case that the size of the overlap between sampling frames is unknown. This is because it is typically easier to obtain frame membership information for the sampled units only than to cross-reference complete frames, e.g. through record linkage. An alignment constraint on the estimates of the size of the overlap between frames is often imposed and has been found to increase precision of the target parameter estimates (Ranalli et al., 2016). We show how such a constraint may be included in the multiplicity adjusted empirical likelihood estimator and discuss its effect on efficiency of the target estimator.

3.5.1 Stratification

The multiplicity adjusted empirical likelihood estimator can be extended to stratified sampling designs using a method similar to the one proposed by Berger and De La Riva Torres (2016). In order to account for stratification, the sample size constraint (3.31) is defined at strata level. Therefore for T samples and H strata we have $T \times H$ constraints.

Suppose that each frame Q_t is divided into H_t strata $U_{t;1}, U_{t;2}, ..., U_{t;H_t}$. Note that

each of the sampling frames can be stratified differently. Let $\boldsymbol{S}_{t;h}$ be the sample of size $n_{t;h}$ selected from strata $U_{t;h}$ in frame Q_t. Constraint (3.31) takes the following form:

$$\sum_{i \in \boldsymbol{S}_{t;h}} m_i \rho_i = n_{t;h} \quad for \;\; t = 1, 2, ..., T; \;\; h = 1, 2, ..., H_t. \tag{3.71}$$

This is equivalent to defining $\boldsymbol{d}_i$ and $\boldsymbol{D}$ in (3.37) and (3.38) as vectors of dimension $\sum_{t=1}^{T} H_t$, with the h-th elements defined respectively by $d_{ih} = \delta_{t;h;i}^{(H)} \rho_i$ and $D_h = n_{t;h}$, where $\delta_{t;h;i}^{(H)}$ is equal to 1 if $i \in U_{t;h}$ and to 0 otherwise.

The point estimator $\widehat{\boldsymbol{\theta}}$ is not influenced by stratification. However, strata information is necessary to obtain correct confidence intervals.

3.5.2 Domain-based constraints

In some situations it may be useful to define a constraint which applies to a domain rather than to the whole population. For example, means or totals of some auxiliary variables may be known only for one sampling frame or a specific socio-demographic group, rather than for the population. Below we show how domain-level constraints can be incorporated into (3.34).

Let $\boldsymbol{\varphi}_U$ be a vector of size r of known parameters $\varphi_{U1}, \varphi_{U2}, ..., \varphi_{Ur}$ of population domains $\mathcal{D}_1, \mathcal{D}_2, ..., \mathcal{D}_r \subset U$ respectively. Let $f_i(x_i, \varphi_{Uj})$ be a function such that

$$\sum_{i \in \mathcal{D}_j} f_i(x_i, \varphi_{Uj}) = 0. \tag{3.72}$$

The domain-level known parameter constraint on the weights m_i takes the following form:

$$\sum_{t=1}^{T} \sum_{i \in \boldsymbol{S}_t} m_i diag(\boldsymbol{\delta}_i^{(\boldsymbol{\mathcal{D}})}) \mathbf{f}_i(\boldsymbol{\varphi}_U) = \mathbf{0}_r, \tag{3.73}$$

where $\boldsymbol{\delta}_i^{(\boldsymbol{\mathcal{D}})}$ is an r-vector with j-th element equal to 1 if $i \in \mathcal{D}_j$ and to 0 otherwise.

This constraint can be handled by adding an n by r sub-matrix $\boldsymbol{c}_{\mathcal{D}} = diag(\boldsymbol{\delta}_i^{(\boldsymbol{\mathcal{D}})})\mathbf{f}_i(\boldsymbol{\varphi}_U)^\top$ into the matrix $\boldsymbol{c}$ (defined in (3.37)) and r zeros to vector $\boldsymbol{C}$ (defined in (3.38)).

When $\mathcal{D}_j = U$, the parameter φ_{Uj} is a population parameter. Note that the constraint may also include domain counts. In particular, constraints involving frame sizes as well as the size of the overlapping domain may be easily defined. These constraints are commonly used in the regression type multiple frame estimators and they have been found to considerably improve precision of the point estimators (e.g. Ranalli et al. (2016)).

3.5.3 Alignment constraints on the overlapping domain

A special case of a domain-level constraint is an alignment-type constraint on the overlapping domain. Some estimators (e.g. Fuller and Burmeister (1972), Skinner and Rao (1996), Skinner (1991), Rao and Wu (2009*b*), Ranalli et al. (2016)) use a constraint on the equality of estimators of a population parameter defined for the domain $\mathcal{D}_{AB}$, obtained from each of the samples, in order to increase precision of estimators for population parameters. Aligning estimates for the common domain may also be convenient for the sake of numerical consistency.

Suppose that two sampling frames Q_A and Q_B overlap. Let $\mathcal{D}_{AB}$ of size $N_{\mathcal{D}_{AB}}$ be the set of units which appear in both frames. Let $\boldsymbol{S}_{AB}^{(A)}$ be the intersection of $\mathcal{D}_{AB}$ and the sample selected from frame Q_A and let $\boldsymbol{S}_{AB}^{(B)}$ be the intersection of $\mathcal{D}_{AB}$ and the sample selected from frame Q_B. Suppose that we want to define an alignment constraint on the estimates of a population mean $\boldsymbol{\xi}_{\mathcal{D}_{AB}} = N_{\mathcal{D}_{AB}}^{-1} \sum_{i \in \mathcal{D}_{AB}} \boldsymbol{\xi}(\boldsymbol{w}_i)$, where $\boldsymbol{\xi}(\boldsymbol{w}_i)$ is a known function of $\boldsymbol{w}_i$, as in the examples given in chapter 2, and $\boldsymbol{w}_i$ are selected components of $\boldsymbol{v}_i$ measured for each sampled unit that belongs to domain $\mathcal{D}_{AB}$. The alignment constraint for the two frames takes the following

form:

$$\sum_{i \in \boldsymbol{S}_{AB}^{(A)}} m_i \boldsymbol{\xi}(\boldsymbol{w}_i) = \sum_{i \in \boldsymbol{S}_{AB}^{(B)}} m_i \boldsymbol{\xi}(\boldsymbol{w}_i) \cdot \tag{3.74}$$

This can be translated to

$$\sum_{t=1}^{T} \sum_{i \in \boldsymbol{S}} m_i \delta_i^{(\mathcal{D}_{\mathcal{AB}})} \delta_i^{(Q_A)} \boldsymbol{\xi}(\boldsymbol{w}_i) = \mathbf{0}_r, \;\; t = (1,2), \tag{3.75}$$

where r is the dimension of vector $\boldsymbol{\xi}(\boldsymbol{w}_i)$, $\delta_t^{(\mathcal{D}_{\mathcal{AB}})}$ equals 1 if unit i belongs to domain $\mathcal{D}_{AB}$ and 0 otherwise, and $\delta_i^{(Q_A)}$ equals 1 if unit i appears in frame Q_A and -1 otherwise.

The effect of the alignment constraint on the precision of the estimator (3.41) depends on the strength of the correlation between $\boldsymbol{\xi}(\boldsymbol{w}_i)$ and $\boldsymbol{g}_i(\boldsymbol{\theta})$ and the variance of the estimates $\hat{\boldsymbol{\xi}}_{\mathcal{D}_{AB}1p}$ and $\hat{\boldsymbol{\xi}}_{\mathcal{D}_{AB}2p}$ obtained from each of the samples for the overlapping domain, where

$$\hat{\boldsymbol{\xi}}_{\mathcal{D}_{AB}tp} = \sum_{i \in \boldsymbol{S}_t} \rho_i^{-1} \delta_t^{(\mathcal{D}_{\mathcal{AB}})} \boldsymbol{\xi}(\boldsymbol{w}_i) \cdot \tag{3.76}$$

This can be seen if we consider that when an alignment constraint is used, the estimator (3.54) takes the following form:

$$\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}) = \{\widehat{\boldsymbol{G}}_{r1}(\boldsymbol{\theta}_1)^\top, \widehat{\boldsymbol{G}}_{r2}(\boldsymbol{\theta}_2)^\top\}^\top, \tag{3.77}$$

where

$$\begin{aligned}
\widehat{\boldsymbol{G}}_{r1}(\boldsymbol{\theta}_1) \;\; := \;\; & \widehat{\boldsymbol{G}}_{1p}(\boldsymbol{\theta}) - \widehat{\boldsymbol{B}}_{1\mathbf{f}1}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)^\top \, \widehat{\mathbf{f}}_{1p}(\boldsymbol{\varphi}_1) - \widehat{\boldsymbol{B}}_{1\mathbf{f}2}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)^\top \, \widehat{\mathbf{f}}_{2p}(\boldsymbol{\varphi}_2) \\
& + \widehat{\boldsymbol{B}}_{1\boldsymbol{\xi}}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)^\top \big(\hat{\boldsymbol{\xi}}_{\mathcal{D}_{AB}2p} - \hat{\boldsymbol{\xi}}_{\mathcal{D}_{AB}1p}\big), \qquad (3.78) \\
\widehat{\boldsymbol{G}}_{r2}(\boldsymbol{\theta}_2) \;\; := \;\; & \widehat{\boldsymbol{G}}_{2p}(\boldsymbol{\theta}) - \widehat{\boldsymbol{B}}_{2\mathbf{f}1}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)^\top \, \widehat{\mathbf{f}}_{1p}(\boldsymbol{\varphi}_1) - \widehat{\boldsymbol{B}}_{2\mathbf{f}2}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)^\top \, \widehat{\mathbf{f}}_{2p}(\boldsymbol{\varphi}_2) \\
& + \widehat{\boldsymbol{B}}_{2\boldsymbol{\xi}}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)^\top \big(\hat{\boldsymbol{\xi}}_{\mathcal{D}_{AB}1p} - \hat{\boldsymbol{\xi}}_{\mathcal{D}_{AB}2p}\big) \qquad (3.79)
\end{aligned}$$

and $\widehat{\boldsymbol{G}}_{tp}(\boldsymbol{\theta}) = \sum_{i\in\boldsymbol{S}_t} \rho_i^{-1}\boldsymbol{g}_{ti}(\boldsymbol{\theta}_{tU})$, $\widehat{\mathbf{f}}_{tp}(\boldsymbol{\varphi}_t) = \sum_{i\in\boldsymbol{S}_t} \rho_i^{-1}\mathbf{f}_{ti}(\boldsymbol{\varphi}_{tU})$. The estimator (3.78) is an extended regression system with an alignment term, analogous in form to estimator (2.92). As it was discussed in chapter 3, the added variance associated with the fact that $\boldsymbol{\xi}_{\mathcal{D}_{AB}}$ is estimated decreases the precision of the estimator (3.78), but the variance of (3.78) is also reduced based on the correlation between $\boldsymbol{\xi}(\boldsymbol{w}_i)$ and $\boldsymbol{g}_i(\boldsymbol{\theta})$. Similarly to the examples given in chapter 2, the function $\boldsymbol{\xi}(\boldsymbol{w}_i)$ can be chosen to maximise this correlation. In practical applications the size of the overlap is likely to be crucial for the decision whether to include the alignment constraint. When the overlap is small, the estimates of $\boldsymbol{\xi}_{\mathcal{D}_{AB}}$ are likely to be unstable. Note that if the parameter $\boldsymbol{\xi}_{\mathcal{D}_{AB}}$ is known, it is always better to use a domain-level calibration constraint for each frame than to impose an alignment constraint.

3.6 Relationship to the aligned empirical likelihood estimator

In this chapter we show how the multiplicity adjusted empirical likelihood estimator relates to the aligned empirical likelihood estimator presented in chapter 2. Suppose that we have a two frames design, where samples are selected independently from each frame. If the two frames overlap completely and at least one variable with unknown population level values is measured in both samples, that is, $Q_1 = Q_2 = U$, then the sampling design is as in chapter 2.

The key difference between the two estimators is of course in their primary aim. The aligned empirical likelihood estimator is used when we want to obtain a single vector of weights which can be used to estimate common and non common population parameters and which produces equal estimates for the common parameters when it is applied to each survey separately. The multiplicity adjusted empirical likelihood estimator is used when the variables of interest are measured in both

samples and we assume that unit level data from both samples will always be available to calculate estimates. Note that in practice the two estimators are likely to be used in different context. Alignment of estimates is more often required when several surveys were carried out independently in the same population, while the multiplicty adjusted empirical likelihood estimator is more likely to be applied in a classical multiple frame settings, when a single survey uses multiple sampling frames because a single frame with a good coverage is unavailable.

It is, however, theoretically interesting to consider how the two estimators compare in the case when there is a complete overlap between the frames and all survey variables are measured in both samples. The difference is in the formulation of the sample size (design) constraint. The aligned empirical likelihood estimator uses a design constraint

$$\sum_{i \in \boldsymbol{S}_t} m_{1i}\ \pi_{1i} = n_t, \quad t = 1, 2\cdot \tag{3.80}$$

The multiplicity adjusted empirical likelihood estimator is based on a sample size constraint

$$\sum_{i \in \boldsymbol{S}_t} m_i \alpha_i^{-1} \pi_i = n_t, \quad t = 1, 2\cdot \tag{3.81}$$

Suppose that we want to estimate a mean $\boldsymbol{\xi}_{_U} = N^{-1} \sum_{i \in U} \boldsymbol{\xi}(\boldsymbol{w}_i)$ of a known function $\boldsymbol{\xi}$ of the common variable $\boldsymbol{w}_i$. The aligned empirical likelihood estimator of $\boldsymbol{\xi}$ is the solution to

$$\sum_{i \in \boldsymbol{S}_1} \widehat{m}_{1i}(\boldsymbol{\varphi}_{_U})_{(ALGN)}\ \boldsymbol{h}_{1i}(\boldsymbol{\xi}) = \boldsymbol{0} \quad \text{and} \quad \sum_{i \in \boldsymbol{S}_2} \widehat{m}_{2i}(\boldsymbol{\varphi}_{_U})_{(ALGN)}\ \boldsymbol{h}_{2i}(\boldsymbol{\xi}) = \boldsymbol{0}, \tag{3.82}$$

where $\boldsymbol{h}_{ti}(\boldsymbol{\xi}) = \boldsymbol{\xi}(\boldsymbol{w}_i) - N n_t^{-1} \pi_{ti}\ \boldsymbol{\xi}$. The solutions of the equations (3.82) are $\widehat{\boldsymbol{\xi}}_1$

and $\widehat{\boldsymbol{\xi}}_2$, where

$$\widehat{\boldsymbol{\xi}}_t = N^{-1} \sum_{i \in \boldsymbol{S}_t} \widehat{m}_{ti}(\boldsymbol{\varphi}_{U})_{(ALGN)} \; \boldsymbol{\xi}(\boldsymbol{w}_i) \cdot \tag{3.83}$$

Because of the alignment constraint (2.44), these solutions are equal, i.e., $\widehat{\boldsymbol{\xi}}_1 = \widehat{\boldsymbol{\xi}}_2$. That is, once the adjusted weights $\widehat{m}_{ti}(\boldsymbol{\varphi}_{U})$ have been calculated, it is sufficient to use one sample data to obtain the estimate $\widehat{\boldsymbol{\xi}}$. The adjusted empirical likelihood weights in (3.82) are equal to

$$\widehat{m}_i(\boldsymbol{\varphi}_{U})_{(ALGN)} = \left(\pi_i + \boldsymbol{\eta}_{(ALGN)}^{\top} \boldsymbol{c}_i\right)^{-1} \cdot \tag{3.84}$$

The multiplicity adjusted empirical likelihood estimator of $\boldsymbol{\xi}$ is the solution to the estimating equation which involves values of both samples:

$$\sum_{t=1}^{T} \sum_{i \in \boldsymbol{S}_t} \widehat{m}_i(\boldsymbol{\varphi}_{U})_{(MLT)} \boldsymbol{h}_{ti}(\boldsymbol{\xi}) = \boldsymbol{0}, \tag{3.85}$$

where the adjusted weights are given by

$$\widehat{m}_i(\boldsymbol{\varphi}_{U})_{(MLT)} = \left\{\rho_i + \boldsymbol{\eta}_{(MLT)}^{\top} \boldsymbol{c}_i\right\}^{-1} = \left\{\alpha_i^{-1} \pi_i + \boldsymbol{\eta}_{(MLT)}^{\top} \boldsymbol{c}_i\right\}^{-1} \cdot \tag{3.86}$$

If the Simple Multiplicity (Mecatti, 2005) adjustment is used and we have two surveys with complete overlap, $\alpha_i^{-1} = T = 2$. However, when full information on selection probabilities from both frames is available, an adjustment α_i leading to a more efficient estimator may be chosen, e.g. Kalton and Anderson's (1986) adjustment. Therefore, if there is a complete overlap of sampling frames, only the common variables are of interest and full frame information is available, the multiplicty adjusted empirical likelihood estimator is likely to be more efficient than the aligned empirical likelihood estimator. On the other hand, the aligned empirical likelihood estimator does not require access to each sample's microdata after the adjusted weights are calculated.

3.7 Simulation study

We evaluate performance of the proposed empirical likelihood multiplicity adjusted (ELM) point estimator and coverage of the empirical likelihood confidence intervals in a series of simulations. For comparison, the generalized multiplicity-adjusted regression estimator (GMREG) (Ranalli et al., 2016) is also calculated.

Each of the estimators requires calculating the multiplicity-adjusted selection probability of the i-th unit, $p_{t;i} = \pi_{t;i}\alpha_{t;i}^{-1}$. The following adjustments are used:

- Simple multiplicity adjustment (Mecatti, 2005)

$$\alpha_{t;i} = M_i^{-1}, \quad M_i = \sum_{t=1}^{T} \kappa_{t;i}, \tag{3.87}$$

- Kalton and Anderson's (1986) adjustment

$$\alpha_{t;i} = \pi_{t;i} \left(\sum_{t=1}^{T} \pi_{t;i}\kappa_{t;i} \right)^{-1}. \tag{3.88}$$

The simulations were performed using the R software (R Core Team, 2015). For the empirical likelihood estimator, new procedures were developed. The GMREG estimator was calculated using the Frames2 package (Arcos et al., 2015).

We use the following study populations:
Population 1 is a synthetic dataset of size $23,500$ units, generated according to the following model: (Ranalli et al., 2016) $y \sim N(5000, 500)$, $x_{1i} = (y_i - \epsilon_{1i})/0.5$, $x_{2i} = (y_i - \epsilon_{2i})/1.2$, $\epsilon_{1i} \sim N(500, 300)$, $\epsilon_{2i} \sim N(700, 500)$.

Population 2 consists of 50,000 units and is generated according to the following model (Wu and Rao (2006)):

$$y_i = 3 + a_i + 8x_i + 6z_i + 0.5e_i, \tag{3.89}$$

where a, x and z follow independent exponential distributions with the rate parameter equal to 1 and $e_i \sim \chi^2_1 - 1$. This produces a dataset of highly skewed variables.

Population 3 is a synthetic dataset *EUSILCP* (Alfons et al., 2010) available within the R (R Core Team, 2015) package simFrame (Alfons, 2013). It contains 58,654 units. *EUSILCP* was modelled based on the Austrian EU-SILC (European Union Statistics on Income and Living Conditions) survey from year 2006 and preserves distributions of the key variables. We use the following variables: equalised household income, household size and age.

Below we show the results of several simulation studies. We start with estimating totals. We first consider a normally distributed variable of interest, generated independently from the frame allocation mechanism. We then use some more skewed variables and check the effect of the correlation between the variable of interest and the frame allocation mechanism as well as the effect of the correlation between the variable of interest and the sampling probabilities. We also check how well the proposed multiplicity adjusted empirical likelihood estimator deals with quantiles of distribution.

3.7.1 Estimation of totals

Normal data

The first simulation tests the performance of the proposed estimator in the relatively favourable conditions of *Population 1*. The simulations follow the conditions tested by Ranalli et al. (2016). Units are allocated to two frames according to the values of variable g_i generated from a binomial distribution. Two levels of overlap are tested. First, $g_i \sim B(2, 0.3)$ and unit i is allocated to frame 1 if $g_i = 0$ or $g_i = 1$ and to frame 2 if $g_i = 1$ or $g_i = 2$. This gives a small overlap of about 9%

of the population size. Second, $g_i \sim B(2, 0.5)$ and unit i is allocated to frame 1 if $g_i = 0$ or $g_i = 1$ and to frame 2 if $g_i = 1$ or $g_i = 2$. This gives large overlap of approximately 50%. Sample 1 is selected from frame 1 using stratified simple random sampling with replacement, sample 2 is selected from frame 2 using Midzuno sampling with $\pi_i \sim y_i - N(300, 200)$. The sample sizes are $n_1 = 201$ and $n_2 = 270$. We assume that the frame totals of variable x are known. We also include constraints on the frame sizes. We test the estimators when the size of the overlap between frames is known and when it is not known. Table (3.1) presents 100 × percent relative mean squared error ($100 \times RMSE\%$) of the tested estimators with the Kalton-Anderson and the Simple Multiplicity adjustments.

Table 3.1: 100 × percent relative mean squared error ($100 \times RMSE\%$) of the proposed Empirical Likelihood Multiplicity adjusted (ELM) estimator and the Generalized Multiplicity-adjusted Regression estimator (GMREG) (Ranalli et al., 2016). *Population 1.* $\boldsymbol{S}_1$: stratified simple random sampling, $\boldsymbol{S}_2$: Midzuno sampling. Based on 1,000 iterations.

α	Small overlap		Large overlap	
	GMREG	ELM	GMREG	ELM
	N_{AB} unknown			
Multiplicity	0.45	0.48	2.57	0.58
Kalton-Anderson	0.50	0.35	2.30	0.67
	N_{AB} known			
Multiplicity	0.10	0.48	0.12	0.58
Kalton-Anderson	0.10	0.34	0.12	0.67

The normality of the variable of interest creates relatively favourable conditions for the regression estimators. When the size of the overlapping domain is known, the GMREG estimator is more precise than the proposed ELM estimator. However, the ELM estimator is more precise when the size of the overlapping domain is unknown, especially when this overlap is large.

Variables with skewed distributions and correlation between the variable of interest and frame allocation

Second simulation attempts to test relative performance of the estimators in a more complex situation. We use *Population 2*, where the variable of interest follows a skewed distribution. We introduce some dependence between the variable of interest and the frame allocation, as well as the sampling probabilities. Units are allocated to frames according to the values of the variable $g_i \sim B(2, y_i^{st})$, where $y_i^{st} = \{y_i - min(\boldsymbol{y})\}\{max(\boldsymbol{y}) - min(\boldsymbol{y})\}^{-1}$. Three frame allocation methods are tested:

- unit i appears in frame 1 when $g_i = 0$ or $g_i = 2$ and in frame 2 when $g_i = 1$ or $g_i = 2$. This gives an overlap between frames of approximately 5%.
- unit i appears in frame 1 when $g_i = 0$ or $g_i = 1$ and in frame 2 when $g_i = 1$ or $g_i = 2$. This gives an overlap between frames of approximately 27%.
- unit i appears in frame 1 when $g_i = 0$ or $g_i = 2$ and in frame 2 when $g_i = 0$ or $g_i = 1$. This gives an overlap between frames of approximately 70%.

1,500 units are selected from each frame by random systematic sampling with sampling probabilities π_i proportional to a size variable τ_i, generated according to the following model $\tau_i =\sim 2y_i + l_i + k_i + 30$, with $l_i \sim \chi^2_{df=1}$ and $k_i \sim N(10, 10)$. This introduces a correlation between the variable of interest y_i and the sampling probabilities π_i of about 0.88.

Table (3.2) presents $100 \times RMSE\%$ of the proposed estimator and the GMREG estimator with the three overlap sizes, when the size of the overlapping domain is known and when it is unknown.

With this skewed dataset, the proposed estimator is slightly more precise than the GMREG estimator in all cases, even when full information on the size of the

Table 3.2: 100 × percent relative mean squared error ($100 \times RMSE\%$) of the proposed Empirical Likelihood Multiplicity adjusted (ELM) estimator and the Generalized Multiplicity-adjusted Regression estimator (GMREG) (Ranalli et al., 2016). *Population 2*. Random systematic sampling design. Based on 1,000 iterations.

α	V. small overlap		Small overlap		Large overlap	
	GMREG	ELM	GMREG	ELM	GMREG	ELM
	N_{AB} unknown					
Multiplicity	0.39	0.17	0.57	0.22	0.68	0.24
Kalton-Anderson	0.38	0.17	0.55	0.24	0.67	0.24
	N_{AB} known					
Multiplicity	0.34	0.17	0.43	0.22	0.41	0.24
Kalton-Anderson	0.33	0.17	0.42	0.24	0.41	0.24

frames overlap is available. We notice that the precision of the ELM estimator does not deteriorate when the size of the overlap is unknown and that it is mildly affected by the size of the overlap. We observe slight deterioration of the GMREG estimator when the size of the overlap is large and unknown. We should note, however, that both estimators have low relative root square errors.

We also test whether the coverage of the multiplicity adjusted empirical likelihood confidence intervals is affected by the dependence between the frame allocation and the values of the variable of interest, the size of the overlap or the fact that the level of overlap is unknown. We calculate the confidence intervals for the ELM estimator, calculated as discussed in chapter 3.4.4. For the GMREG estimator, symmetric confidence intervals are calculated based on the Deville and Särndal's (1992*a*) variance estimator available within the Frames2 package (Arcos et al., 2015).

Table (3.3) presents converges of confidence intervals of the two tested estimators.

Both confidence intervals show some over-coverage, especially when the overlap is small. The coverage of the empirical likelihood confidence interval is similar to that of the normality based confidence interval, but typically slightly closer to the nominal level and in most cases is not significantly different from 95%. The coverage of confidence intervals is not affected by including the constraints on the

Table 3.3: Coverage of confidence intervals (%) of the the proposed Empirical Likelihood Multiplicity adjusted (ELM) estimator and the Generalized Multiplicity-adjusted Regression estimator (GMREG) (Ranalli et al., 2016) in *Population 2*. Based on 1,000 iterations. †: values significantly different from the nominal level.

α	V small overlap		Small overlap		Large overlap	
	GMREG	ELM	GMREG	ELM	GMREG	ELM
			N_{AB} unknown			
Multiplicity	97.1†	96.2	98.0†	95.7	95.0	95.3
Kalton-Anderson	97.5†	96.3	98.1†	95.9	95.0	95.3
			N_{AB} known			
Multiplicity	97.4†	96.2	97.3 †	95.9	96.2	95.3
Kalton-Anderson	97.7†	96.3	97.3†	95.7	95.9	95.3

size of the overlapping domain.

Correlation between the variable of interest and the sampling probabilities

The third simulation focuses on the effect of the correlation between the variable of interest and the selection probabilities. We use data from *Population 3*, restricted to only include people over 16 years old. Units are allocated to frames as in the three scenarios listed above. We estimate the total of the equalised household income and use the total of the household size from frame 1 and the total of age from frame 2 as auxiliary information. We also incorporate information on the frame sizes and the size of the overlap between frames. Samples of size 1,000 are selected in each frame using random systematic sampling with sampling probabilities proportional to a size variable τ_i. We first check performance of the estimators when τ_i is generated from an independent normal distribution, i.e., $\tau_i \sim N(100, 20)$. We then introduce a correlation between the sampling probabilities π_i and the variable of interest by using $\tau_i \sim 0.7 * y_i + \gamma + \delta$, where $\gamma \sim \chi^2_{df=1}$ and $\delta \sim N(100000, 10000)$. This gives $cor(\pi_i, y_i) \approx 0.6$.

The $100RMSE\%$ of the proposed estimator and the GMREG estimator are pre-

sented in tables (3.4) and (3.5). Table (3.4) shows results obtained when the π_i are generated independently from the variable of interest. Table (3.5) show results obtained with $cor(\pi_i, y_i) \approx 0.6$.

Table 3.4: 100 × percent relative mean squared error ($100 \times RMSE\%$) of the proposed Empirical Likelihood Multiplicity adjusted (ELM) estimator and the Generalized Multiplicity-adjusted Regression estimator (GMREG) (Ranalli et al., 2016). *Population 3*. Sampling probabilities generated from an independent normal distribution. Random systematic sampling design. Based on 1,000 iterations.

α	V. small overlap		Small overlap		Large overlap	
	GMREG	ELM	GMREG	ELM	GMREG	ELM
	N_{AB} unknown					
Multiplicity	1.80	2.14	2.04	2.24	2.45	2.05
Kalton-Anderson	1.75	2.06	1.91	2.06	2.45	2.05
	N_{AB} known					
Multiplicity	1.78	2.14	1.99	2.24	1.76	2.05
Kalton-Anderson	1.74	2.06	1.87	2.06	1.75	2.04

Table 3.5: 100 × percent relative mean squared error ($100 \times RMSE\%$) of the proposed Empirical Likelihood Multiplicity adjusted (ELM) estimator and the Generalized Multiplicity-adjusted Regression estimator (GMREG) (Ranalli et al., 2016). *Population 3*, $cor(y_i, \pi_i) \approx 0.6$. Random systematic sampling design. Based on 1,000 iterations.

α	V. small overlap		Small overlap		Large overlap	
	GMREG	ELM	GMREG	ELM	GMREG	ELM
	N_{AB} unknown					
Multiplicity	1.37	1.09	1.68	1.30	1.89	1.34
Kalton-Anderson	1.34	1.09	1.55	1.27	1.88	1.33
	N_{AB} known					
Multiplicity	1.36	1.09	1.62	1.30	1.56	1.34
Kalton-Anderson	1.34	1.09	1.51	1.27	1.54	1.27

When the sampling probabilities follow an independent normal distribution, the proposed estimator has slightly higher relative mean square error than the GMREG estimator in all cases when the full frame information is available and with small and medium overlap sizes when the size of the overlap is unknown. When the overlap is large and its population size is unknown, the ELM estimator is slightly more precise. When the π_i and the target variable are correlated, the ELM estimator performs better in all cases.

Table (3.6) presents converges of confidence intervals of the two tested estimators. The same method of calculating the lower and upper bounds of confidence intervals for the ELM and the GMREG estimators as described in the previous chapter was used. The coverage of the proposed empirical likelihood confidence interval is acceptable across all tested scenarios. When the overlap is large, the empirical likelihood confidence intervals show very slight over-coverage, significantly different from the nominal value in a couple of cases. The GMREG confidence intervals also have good coverage, with only one case of under-coverage significantly different from 95%.

Table 3.6: Coverage of confidence intervals (%) of the the proposed Empirical Likelihood Multiplicity adjusted (ELM) estimator and the Generalized Multiplicity-adjusted Regression estimator (GMREG) Ranalli et al. (2016). *Population 3*. Based on 1,000 iterations. †: values significantly different from the nominal coverage level.

α	V small overlap		Small overlap		Large overlap	
	GMREG	ELM	GMREG	ELM	GMREG	ELM
$cor(y_i, \pi_i) \approx 0.6$						
			N_{AB} unknown			
Multiplicity	94.7	95.3	94.1	94.6	94.5	96.3
Kalton-Anderson	94.7	94.6	95.3	95.0	94.0	96.2
			N_{AB} known			
Multiplicity	95.6	95.2	94.1	94.8	95.6	96.6†
Kalton-Anderson	94.9	94.6	94.7	95.1	95.4	96.3
$cor(y_i, \pi_i) \approx 0$						
			N_{AB} unknown			
Multiplicity	93.6†	95.0	94.1	94.6	94.5	96.3
Kalton-Anderson	94.0	95.4	95.3	95.0	94.0	96.2
			N_{AB} known			
Multiplicity	93.8	95.0	94.1	94.8	95.6	96.6†
Kalton-Anderson	94.1	95.2	94.7	95.1	95.4	96.3

3.7.2 Estimation of quantiles of distribution

One of the benefits of the proposed approach is its flexibility to handle a wide class of population parameters of interest other than means or totals. Table (3.7) shows

relative absolute root mean square errors (%) (rrmse), 100 × percent relative mean squared error (rmse), left tail error rates, right tail error rates and coverages of confidence intervals for the proposed estimator of the 10th, 20th, 80th and 90th quantile of distribution of equalised household income in *population 2*. We test performance of the estimator with two different sample sizes: $n_1 = 1000, n_2 = 1500$ and $n_1 = 2000, n_2 = 2500$, with an overlap of approximately 50% of the population size. The random systematic sampling design is used in each frame and the sampling probabilities are proportional to the household size. We assume that the population totals of age and household size are known.

The error rates of the proposed estimator are all of acceptable size, although, understandably, the 100 × percent relative mean squared error rates are noticeably larger than those for the estimators of totals presented in the previous tables. The coverages of the confidence intervals are close to the nominal levels in almost all the cases. The tail error rates are unbalanced, especially when high quantiles are estimated. This is likely to be caused by the skewness of the data. Overall, however, we can say that the multiplicity adjusted empirical likelihood confidence intervals perform well, especially considering the skewness of the variables of interest.

Table 3.7: Relative absolute root mean square errors (%) (rrmse), 100 × percent relative mean squared errors (rmse), left tail error rates (l. t.e.r.), right tail error rates (r. t.e.r.) and coverages of confidence intervals (cov.) of the proposed Empirical Likelihood Multiplicity-Adjusted estimator. *EUSILCP* data. Estimation of quantiles of distribution and the mean of equalised household income. Stratified random systematic sampling design, stratification by household size with proportional allocation. KA: Kalton-Anderson's adjustment, ML: Multiplicity adjustment. †: values significantly different from the nominal coverage level.

θ	α	$n_1 = 1000$, $n_2 = 1500$					$n_1 = 2000$, $n_2 = 2500$				
		rrmse	rmse	l. t.e.r	r. t.e.r.	cov.	rrmse	rmse	l. t.e.r	r. t.e.r.	cov.
	KA	3.4	11.6	2.6	2.0	95.4	2.3	5.3	$3.5^{\dagger}$	$1.1^{\dagger}$	95.4
q_{10}	*ML*	3.4	11.6	2.3	2.1	95.6	2.3	5.3	2.7	$1.2^{\dagger}$	96.1
	KA	2.8	7.8	$1.4^{\dagger}$	3.4	95.2	1.9	3.6	2.5	1.9	95.6
q_{20}	*ML*	2.8	7.8	0.8	$4.2^{\dagger}$	95.0	1.9	3.6	2.2	2.0	95.8
	KA	2.0	4.0	$1.0^{\dagger}$	$4.9^{\dagger}$	94.1	1.4	2.0	$1.1^{\dagger}$	$4.4^{\dagger}$	94.5
q_{80}	*ML*	2.0	4.0	$1.5^{\dagger}$	$3.8^{\dagger}$	94.7	1.3	1.7	$1.2^{\dagger}$	3.2	95.6
	KA	2.8	7.8	$1.0^{\dagger}$	$4.7^{\dagger}$	94.3	3.1	9.6	$1.2^{\dagger}$	$3.5^{\dagger}$	95.3
q_{90}	*ML*	2.9	8.4	$1.1^{\dagger}$	$3.7^{\dagger}$	95.2	3.1	9.6	1.9	2.8	95.3

3.8 Conclusions

We propose an Empirical Likelihood approach to finite population parameter estimation in the multiple frames context. The estimator is based on the multiplicity adjustment principle (Singh and Mecatti, 2011; Mecatti and Singh, 2014; Rao and Wu, 2009*b*), and can accommodate various multiplicity adjustment factors. Additional benchmark constraints constructed around known population level parameters may be incorporated easily. In particular, constraints on the frame size and size of the overlapping domain can be included. Previous research (Ranalli et al., 2016) have shown that this type of constraints often lead to considerable gains in precision. Alignment type constraints, requiring that both frames produce the same point estimates for parameters of the overlapping domain, can also be defined. The alignment constraint can be formulated for a mean of a function of the common variable. A function which maximises the correlation with the variable of interest should be selected.

A wide class of parameters, expressed as solutions to population estimating equations, can be estimated through the proposed estimator. A single weight, which can be used for estimation of various parameters, is obtained for every unit. The weights are positive by definition.

Empirical likelihood multiplicity adjusted confidence intervals for finite population parameters are constructed based on the χ^2 approximation of the distribution of the empirical likelihood ratio statistic under the null hypothesis. As in the case of the aligned empirical likelihood estimator discussed in the previous chapter, the confidence intervals do not require variance estimation, are range-preserving and asymmetric.

We consider the flexibility in terms of the type of multiplicity adjustment used, type of parameters of interest and type of constraints imposed the main benefit of the proposed method. In the simulations performed, the proposed estimator

has acceptable precision. In some cases, it tends to have a lower relative root mean square error than the generalized multiplicity adjusted regression estimator with the same multiplicity adjustment. This happens when the variable of interest follows a skewed distribution. In particular, we notice that the proposed estimator performs particularly well when the size of the frames overlap is unknown. It is also relatively insensitive to the actual size of the overlap, while the GMREG estimator often has a higher relative root mean square error when the overlap between sampling frames is large and when the size of the overlap is unknown. Coverage of the proposed empirical likelihood confidence intervals is close to the nominal level in most cases, with just a few cases of slight over-coverage or under-coverage. This holds also for estimation of quantiles of distribution.

The multiplicity adjusted estimator has a similar structure to the aligned empirical likelihood estimator discussed in chapter 2. It differs, however, in the formulation of the design constraint, in that it allows to include custom adjustment factors. We do not discuss the choice of the multiplicity adjustment factor. Several adjustments have been proposed. In practical applications, the choice of the adjustment factor is likely to be driven by the availability of information.

Chapter 4

Using empirical likelihood to obtain range-respecting confidence intervals for census coverage

4.1 Introduction

Censuses aim to obtain an almost perfect coverage of the population. However, although efforts are made to maximise the response rates, a small proportion of the population is usually missed. A second register or survey can be used to assess the overage of a census. Several countries use a survey carried out after the census. The dual system estimator (DSE) is then used in order to estimate the census coverage. Examples of such countries include the United Kingdom (Brown et al., 1999; Abbott, 2009), New Zealand (Statistics New Zealand, 2014) or Brazil (da Silva et al., 2015).

In the UK, the census coverage survey (CCS) is carried out shortly after the census. The census coverage survey has a large sample size and is used to estimate the population size by correcting the census totals in geo-demographic groups based on the estimated under-coverage, as well as to estimate the coverage of the census. The estimation procedure is undertaken separately in 106 Estimation

Areas (defined based on a geographic split of the country) and estimates are produced for 45 age-sex groups within each Estimation Area. The census coverage rate is defined as the proportion of the number of people enumerated in census to the population size estimated through the dual system estimator. In 2011, the overall coverage across England and Wales was 94% (Office for National Statistics, 2017).

Estimating the uncertainty around the census coverage estimate is not a straightforward task. The methodology used by the Office for National Statistics after the 2001 census relied on constructing symmetric confidence intervals, based on the jackknife variance estimator. The 2011 census used bootstrap bias corrected and accelerated (BCa) confidence (Efron, 1987; Baillie et al., 2011) intervals (Kabzinska et al., 2017). However, construction of the confidence intervals is difficult because of the distribution of the census coverage rate. The coverage follows the binomial distribution and varies hugely between regions and age-sex groups. There are groups where the estimated coverage rate is very close to 1. In such context confidence intervals are known to be difficult to construct (Liu and Kott, 2009). Moreover, when the estimated coverage rate is close to 1, the symmetric confidence intervals have upper bounds above 1, which might be confusing for end users.

In this chapter we consider an empirical likelihood approach for the estimation of census coverages. Empirical likelihood gives confidence interval bounds between 0 and 1. It can also easily be extended to include any benchmark (calibration) constraints. Specifically, we consider design-based empirical likelihood (Berger and De La Riva Torres, 2016) confidence intervals which are constructed by directly inverting the log-likelihood ratio function. This means that confidence intervals can be obtained without variance estimates.

This chapter begins with a summary of the design of the census coverage survey and its estimation procedure. Then chapter 4.3 explains how empirical likelihood can be applied to estimate census coverage and specifies the relevant estimating functions and constraints. Chapter 4.4 gives numerical results based on applying

empirical likelihood to data from the census coverage survey carried after the 2011 England and Wales population census. In chapter 4.5, we present the results of a simulation study which compares empirical likelihood confidence intervals with those derived from an approximation to the existing approach.

4.2 Sampling design of the census coverage survey and the current estimation procedure

In this chapter we summarise the design of census coverage survey and describe the current estimation procedure. We discuss the properties of the dual system estimator and show how the estimates of the population size are used to produce census coverage estimates.

4.2.1 Sample selection

Census coverage survey uses a stratified cluster sampling design. A separate sample is taken in each Estimation Area, which consists of roughly 1 million people. The primary sampling units are small geographical entities, called Output Areas, stratified by Local Authority and a proxy measure of how likely the local population is to respond in a census, called Hard to Count index. For each of the sampled Output Areas, a sample of postcodes is taken (Brown et al., 2011). Figure (4.1) shows this geographical division on a diagram.

In each of the sampled postcodes, full enumeration of households is attempted. The total number of households and people is measured and some additional household and person level information is gathered for each household.

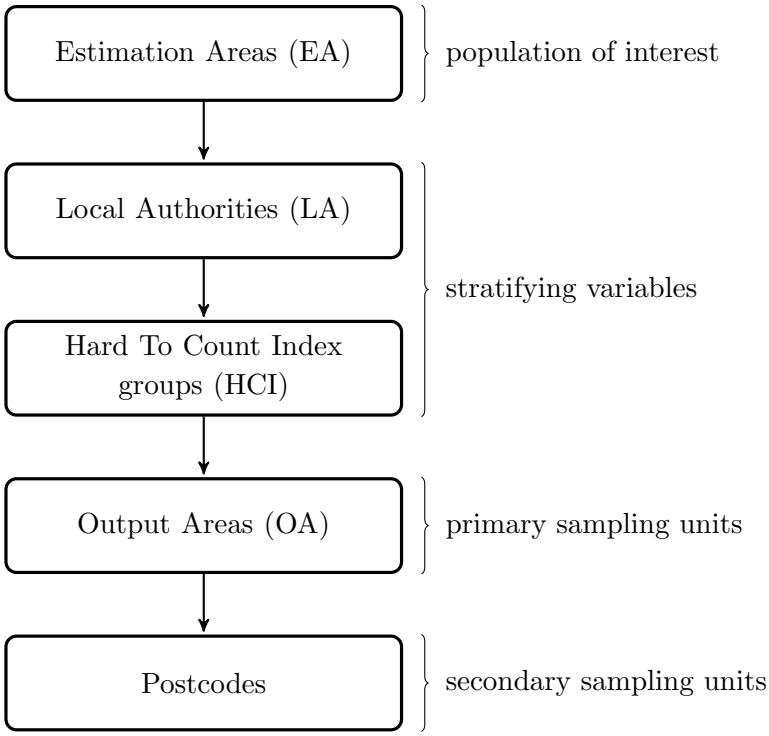


Figure 4.1: Geographical entities in England and Wales used in the design of the Census Coverage Survey

4.2.2 Dual system population size estimation

The total number of households and persons measured in the sampled postcodes is used to estimate the population size. The following dual system estimation formula is applied:

$$\widehat{DSE}_N = \frac{(N_{C^+S^+} + N_{C^+S^-})(N_{C^+S^+} + N_{C^-S^+})}{N_{C^+S^+}}, \tag{4.1}$$

where $N_{C^+S^+}$ is the number of households (or people) who were enumerated both in census and in CCS, $N_{C^+S^-}$ is the number of households (or people) who were enumerated in census but not in CCS and $N_{C^-S^+}$ is the number of households (or people) who were not enumerated in census but were enumerated in CCS.

The estimator (4.1) relies on several assumptions: (Abbott, 2009, 2011)

1. *Independence between the census and the CCS*, meaning that the proportion

of households (persons) enumerated in census is the same among CCS respondents and among CCS non-respondents and that the proportion of CCS respondents is the same among households (persons) enumerated in census and households (persons) not enumerated in census;

2. *A closed population*, meaning that no in- or out - migration happens between the census and the Census Coverage Survey;

3. *Homogeneity of selection probabilities within Output Areas*, meaning that households (persons) within the same stratum have equal chances of being enumerated in the census or in CCS;

4. *Perfect matching*, meaning that the values $N_{C^+S^+}$, $N_{C^+S^-}$ and $N_{C^-S^+}$ are correct, i.e., calculated without error.

If the assumptions above are not met, which is likely in real life conditions, the DSE is known to be negatively biased (Brown et al., 2006). Specifically, violation of the assumption of independence between the response probabilities in the Census Coverage Survey and in the census can lead to a high bias in the estimator (Brown et al., 2011). Further to this, while the DSE is used to assess census under-coverage, adjustment for over-coverage, resulting e.g. from double counting of people or households, is also necessary. The ONS applies a series of adjustments based on the estimated census over-count and on known statistics, such as sex ratios in age groups, derived from administrative records. Specifically, the following adjustments are applied: (Office for National Statistics, 2012)

1. sample balance adjustment

   This adjustment is applied if the sample is considered unbalanced. This is assessed based on a comparison between census dummy questionnaires, i.e., questionnaires filled in by census enumerators if no response from a household is obtained and the response rates estimated through the DSE. A sample is considered unbalanced if the two rates are significantly different.

2. DSE bias adjustment by age-sex group and HTC

   Adjustment for person and household level bias is applied based on social surveys and the Alternative Household Estimate, which is a national register of households created by compiling several administrative resources, such as the NHS Patient Register, the Department for Work and Pensions Customer Information System, English School Census, Welsh School Census and the Higher Education Statistics Agency data.

3. census over-count adjustment

   The over-count adjustment corrects for the fact that some people or households may be enumerated more than once. This might be because of duplicate returns for the same person, people being counted in the wrong location, or erroneous returns.

4. national level adjustment for residual bias

   The total population estimates by sex and age groups at the national level are compared to the sex ratios available from administrative sources. A national level adjustment is then applied to account for any residual bias if the population totals are implausible.

4.2.3 Population size and census coverage estimation

The second stage of estimation consists of producing Estimation Area estimates based on the values of $\widehat{DSE}_N$ and the census counts observed for the sampled postcodes. First, the census coverage, defined as the ratio of the census count to the population size ($\widehat{DSE}_N$), is estimated. This is done by fitting a straight regression line (with no intercept) to the age-sex specific values of $\widehat{DSE}_N$ and census counts for the sampled postcodes. Second, the population size in age-sex

groups is estimated through a ratio estimator (Abbott, 2009).

After the population estimates are obtained for each Estimation Area, a synthetic small area estimator (e.g. Rao, 2015) is used to obtain population estimates for each Local Authority (Brown et al., 2011). Finally, the national database is adjusted through imputation for the estimated under-count (Abbott, 2009).

After the 2001 census, symmetric confidence intervals for census coverage were obtained through the Jacknife variance estimator (Kabzinska et al., 2017). Following the 2011 census, symmetric bootstrap bias corrected and accelerated confidence intervals (Efron, 1987) were used (Kabzinska et al., 2017). Confidence interval half-widths were then published, even though some of them resulted in the upper bound larger than 1.

Note that in any estimates produced from the census coverage survey there are two sources of variance. First, there is the sampling variance associated with the fact that Estimation Area level estimates are obtained from a sample of postcodes. Second, there is the variance of the dual system estimator. This variance is associated with the fact that the probabilities of responding in the census and in the census coverage survey are random quantities. The current methodology for estimation of census coverage treats the dual system estimates of the population size as fixed quantities, i.e., the confidence intervals are constructed based on the estimate of the sampling variance in the census coverage survey. We follow this approach in this work, although we acknowledge that further development which would account for this uncertainty would be desirable.

4.3 Applying empirical likelihood to census coverage

In this chapter we consider an empirical likelihood approach to census coverage estimation. We focus specifically on the second step in the current approach, when

the postcode level census coverages are used to produce census coverage estimates at the Estimation Area level. In line with the current approach, we treat the DSE population sizes in sampling units (postcodes) as fixed, non-random quantities. We also assume that any adjustments have already been applied to the DSE. We do not extend the method to allow for small area estimates at the Local Authority level. This would require some further theoretical development in the empirical likelihood methodology. We therefore do not claim that the entire methodology for population size estimation can be substituted with the presented empirical likelihood method. Instead, we suggest that the empirical likelihood method could be used to produce the confidence intervals for census coverage estimates at the Estimation Area level, which are currently reported as half-widths and occasionally exceed 1.

In line with the design described in chapter 4.2.1, we consider T populations of interest, called Estimation Areas and denoted U_t, where $t = 1, \ldots, T$. Each Estimation Area U_t is divided into H disjoint strata $h = 1, 2, ..., H$. An independent sample $\boldsymbol{S}_{t;h}$, of size $n_{t;h}$, of output areas is taken from each of the H strata. Let $\boldsymbol{S}_t$ be the collection of labels of the selected output areas; that is $\boldsymbol{S}_t = \cup_{h=1}^{H} \boldsymbol{S}_{t;h}$. Let $\pi_{t;h;i}$ denote the selection probability for output area i in stratum h. Within each of the selected output areas, $n_{t;h;i}$ postcodes are selected with unequal probabilities $\pi_{t;h;i;k}$, where k is the postcode index. Let $\boldsymbol{S}_{t;h;i}$ be the collection of labels of the postcodes selected in output area i in stratum h. To simplify the notation, we drop the stratum index h in the following text wherever the stratum membership can be ignored.

Suppose that for every sampled postcode k, we know the DSE estimate of the population size, denoted by $y_{t;i;k}$, and the number of individuals enumerated in census, denoted by $x_{t;i;k}$. Suppose also that for every output area i, the number of individuals enumerated in the census, denoted by $x_{t;i}$, is known.

Let θ_{U_t} denote the parameter of interest, that is, the census coverage at the Estimation Area level. We define θ_{U_t} as the ratio of the census count to the dual

system estimate of the population size. The parameter θ_{U_t} can be expressed as the unique solution of the population estimating equation:

$$\sum_{i \in U_t} g(y_{t;i}, x_{t;i}, \theta_{U_t}) = 0, \tag{4.2}$$

where

$$g(y_{t;i}, x_{t;i}, \theta_{U_t}) = x_{t;i} - \theta_{U_t} y_{t;i} \tag{4.3}$$

and $y_{t;i}$ denotes the population size in output area $i \in U_t$. The value $y_{t;i}$ is not assumed to be known.

The design-based empirical likelihood function is defined as: (Berger and De La Riva Torres, 2016)

$$\ell(m_{t;i}) = \sum_{i \in \boldsymbol{S}_t} \log(m_{t;i}), \tag{4.4}$$

where the $m_{t;i}$ are unknown scale loads associated with each output area.

We propose to use the following constraints on the $m_{t;i}$:

1. *sample size constraint*:

$$\sum_{i \in \boldsymbol{S}_{th}} m_{t;h;i}\ \pi_{t;h;i} = n_{t;h}, \qquad h = 1, 2, ..., H, \tag{4.5}$$

2. *unknown parameter constraint*:

$$\sum_{i \in \boldsymbol{S}_t} m_{t;i}\ \widehat{g}(y_{t;i}, x_{t;i}, \theta_t) = 0\cdot \tag{4.6}$$

Note that constraint (4.5) is defined at the stratum level, while constraint (4.6) is defined at the sample level. Constraint (4.5) is a typical constraint used in the design-based empirical likelihood methodology under unequal probability sam-

pling designs (see chapter 2 and (Berger and De La Riva Torres, 2016)). Constraint (4.6) includes the unknown population parameter θ_t. The $\widehat{g}(y_{t;i}, x_{t;i}, \theta_t)$ is an estimate of $g(y_{t;i}, x_{t;i}, \theta_t)$ for each output area; that is,

$$\widehat{g}_{t;i}(y_{t;i}, x_{t;i}, \theta_t) = n_{t;i}^{-1} \sum_{k \in \boldsymbol{S}_{t;i}} \pi_{t;i;k}^{-1} \; g(y_{t;i;k}, x_{t;i;k}, \theta_{t;i}), \tag{4.7}$$

where

$$g(y_{t;i;k}, x_{t;i;k}, \theta_{t;i}) = x_{t;i;k} - \theta_{t;i} \; y_{t;i;k}. \tag{4.8}$$

Constraint (4.6) is based on the ultimate cluster approach (Hansen et al., 1953). Using this approach for empirical likelihood inference in complex sampling designs has been proposed by Oguz-Alper and Berger (2016).

As it was discussed in the previous chapters, empirical likelihood can handle additional calibration (Deville and Särndal, 1992a) constraints based on a known population level characteristic, e.g. a total or a mean of a variable which is also measured for the sampled units. A natural variable which can be used to construct a calibration type constraint is the census count for each output area. Note that in order to define this constraint we only need to know the total or mean census count in Estimation Area U_t and the census counts within the sampled output areas.

We denote the known mean number of persons enumerated in the census in Estimation Area U_t by Ψ_{U_t}. Parameter Ψ_{U_t} can be expressed as the solution of the following estimating equation:

$$\sum_{i \in U_t} f(x_{t;i}, \Psi_{U_t}) = 0, \tag{4.9}$$

where $f(x_{t;i}, \Psi_{U_t}) = x_{t;i} - \Psi_{U_t}$.

Translating (4.9) into a sample-level constraint gives

$$\sum_{i \in \boldsymbol{S}_t} m_{t;i} \; f(x_{t;i}, \Psi_{U_t}) = 0 \cdot \tag{4.10}$$

Constraint (4.10) is optional, but it is likely to improve precision of the estimator of the target variable. Note that we use the raw (unadjusted) census count in the constraint. Therefore, while the census count itself may be impacted by under-enumeration, the calibration constraint is not, in that the parameter Ψ_{U_t} is indeed the population total of values $x_{t;i}$.

Let $\widehat{m}^*_{t;i}(\theta_t)$ be the vector of values which maximise expression (4.4), for a given vector θ_t, under $m_{t;i} > 0$ and constraints (4.5), (4.6) and (4.10). The *maximum empirical likelihood point estimator* of θ_{N_t} is defined as the value $\widehat{\theta}_t$ which maximises the following function:

$$\ell(\theta_t) = \sum_{i \in \boldsymbol{S}_t} \log\{m^*_{t;i}(\theta_t | \Psi_{U_t})\} \cdot \tag{4.11}$$

Following an argument presented by Berger and De La Riva Torres (2016), we notice that the estimator $\widehat{\theta}_t$ is also given by the unique solution of the sample estimating equation:

$$\widehat{G}(\theta) = \sum_{i \in \boldsymbol{S}_t} \widehat{m}_{t;i}(\Psi_{U_t}) \widehat{g}(y_{t;i}, x_{t;i}, \theta_t) = 0, \tag{4.12}$$

where the vector $\{\widehat{m}_{t;i}(\Psi_{U_t}) : i \in \boldsymbol{S}\}$ maximises function (4.4) under constraints (4.5) and (4.10) (see chapter 2.5 for further explanation).

Following (4.7) and (4.8), the solution $\widehat{\theta}_t$ can be derived as:

$$\widehat{\theta}_t = \left\{ \sum_{i \in \boldsymbol{S}_t} \left(\widehat{m}_{t;i} n_t^{-1} \sum_{k \in \boldsymbol{S}_{t;i}} \pi^{-1}_{t;i;k} x_{t;i;k} \right) \right\} \left\{ \sum_{i \in \boldsymbol{S}_t} \left(\widehat{m}_{t;i} n_t^{-1} \sum_{k \in \boldsymbol{S}_{t;i}} \pi^{-1}_{t;i;k} y_{t;i;k} \right) \right\}^{-1} \tag{4.13}$$

The design-based empirical likelihood ratio statistic is defined as (Berger and De

La Riva Torres, 2016):

$$\hat{r}(\theta_t) = 2\left\{\ell(\widehat{\boldsymbol{m}}) - \ell(\theta|\Psi_{U_t})\right\}, \tag{4.14}$$

where $\ell(\widehat{\boldsymbol{m}}) = \sum_{i\in\boldsymbol{S}_t} \log\{\widehat{m}_{t;i}(\Psi_{U_t})\}$ and $\ell(\theta)\sum_{i\in\boldsymbol{S}_t} \log\{\widehat{m}^*_{t;i}(\theta_t|\Psi_{U_t})\}$. Note that when a benchmark constraint is used,function (4.14) depends also on the known parameter Ψ_{U_t}. The statistic (4.14) is pivotal and follows a χ^2 distribution with one degree of freedom asymptotically when $\theta_t = \theta_{Ut}$ (Berger and De La Riva Torres, 2016). We use the empirical likelihood ratio statistic to construct confidence intervals for the parameter θ_t by taking the values θ_t such that $r(\theta_t) < \chi^2_{df=1;\alpha}$, where $\chi^2_{df=1;\alpha}$ is the upper α-quantile of the χ^2 distribution with 1 degree of freedom.

4.4 Numerical illustration

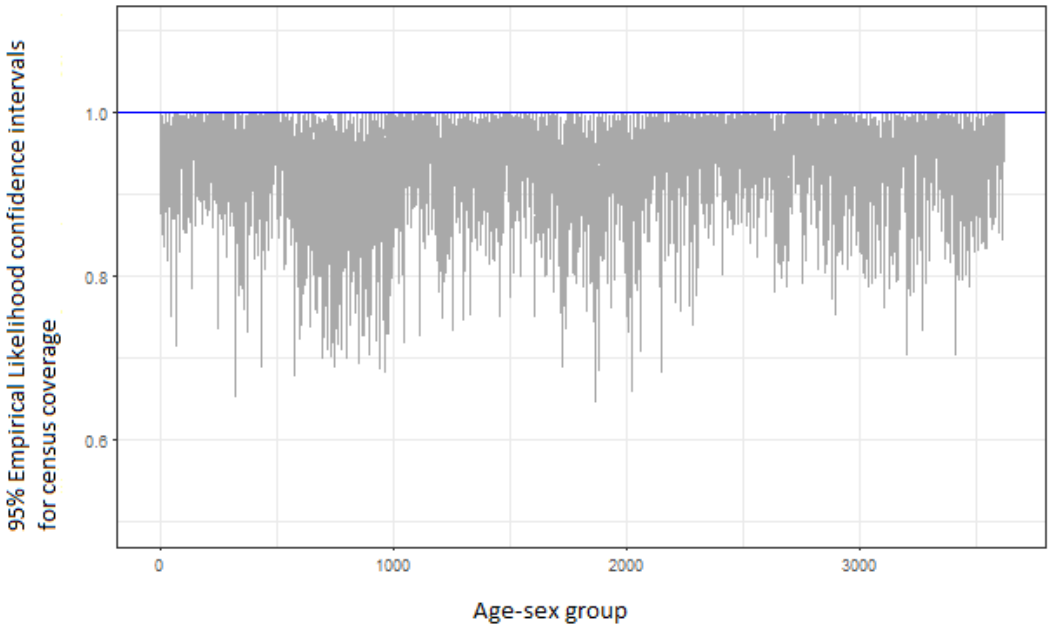
In this chapter we discuss results of applying the proposed approach to the 106 Estimation Areas enumerated in the 2011 England and Wales census. For comparison, we also constructed symmetric confidence intervals for the census coverage based on three variance estimation approaches: linearisation, jackknife and Canty and Davison's (1999) bootstrap with 100 replicates.

We estimate the census coverage in 35 age-sex groups within each of the Estimation Areas. The sample sizes within each Estimation Area by age-sex group are quite small, ranging from 17 to 118 with an average of 44.5 output areas. The estimated census coverage ranges from 72% to 100%, with several estimates very close to 100%. No auxiliary information is used.

The main purpose of this numerical study was to confirm that the empirical likelihood approach yields confidence intervals within the desired range and to see how the empirical likelihood confidence intervals compare to the symmetric confi-

dence intervals. The study confirmed that indeed, empirical likelihood confidence intervals never exceed 1. This can be seen in Figure 4.2, which shows 95% empirical likelihood confidence intervals obtained for different Estimation Areas and age-sex groups. For confidentiality reasons, the age-sex groups are not named on the graph and are labelled by meaningless consecutive integers.

Figure 4.2: Empirical likelihood 95% confidence intervals for the census coverage in different Estimation Areas and age-sex groups.



Many of the empirical likelihood confidence intervals are asymmetric. Figure 4.3 shows some examples of empirical likelihood confidence intervals and symmetric confidence intervals calculated for the same groups based on the jackknife variance estimator. We can see some cases when the upper bound of the symmetric confidence interval exceeds 1, while the empirical likelihood confidence interval remains within the $(0, 1)$ limits. The lower bounds of the empirical likelihood confidence intervals are sometimes lower than the lower bounds of the symmetric confidence intervals. In these cases the empirical likelihood confidence interval is also clearly asymmetric. Overall, the average width of empirical likelihood confidence intervals is similar to the average width of the symmetric confidence intervals.

Figure 4.3: Empirical likelihood (EL) and Symmetric (jackknife) (SYM) 95% confidence intervals for the census coverage in selected age-sex groups.

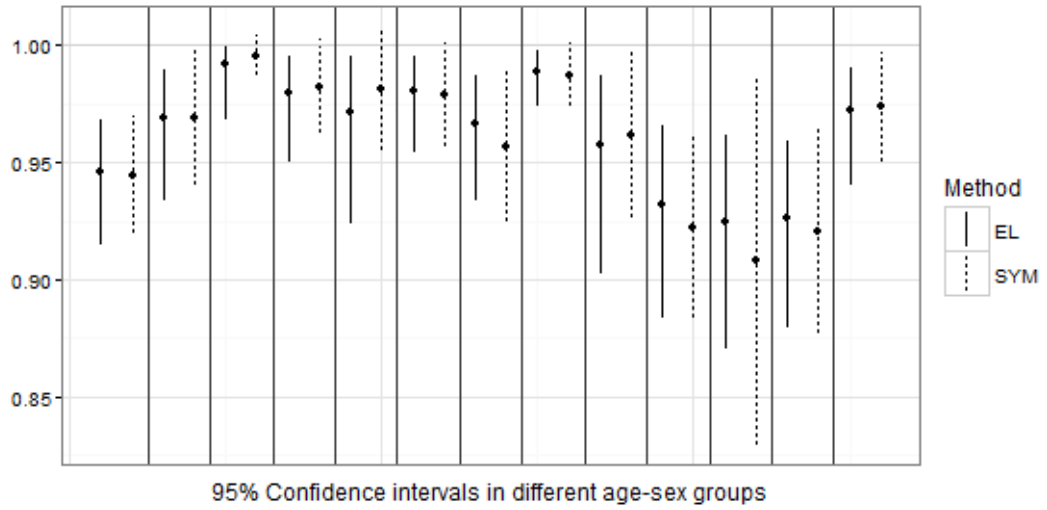
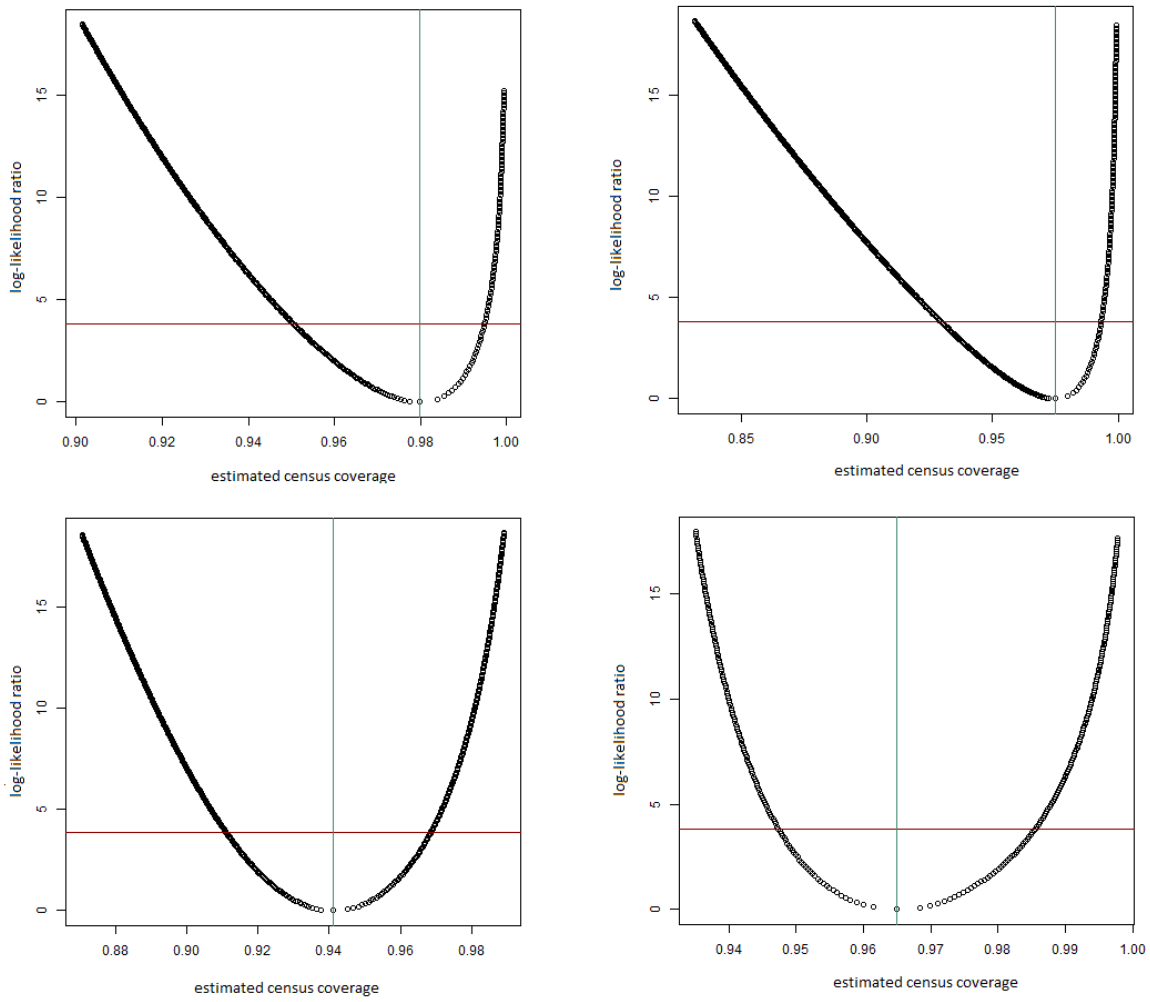


Figure 4.4 shows some examples of the empirical likelihood ratio function (4.14) plotted as a function of θ, based on data from the 2011 England and Wales census. The horizontal lines correspond to the threshold $\chi^2_\alpha = 3.84$. The vertical lines show the point estimate obtained at the minimum of the log-likelihood ratio function. The shape of function (4.14) depends on the distribution of the sample data, the sampling weights, the constraints imposed by the parameter space and, if used, any additional constraints. The functions presented on the graphs in the top row of figure 4.4 give highly asymmetric confidence intervals, with the 'upper' parts much shorter than the 'lower' parts. The function presented in the bottom left corner yields an approximately symmetric confidence interval. The function presented in the bottom right corner gives a slightly asymmetric confidence interval, with the 'upper' part longer than the 'lower' part.

Figure 4.4: Examples of the log-likelihood ratio plotted as a function of the point estimate for census coverage in some selected age-sex groups.



4.5 Simulation study

Applying empirical likelihood to real data allows us to see how the empirical likelihood confidence intervals are shaped in a realistic situation. However, in order to assess the coverage of the confidence intervals, we need to apply the proposed method to a population with a known value of the parameter of interest. Therefore in this chapter we evaluate the performance of the proposed method in a series of simulation studies on synthetic populations. The simulation is designed

to resemble the design described in chapter 4.2.1. We use the sample data from the 2011 census coverage survey together with the population sizes as enumerated in the 2011 England and Wales census. This dataset is used to create a number of synthetic populations which are then used to evaluate the performance of the proposed empirical likelihood estimator. Note that the synthetic populations are not aimed to represent the true composition of the Estimation Areas and should not be interpreted as such. We select four Estimation Areas of varying census coverage to create synthetic populations: Kensington and Chelsea; Southwark; Cornwall and Isles of Scilly; and Merseyside. The 2011 census coverages of these Estimation Areas are respectively 85.4%, 87.2%, 93.6% and 93.2%. Within each Estimation Area, there are 35 age-sex groups which also vary hugely in terms of census coverage.

Modelling a synthetic population based on the sample data has to follow two steps: creating a number of synthetic postcodes and creating synthetic output areas. First, within each output area, ten synthetic postcodes are generated. We generate three variables: the number of people enumerated in the census only, the number of people enumerated in the census coverage survey only and the number of people enumerated in both census and census coverage survey are generated from normal distributions with means and standard deviations equal to those observed in the actual sample of postcodes. The generated values are then rounded to integers, as they represent numbers of people. Due to the random procedure of generating values, it is possible that a postcode with no people enumerated in either census or census coverage survey is generated. To avoid this, any 0 values are replaced by 1. After the first step, our synthetic population has only as many output areas as were available in the sample, but each of the output areas consists of exactly ten synthetic postcodes. In the second step, synthetic output areas are created. Each of the output areas obtained in step 1 is replicated 100 times. Then values of each of the variables are modified by adding a random noise generated from a normal distribution with mean and standard deviation set to 3% of the mean and standard deviation observed in the output area. Note that the random

error is always positive. This creates populations with slightly higher counts than in the original sample data, but allows us to avoid problems with very low counts at stratum level. We do not change the stratifying variables, that is, the synthetic output areas have the same hard to count index and the same local authority as the output area that was used to generate them.

Following the sampling design of the census coverage survey, the synthetic Estimation Areas are stratified by the hard to count index and local authority. Then two stage cluster sampling without replacement is used, with output areas as the primary sampling units and postcodes as the secondary sampling units. In each of the strata, a sample of 5% of the output areas is selected by simple random sampling without replacement. Within each of the sampled output areas, 5 postcodes are selected by simple random sampling without replacement. Selecting 50% of postcodes corresponds with the design of the census coverage survey.

We select 1,000 samples in each of the synthetic populations. Census coverage, as well as the lower and upper bounds of the confidence intervals for the census coverage, are estimated from each of the samples. We use four methods of obtaining confidence intervals. First, we apply the proposed empirical likelihood methodology and obtain empirical likelihood confidence intervals, with a benchmark constraint on the mean census count. Second, we calculate symmetric confidence intervals around an estimator of the ratio of census count and population size using three methods of variance estimation: linearisation, Jackknife and Canty and Davison's (1999) bootstrap with 100 replicates. For computation of the symmetric confidence intervals of the ratio estimator, we use the svyratio function from the survey package (Lumley et al., 2004; Lumley, 2016). The bootstrap method used draws a sample of PSUs (Output Areas) from each stratum (see (Lumley and Lumley, 2018) and (Preston, 2009) for details). Alternatively, Preston's (2009) multi-stage rescaled bootstrap method could be used.

Figures 4.5, 4.6, 4.7 and 4.8 show the observed coverage level of confidence intervals of the four different methods for census coverage in different age-sex groups within

the tested Estimation Areas. The plots give an overview of how close the empirical coverage of confidence intervals obtained in various groups is to the nominal level of 95%.

All of the methods give confidence intervals with acceptable coverage levels, considering the relatively small sample sizes, complex sampling design and relatively high variability of the parameter of interest. All confidence intervals suffer from under or over-coverage in several age-sex groups. The empirical likelihood confidence intervals behave similarly to the symmetric confidence intervals and in some cases have coverage closer to the nominal level. For example, in Figure 4.5, the minimum level observed for the empirical likelihood confidence interval is 92.5%. For the other approaches, the level observed can be as low as 90%. The empirical likelihood confidence intervals are also, on average, marginally shorter than the symmetric confidence intervals, with average length of 5.33% for empirical likelihood intervals and between 5.66% and 5.76% for symmetric intervals, even if they are truncated at 1. Figures 4.9 - 4.12 show the average length of confidence intervals obtained in each age-sex group of the four synthetic populations considered.

Figure 4.5: Coverage of empirical likelihood and symmetric confidence intervals in various age-sex groups, in population *synthIL06KENS*

Figure 4.6: Coverage of empirical likelihood and symmetric confidence intervals in various age-sex groups, in population *synthIL09SOUT*

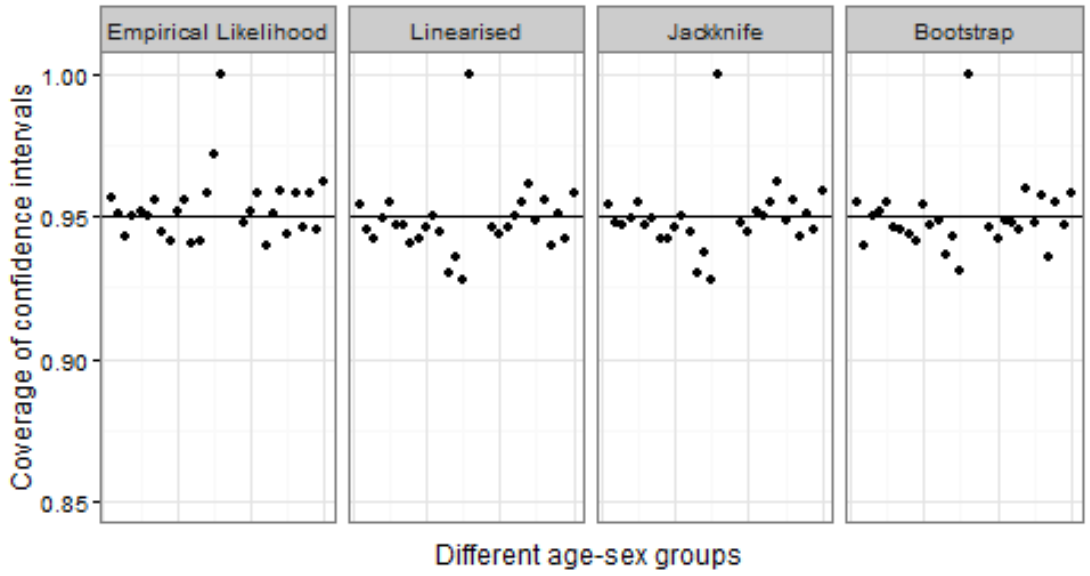


Figure 4.7: Coverage of empirical likelihood and symmetric confidence intervals in various age-sex groups, in population *synthSW03CORN*

Figure 4.8: Coverage of empirical likelihood and symmetric confidence intervals in various age-sex groups, in population *synthNW06MERS*

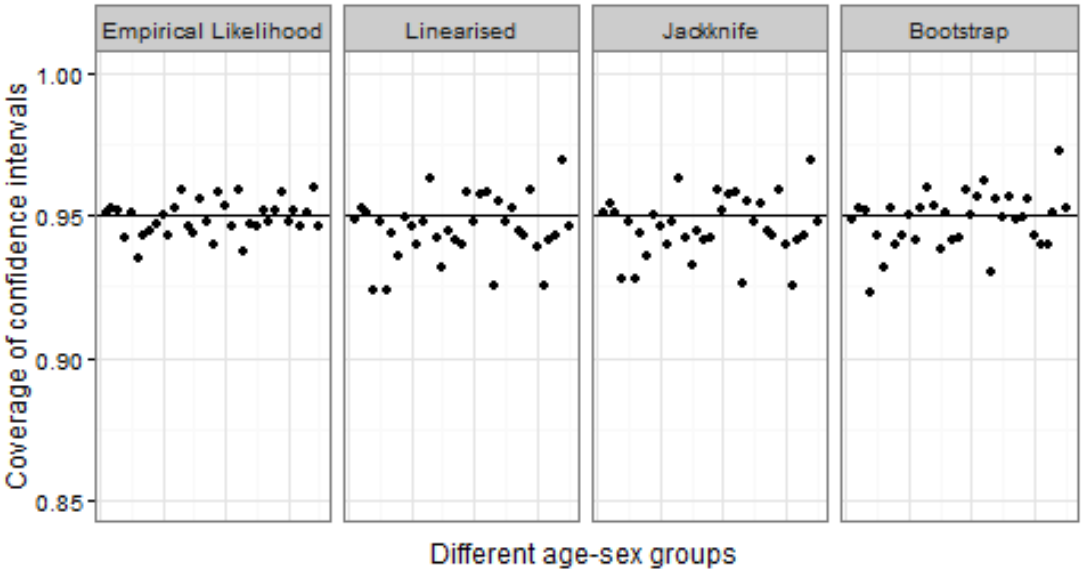


Figure 4.9: Average length of confidence intervals in various age-sex groups, in population *synthIL06KENS*

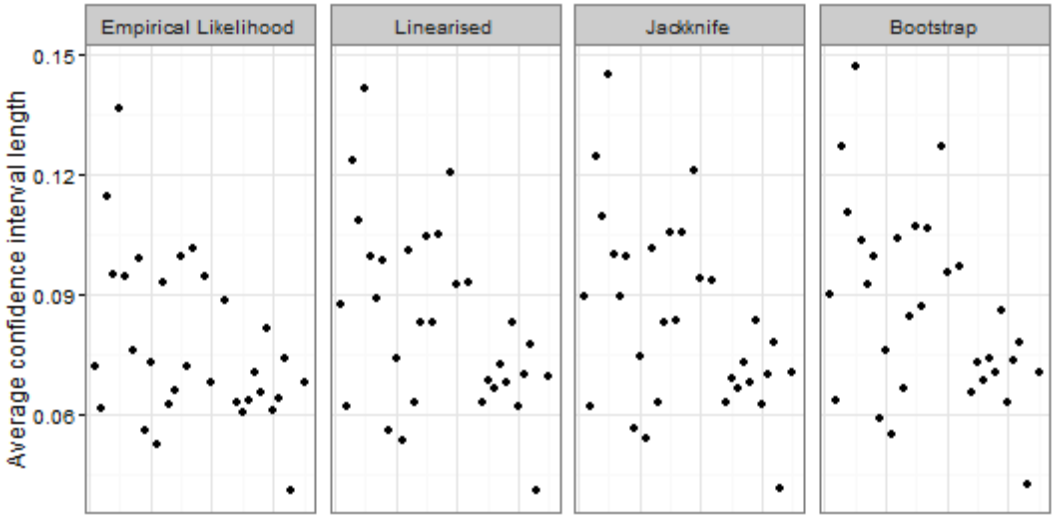


Figure 4.10: Average length of confidence intervals in various age-sex groups, in population *synthIL09SOUT*

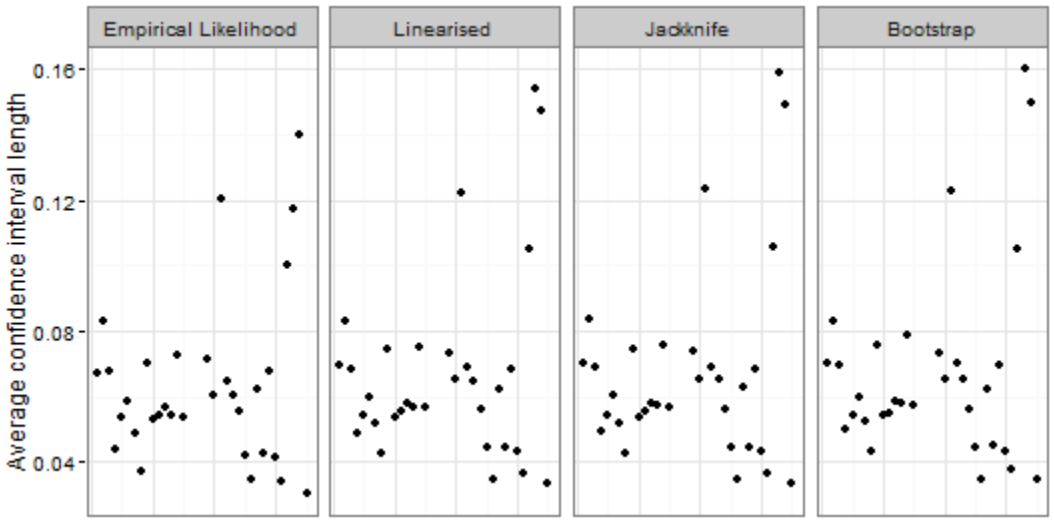


Figure 4.11: Average length of confidence intervals in various age-sex groups, in population *synthSW03CORN*

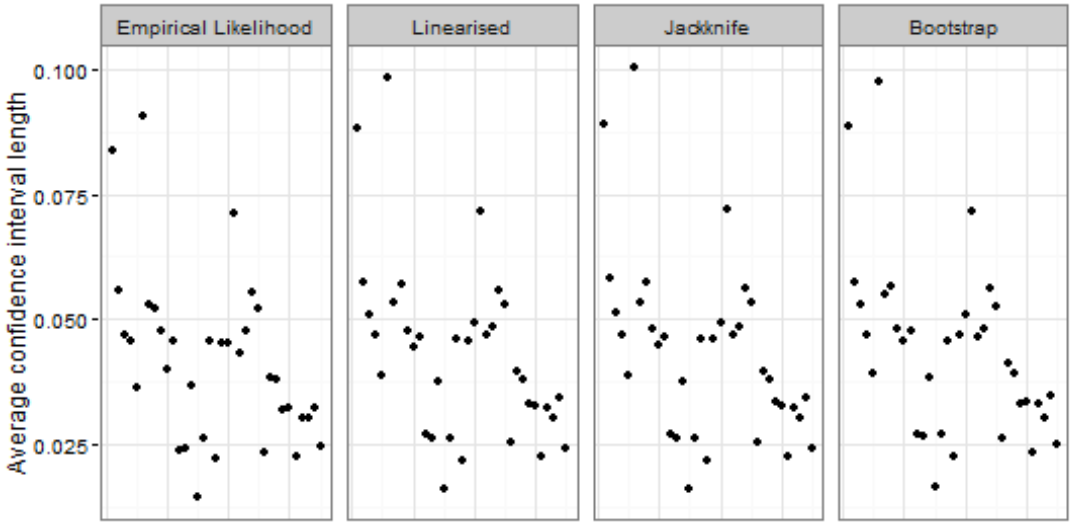
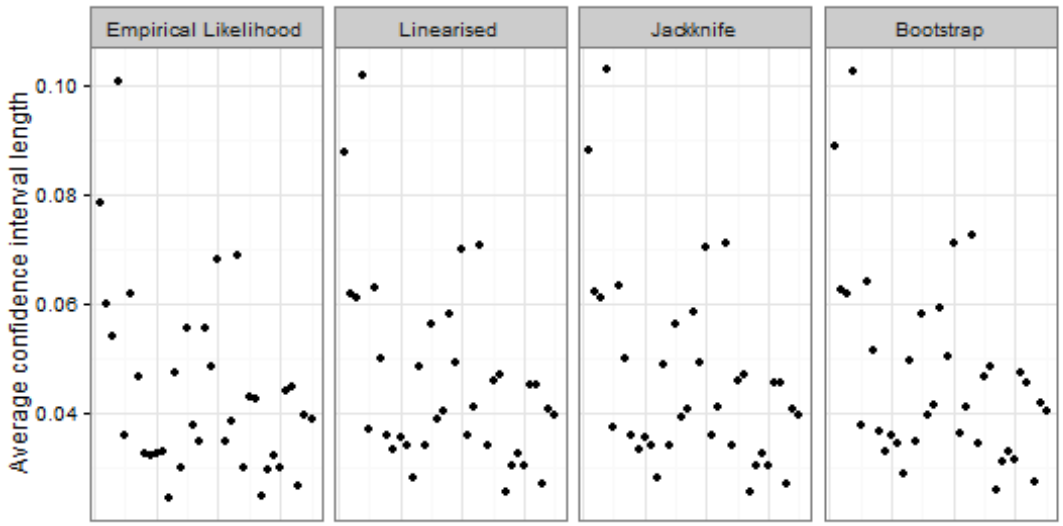


Figure 4.12: Average length of confidence intervals in various age-sex groups, in population *synthNW06MERS*



4.6 Conclusions

The main practical advantage of the empirical likelihood confidence intervals in relation to census coverage estimation, or binomial parameter estimation more generally, is that they do not exceed 1. It is worth noting that the numerical algorithm used to obtain empirical likelihood weights does not involve any explicit constraints on the upper bound of the confidence intervals. Instead, the empirical likelihood ratio function naturally yields confidence intervals and point estimates lower than 1. It is of course possible to trim the symmetric confidence intervals at 1, which does not influence their asymptotic coverage, as the true parameter value can never exceed 1. However, the lower bounds of the symmetric confidence intervals might then still remain too high, as they do not account for different levels of variability of the data on the two sides of the point estimate. The empirical likelihood confidence intervals are based on the likelihood ratio function which is defined by the shape of the sample data. They will, therefore, correctly account for larger variability below the point estimate than above it, if this is the case.

Empirical likelihood allows to easily incorporate calibration type constraints, which might be constructed using an arbitrarily chosen function of the known parameters. In particular, the function can be selected so as to maximise the correlation between the calibration variable and the parameter of interest.

The numerical simulations presented in the previous chapter show that empirical likelihood confidence intervals indeed remain within the range of the parameter of interest. The coverage of empirical likelihood confidence intervals is comparable to the coverage of the tested symmetric confidence intervals. The empirical likelihood confidence intervals also have comparable width to the symmetric confidence intervals. However, we can see a few examples when the confidence intervals are clearly asymmetric and the lower bound of the empirical likelihood confidence interval is lower than that of the symmetric confidence interval. This suggests that empirical likelihood confidence intervals might be well suited for generation

of confidence intervals when variables are highly skewed.

The approach presented in this paper was designed specifically as an alternative for the confidence interval estimation at the census coverage estimation stage in the current coverage estimation procedure. That is, it is applied to the population sizes estimated through DSE in order to obtain population level estimates of census coverage. In line with the current approach, it treats the dual system estimates as fixed. Estimating the uncertainty around the DSE estimates and incorporating it into the calculation of the empirical likelihood confidence intervals would be an interesting direction of future research.

We should note that there are multiple alternative ways of estimating the population size than the Dual System Estimator. For example, Chipperfield et al. (2017) describe a method applied by the Australian Bureau of Statistics, where the population size is estimated through a generalized regression type model, called the PREG (population regression), accounting for over and under coverage in census and considering non-response. Zhang (2015) gives a comprehensive overview of modelling approaches for undercoverage and overcoverage in registers. Moreover, triple system estimators, where administrative registers are used alongside a census and a post-enumeration survey, have been proposed. This allows the dependence between response probabilities in the census and census coverage survey to be modelled (see (Baffour et al., 2013) and (Griffin, 2014)). Extending empirical likelihood to work with such methods would require considerable developments, especially if any modelling was involved.

Finally, empirical likelihood is of course not the only method which gives asymmetric and range-preserving confidence intervals. Investigating the properties of other such methods, for example bootstrap-based approaches, in the context of estimation of census coverage, would be an interesting future research direction.

Chapter 5

Numerical aspects of empirical likelihood

Empirical Likelihood estimation procedure involves numerical operations at several stages. Specifically, numerical optimisation (finding minimum or maximum of a function) and root finding methods are applied in order to:

1. obtain the vectors of adjusted weights $\widehat{m}_i(\boldsymbol{\varphi}_{U})$ and $\widehat{m}_i^{\star}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{U})$, such that the constraints based on sample data and population parameters are met,

2. obtain the point estimate $\hat{\boldsymbol{\theta}}$,

3. obtain the lower and upper bounds of confidence intervals, which requires evaluating the log-likelihood ratio function for multiple values of the parameter of interest $\boldsymbol{\theta}$.

In this chapter we discuss the computational aspects of these numerical operations. We consider the three computational tasks listed above in a general way, so that the discussion presented here is relevant to all three empirical likelihood applications discussed in the previous chapters and to design-based empirical likelihood estimation in general. We keep in mind the specific conditions imposed by the problems discussed in chapters 2, 3 and 4, such as a typically large dimension of the constraint matrix or confidence interval bounds laying close to the boundary of the parameter space.

5.1 Obtaining the vectors of adjusted weights $\widehat{m}_i(\boldsymbol{\varphi}_U)$ and $\widehat{m}_i^{\star}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)$

Finding the adjusted weights $\widehat{m}_i(\boldsymbol{\varphi}_U)$ is, in principle, a task of optimisation under constraints. The task consists of finding a vector of values $\widehat{m}_i(\boldsymbol{\varphi}_U)$ which maximise the value of a function $\ell(\boldsymbol{m})$ and are such that a constraint of the form $\boldsymbol{m}^\top \boldsymbol{c}^\top = \boldsymbol{C}^\top$ holds, where $\boldsymbol{c}$ is a matrix and $\boldsymbol{C}$ is a column vector. These adjusted weights $\widehat{m}_i(\boldsymbol{\varphi}_U)$ might be of interest themselves, but are also necessary to obtain point estimates and to evaluate the log-likelihood ratio function $\widehat{r}(\boldsymbol{\theta}|\boldsymbol{\varphi}_U)$, which is required for calculation of confidence intervals. The vector of adjusted weights $\widehat{m}_i^{\star}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)$, which maximises $\ell(\boldsymbol{m})$ under the extended system of constraints $\boldsymbol{m}^\top \boldsymbol{c}^{\star}(\boldsymbol{\theta})^\top = \boldsymbol{C^*}^\top$, where $\boldsymbol{c}$ is a sub-matrix of $\boldsymbol{c}^{\star}(\boldsymbol{\theta})$ and $\boldsymbol{C}$ is a sub-vector of $\boldsymbol{C^*}$, is found in an analogous way. We explain the numerical aspects of this estimation process using $\widehat{m}_i(\boldsymbol{\varphi}_U)$ as an example.

Following Berger and De La Riva Torres (2016), Wu (2004*a*) and other authors, we use the Lagrange's multipliers method to solve the optimisation problem, which is therefore translated into solving a system of non-linear equations of the following general form:

$$\sum_{i \in \boldsymbol{S}} (\boldsymbol{\lambda}^\top \boldsymbol{c}_i)^{-1} \boldsymbol{c}_i - \boldsymbol{C} = \boldsymbol{0}, \tag{5.1}$$

where $\boldsymbol{\lambda}$ is a vector of Lagrange's multipliers. The equation (5.1) is solved with respect to $\boldsymbol{\lambda}$. The number of unknowns (the Lagrange's multipliers) is the same as the dimension of vector $\boldsymbol{C}$ and in practical applications can vary between 1 and a few hundreds. The equation (5.1) is easily derived if we write the Lagrange function for maximization of $\ell(\boldsymbol{m})$ in the following form:

$$Q(\boldsymbol{\lambda}) = \sum_{i \in \boldsymbol{S}} \log(m_i) - \boldsymbol{\lambda}^\top (\sum_{i \in \boldsymbol{S}} m_i \boldsymbol{c_i} - \boldsymbol{C}) \cdot \tag{5.2}$$

We find the adjusted weights by solving the following equation:

$$\frac{dQ}{dm_i} = \frac{1}{m_i} - \boldsymbol{\lambda}^\top \boldsymbol{c}_i = \mathbf{0}, \tag{5.3}$$

which gives

$$\widehat{m}_i(\boldsymbol{\varphi}_U) = (\boldsymbol{\lambda}^\top \boldsymbol{c}_i)^{-1}. \tag{5.4}$$

Equation (5.1) is obtained after substituting (5.4) in the following equation:

$$\frac{dQ}{d\boldsymbol{\lambda}} = \sum_{i \in \boldsymbol{S}} \widehat{m}_i(\boldsymbol{\varphi}_U)\boldsymbol{c_i} - \boldsymbol{C} = \mathbf{0}. \tag{5.5}$$

Equation (5.1) is often solved through application of a modified version of the Newton - Raphson algorithm. The $k+1$-th iteration of the Newton-Raphson algorithm for solving the system of non-linear equations (5.1) consists of taking

$$\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k - \{Q'(\boldsymbol{\lambda}_k)\}^{-1} Q(\boldsymbol{\lambda}_k), \tag{5.6}$$

where $Q'(\boldsymbol{\lambda}_k)$ is the Jacobian of $Q(\boldsymbol{\lambda})$ calculated at $\boldsymbol{\lambda} = \boldsymbol{\lambda}_k$.

The Jacobian $Q'(\boldsymbol{\lambda}_k)$ is a p by p matrix, where p is the dimension of $\boldsymbol{\lambda}$, with the $(j;h)$-th element equal to

$$\begin{aligned} J_{(j;h)} &= \frac{\delta q_j(\boldsymbol{\lambda})}{\delta \lambda_h} && (5.7)\\ &= \sum_{i \in \boldsymbol{S}} c_{i;j} \frac{\delta}{\delta \lambda_h} (\lambda_1 c_{i;1} + \lambda_2 c_{i;2} + ... + \lambda_p c_{i;p})^{-1} && (5.8)\\ &= -\sum_{i \in \boldsymbol{S}} \frac{c_{i;j} c_{i;h}}{(\lambda_1 c_{i;1} + \lambda_2 c_{i;2} + ... + \lambda_p c_{i;p})^2}, && (5.9) \end{aligned}$$

where $q_j(\boldsymbol{\lambda})$ is the j-th element of vector (5.2), λ_j is the j-th element of the vector of Lagrange's multipliers $\boldsymbol{\lambda}$ and $c_{i;j}$ is the element in the i-th row and j-th column of the matrix $\boldsymbol{c}$.

In practice, it is not necessary to invert the Jacobian $Q'(\boldsymbol{\lambda}_k)$. Instead, a linear

equation

$$Q(\boldsymbol{\lambda}_k) + Q'(\boldsymbol{\lambda}_k)(\boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k) = \mathbf{0} \tag{5.10}$$

is solved with respect to $\boldsymbol{\delta} = \boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k$ and then $\boldsymbol{\lambda}_{k+1} = \boldsymbol{\delta} - \boldsymbol{\lambda}_k$ is calculated.

The algorithm (5.6) converges locally with the quadratic rate. One of the possible modifications of the Newton-Raphson algorithm which ensures global convergence consists of decreasing the step size by multiplying it by a factor q, where $0 < q < 1$ (Polyak, 1987).

The approach consisting of modifying the step size in the Newton - Raphson algorithm was applied in statistics to ensure that the solution to equation (5.1) yields non-negative weights. This method is often referred to as the *Polyak correction*. In particular, Chen et al. (2002) proposed an algorithm which adjusts the step in the Newton - Raphson method by a parameter $q_k = k^{-\frac{1}{2}}$, where k is the number of the current iteration. If the solution found in the k-th step yields negative weights, the parameter q is adjusted to $q_{k+1} := q_k/2$ and the condition is checked again. These steps are repeated until the non-negativity condition is met. Chen et al. (2002) further extended this method to accommodate weights range restrictions in presence of benchmark constraints, under pseudo empirical likelihood. This is obtained by iterative relaxation of benchmark constraints until the obtained weights are within a desired range.

The algorithm of Chen et al. (2002) was used e.g. by Wu (2004*a*) to find pseudo empirical likelihood weights in two samples context and by Berger and De La Riva Torres (2016) to obtain design-based empirical likelihood weights from single sample complex designs. Wu (2004*b*) extended the algorithm of Chen et al. (2002) to pseudo empirical likelihood under stratified sampling.

The algorithms of Polyak (1987) and Chen et al. (2002) are guaranteed to converge to a unique solution, if such a solution exists. The Newton-Raphson algorithm, as well as its both adjustments described above, require calculation of the Jacobian (5.7) at each iteration. For large parameter sizes and large number of constraints

this becomes computationally expensive. In common survey sampling settings, the dimension of the parameter of interest is likely to be large, as is the number of benchmark constraints. If stratification is used, the size of the matrix of constraints is further increased by a design constraint created for every stratum. The computational complexity might be reduced by use of a quasi-Newton method instead of the Newton-Raphson algorithm. The quasi-Newton methods were invented to avoid calculation of the Jacobian at every step. A range of methods have been proposed. No method is considered to be the best for all purposes. Algorithms which perform well in some settings are known to perform poorly in others. In fact, as stated in the famous *no free lunch theorem for optimization* (Wolpert and Macready, 1997), a general-purpose universal optimisation strategy is impossible (Ho and Pepyne, 2001). For instance, there often is a trade-off between the speed of convergence and the sensitivity to poor choices of the starting point.

5.2 Obtaining the point estimate $\widehat{\boldsymbol{\theta}}$

Obtaining the point estimate for the parameter of interest $\boldsymbol{\theta}$ requires finding the value $\widehat{\boldsymbol{\theta}}$ which minimises the log-likelihood ratio function defined as $\widehat{r}(\boldsymbol{\theta}|\boldsymbol{\varphi}_U) = 2\left\{\ell(\widehat{\boldsymbol{m}}) - \ell(\boldsymbol{\theta}|\boldsymbol{\varphi}_U)\right\}$, where $\ell(\widehat{\boldsymbol{m}}) = \sum_{i\in\boldsymbol{S}} \log(\widehat{m}_i(\boldsymbol{\varphi}_U))$ and $\ell(\boldsymbol{\theta}|\boldsymbol{\varphi}_U) = \sum_{i\in\boldsymbol{S}} \log(\widehat{m}_i^{\star}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U))$, for a given vector $\widehat{m}_i(\boldsymbol{\varphi}_U)$. The estimates $\widehat{m}_i(\boldsymbol{\varphi}_U)$ do not depend on $\boldsymbol{\theta}$ and only need to be calculated once for given sample data and a system of constraints. The vector of values $\widehat{m}_i^{\star}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)$ depends on $\boldsymbol{\theta}$ in that it is a result of maximising $\ell(\boldsymbol{\theta}|\boldsymbol{\varphi}_U)$ under a system of constraints which include $\boldsymbol{\theta}$.

The parameter estimate $\widehat{\boldsymbol{\theta}}$ can be found in two ways. First, a numerical optimisation algorithm may be applied directly to the empirical log-likelihood ratio function $\widehat{r}(\boldsymbol{\theta}|\boldsymbol{\varphi}_U)$. A simpler solution, however, is found by translating this optimisation into a root finding problem. Functions $\ell(\widehat{\boldsymbol{m}})$ and $\ell(\boldsymbol{\theta}|\boldsymbol{\varphi}_U)$ differ in the

additional constraint imposed on the admissible values of the vector of adjusted weights $\widehat{m}_i^{\star}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{U})$. This constraint takes the general form

$$\sum_{i \in \boldsymbol{S}} \widehat{m}_i^{\star}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{U}) \; \boldsymbol{g}_i(\boldsymbol{\theta}) = \mathbf{0}_{\nu} \cdot \tag{5.11}$$

Berger and Kabzinska (2017) showed that the value $\widehat{\boldsymbol{\theta}}$ that is the unique solution of the equation

$$\widehat{\boldsymbol{G}}(\boldsymbol{\theta}) = \sum_{i \in \boldsymbol{S}} \widehat{m}_i(\boldsymbol{\varphi}_{U}) \; \boldsymbol{g}_i(\boldsymbol{\theta}) = \mathbf{0}_{\nu}, \tag{5.12}$$

is also the value that maximises $\ell(\boldsymbol{\theta}|\boldsymbol{\varphi}_{U})$ and minimises $\widehat{r}(\boldsymbol{\theta}|\boldsymbol{\varphi}_{U})$ (see chapter 2.5 for details). This means that, in practice, it is not necessary to estimate the values $\widehat{m}_i^{\star}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{U})$ in order to obtain the point estimate $\widehat{\boldsymbol{\theta}}$, as this can be obtained from equation (5.12) which only contains $\widehat{m}_i(\boldsymbol{\varphi}_{U})$.

In simple cases, e.g. when $\boldsymbol{\theta}$ is a total, a mean or a ratio, for a given vector of adjusted weights $\widehat{m}_i(\boldsymbol{\varphi}_{U})$, equation (5.12) can be solved analytically. Otherwise, e.g. when $\boldsymbol{\theta}$ is a vector of quantiles, a root finding algorithm has to be applied to solve (5.12). The complexity of this task depends on the type of the parameter $\boldsymbol{\theta}$ used.

Note that if the function $\widehat{r}(\boldsymbol{\theta}|\boldsymbol{\varphi}_{U})$ is minimised directly, the vector of adjusted weights $\widehat{m}_i^{\star}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{U})$ has to be estimated for each candidate value of $\widehat{\boldsymbol{\theta}}$. This is much more computationally demanding than solving the estimating equation (5.12).

5.3 Obtaining the lower and upper bounds of confidence intervals

Let us consider constructing a confidence interval for a scalar parameter of interest first. The upper and lower bounds of an empirical likelihood confidence interval

are the values of the parameter of interest $\boldsymbol{\theta}$ such that $\widehat{r}(\boldsymbol{\theta}|\boldsymbol{\varphi}_U) = \chi^2$, where χ^2 is the limiting value from the χ^2 distribution. Finding these two points requires evaluating $\widehat{r}(\boldsymbol{\theta}|\boldsymbol{\varphi}_U)$ for different candidate values $\widehat{\boldsymbol{\theta}}$ in the neighbourhood of the lower and upper bounds, so that the equation

$$\widehat{r}(\boldsymbol{\theta}|\boldsymbol{\varphi}_U) - \chi^2 = \mathbf{0} \tag{5.13}$$

is solved with respect to $\boldsymbol{\theta}$. The function $\widehat{r}(\boldsymbol{\theta}|\boldsymbol{\varphi}_U)$ is convex. However, this is not a straightforward root finding problem. In fact, it requires solving two nested root finding problems. Evaluation of $\widehat{r}(\boldsymbol{\theta}|\boldsymbol{\varphi}_U)$ for a specific value of $\boldsymbol{\theta}$ requires finding the adjusted weights $\widehat{m}_i^\star(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)$ and comparing them with the previously estimated values $\widehat{m}_i(\boldsymbol{\varphi}_U)$ (see chapter 5.1). Note that while $\widehat{m}_i(\boldsymbol{\varphi}_U)$ do not depend on the parameter of interest and hence only need to be estimated once (i.e., for a given sample and system of benchmark and consistency constraints, there is one vector $\widehat{\boldsymbol{m}}$), the vector $\widehat{m}_i^\star(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)$ vary in function of $\boldsymbol{\theta}$. Therefore, for every candidate value $\boldsymbol{\theta}$, the equation

$$\sum_{i \in \boldsymbol{S}} \widehat{m}_i^\star(\boldsymbol{\theta}, \boldsymbol{\varphi}_U) \boldsymbol{c}_i^\star(\boldsymbol{\theta}) - \boldsymbol{C}^* = \mathbf{0} \tag{5.14}$$

has to be solved. Finding the vector of weights $\widehat{m}_i^\star(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)$ for different candidate values of $\boldsymbol{\theta}$ is the most computationally expensive part of empirical likelihood estimation. In particular, when the candidate value of the parameter $\boldsymbol{\theta}$ differs substantially from the true parameter value $\boldsymbol{\theta}_U$, solving (5.14) may require many iterations. Note that for the solution $\widehat{m}_i^\star(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)$ to exist, we assume that $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}_U$ are such that $\boldsymbol{C}^\star$ is an inner point of the convex hull formed by the sample observations $\{\boldsymbol{c}_i^\star(\boldsymbol{\theta}) : i \in \boldsymbol{S}\}$ (see e.g. chapter 2.4). Therefore, if a candidate value $\boldsymbol{\theta}$ that is outside of the convex hull formed by the sample observations is taken, the solution will not be found at all. This can occur in practice e.g. in the census coverage estimation problem discussed in chapter 4, when the upper bound of the confidence interval is close to 1.

Suitable selection of the candidate values of $\boldsymbol{\theta}$ is crucial for the performance of the algorithm. The search space has to be sufficiently larger than the actual

confidence interval so that the bounds can be found, but ideally not too much larger as the farther we get from the confidence interval, the more difficult the evaluation of $\widehat{r}(\boldsymbol{\theta}|\boldsymbol{\varphi}_U)$ is. Experience from obtaining the confidence intervals for different problems discussed in the earlier chapters suggests that restricting the search space on the inner side of the confidence interval (i.e., not searching in the closest neighbourhood of the point estimate), does not contribute much to the improvement in performance, because evaluating $\widehat{r}(\boldsymbol{\theta}|\boldsymbol{\varphi}_U)$ in this area is typically fast.

The problem of finding the confidence interval bounds is easier than a general case of finding roots of a function in that we know that we are searching for two points which lay on both sides of the point estimate. Therefore it is natural to take the point estimate as the starting point and search in both directions from it. At each side of the point estimate, $\widehat{r}(\boldsymbol{\theta}|\boldsymbol{\varphi}_U)$ is a strictly increasing function. It can, therefore, be evaluated for a selection of parameter values and then interpolated in between, e.g. by splines. Because we only need precise values in the neighbourhood of the confidence interval bounds, it is reasonable to adjust the distance between the points at which the log-likelihood ratio function is evaluated relative to how far we are from the bound. In most cases, bisection can be used to find the confidence interval bounds. However, bisection requires evaluating points that are outside of the confidence intervals. In some particularly difficult cases, such as the census coverage example, this might be computationally difficult, that is, finding the values $\widehat{m}_i^{\star}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)$ for values of $\boldsymbol{\theta}$ substantially different from $\widehat{\boldsymbol{\theta}}$ might require a very large number of iterations.

In such cases, we propose to use algorithm (1) described below. The algorithm is loosely inspired by bisection. We take the point estimate as the starting point and evaluate $\widehat{r}(\boldsymbol{\theta}|\boldsymbol{\varphi}_U)$ at each step, until the value of $\widehat{r}(\boldsymbol{\theta}|\boldsymbol{\varphi}_U)$ is larger than the limiting value from the χ^2 distribution. Then the step is decreased and the search continues in the opposite direction, again until the value of $\widehat{r}(\boldsymbol{\theta}|\boldsymbol{\varphi}_U)$ is at the opposite side of the limiting value from the χ^2 distribution. This procedure continues until

the parameter value which gives a value of $\widehat{r}(\boldsymbol{\theta}|\boldsymbol{\varphi}_U)$ close enough to the limiting value from the χ^2 distribution is found. The main difference between the proposed method and bisection is that we avoid evaluating the log-likelihood ratio function $\widehat{r}(\boldsymbol{\theta}|\boldsymbol{\varphi}_U)$ for values far outside of the confidence interval, where this might be slow or impossible. This is at the expense of evaluating the function at more points within the confidence interval.

The initial step value might be adjusted for the particular problem. If the initial step is too large, it might be difficult to obtain the vector of values $\widehat{m}_i^\star(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)$ for such an extreme parameter value. In such a case decreasing the step size should help. Too small step size will result in slow convergence. For well behaved functions $\widehat{r}(\boldsymbol{\theta}|\boldsymbol{\varphi}_U)$, a larger step might be selected and spline interpolation might be used to interpolate function values between these points. This is often faster than evaluating $\widehat{r}(\boldsymbol{\theta}|\boldsymbol{\varphi}_U)$, especially if the dimension of the constraints matrix is large.

Algorithm 1 Finding confidence interval bounds

1. Start with $\breve{\boldsymbol{\theta}}_k = \widehat{\boldsymbol{\theta}}$ and $\delta = \widehat{\boldsymbol{\theta}}10^{-4}$, $\tau = 1$, $\epsilon = 2.2210^{-16}$
2. While $|\widehat{r}(\breve{\boldsymbol{\theta}}_k|\boldsymbol{\varphi}_U) - \chi^2| > 0 + \epsilon$:
   (a) If $\widehat{r}(\breve{\boldsymbol{\theta}}_k|\boldsymbol{\varphi}_U) > \chi^2$, then $\tau = (-1)\tau$,
   (b) If $\widehat{r}(\breve{\boldsymbol{\theta}}_{k-1}|\boldsymbol{\varphi}_U) > \chi^2 > \widehat{r}(\breve{\boldsymbol{\theta}}_k|\boldsymbol{\varphi}_U)$ or $\widehat{r}(\breve{\boldsymbol{\theta}}_{k-1}|\boldsymbol{\varphi}_U) < \chi^2 < \widehat{r}(\breve{\boldsymbol{\theta}}_k|\boldsymbol{\varphi}_U)$, then $\delta = \delta/2$,
   (c) For upper bound: $\breve{\boldsymbol{\theta}}_{k+1} = \breve{\boldsymbol{\theta}}_k + \tau\delta$, for lower bound: $\breve{\boldsymbol{\theta}}_{k+1} = \breve{\boldsymbol{\theta}}_k - \tau\delta$,
   (d) $k = k + 1$

If the parameter of interest $\boldsymbol{\theta}_U$ is multidimensional, but the components are independent, i.e., each scalar parameter of interest is entirely defined by a single estimating equation, which does not contain any unknown parameters, a confidence interval for each scalar parameter can be obtained separately, by selecting the relevant estimating equation. This will yield a number of confidence intervals of which each has the nominal coverage and which do not depend on each other. For some applications a joint confidence region for a multidimensional parameter might be of interest. A joint confidence region has joint coverage equal to the nominal level. This confidence region can be found by applying the same principle

as for the search of confidence interval bounds. However, in this case we would be searching for a contour in a space rather than for two points on one axis. This makes the task much more demanding computationally.

5.4 Simulation study: execution times

In this section we show results of a small simulation study evaluating the execution times for calculation of empirical likelihood point estimates and confidence intervals in samples of different sizes. We use the same population as in chapter 2.12.3. We take domain (ii) *Distribution, hotels and restaurants* as the example test case. The parameters of interest, auxiliary variables and other conditions are the same as in chapter 2.12.3, that is, we select two independent samples and calculate an estimate of a total in he presence of a benchmark constraint and an alignment constraint.

Table 5.1 shows the distribution of the *user* components of the execution times for calculation of point estimates and confidence intervals for different sizes of the samples, obtained in 100 iterations. The *system* components of the execution times were negligible compared to the *user* times. A garbage collector operation was performed before calculation of each estimator. For comparison, we show the user execution times for calculation of the single sample calibration estimator, the composite regression estimator and the pseudoempirical likelihood estimator and for calculation of confidence intervals based on the variance estimator for the composite regression estimator. The pseudoempirical likelihood relies on a similar variance estimation method, therefore we expected results to be the same. The single sample calibration estimator operates on one sample and only considers the benchmark constraint, while all the other estimates use pooled data from both samples and consider both the benchmark and the alignment constraint.

The execution times of each of the method depend on several factors which are

difficult to control in a simulation study and vary between machines. Therefore results of this simulation should be treated as indicative only. The composite regression estimator and the calibration estimator require performing operations (adding, subtracting, multiplying) on large matrices. The speed of these operations is likely to depend on the available RAM and on the implementation of the matrix operations used in the software. The calculation of the point estimates has been implemented using the base R functions. This results in typically large operation times. The calculation of variance of the composite regression estimator was implemented using the Matrix package (Bates, 2018), which provides much more time efficient matrix storage and manipulation. This resulted in considerably shorter execution times.

The empirical likelihood and pseudoempirical likelihood methods require running a numerical optimisation in order to calculate the adjusted weights. Time necessary for the optimisation will very depending on the algorithm used, the starting point, the required precision, the particular sample data and the shape of the convex hull defined by the constraints. Therefore these times are likely to be much more varied than the calculation times of the regression-type estimators. Indeed, we can see that the coefficients of variation for these methods are large.

The execution times understandably increase with the growing size of the sample. We also notice that the average execution time of the calculation of empirical likelihood confidence intervals is much higher than the average execution time for the calculation of the empirical likelihood point estimate. This is understandable, because obtaining confidence intervals requires calculating the adjusted weights for various candidate values of the parameter of interest, as it was explained in earlier chapters. The execution times for the calculation of the variance of the composite regression estimator are lower than the times for calculation of the point estimates because of the differences in implementation. The calculation of the composite regression point estimates relies on the base R functions (as do the matrix operations in the calculation of the empirical likelihood and pseudoempirical likelihood point

estimates), which results in high execution times. It is interesting to see that in small samples, the execution times for the empirical likelihood methods are considerably higher than those for the composite regression method. However, for very large samples the opposite is true. We should note that a machine with only 4 GB of RAM was used to perform the simulations. Running this comparison on a machine with more available RAM would likely yield similar results, but the relationship between the execution times of the two estimators would change with a larger sample size. We should also notice that the execution times of the calculation of the empirical likelihood confidence intervals follow a highly skewed distribution, with a large difference between the mean and the median. This is caused by a small number of cases with particularly high execution times. All in all, we should note that while comparison of the these execution times is interesting, the execution times are not prohibitive and should not cause problems in practical applications.

Table 5.1: Distribution of the user execution times in seconds for calculation of point estimates and confidence intervals using various methods. Two samples with one alignment constraint and one benchmark constraint. 'AEL': proposed aligned empirical likelihood estimator. 'PEL': pseudoempirical likelihood approach (Wu, 2004*a*). 'Com.': composite regression estimator (Merkouris, 2004). 'Reg.': single sample calibration estimator (Deville and Särndal, 1992*a*). 'p.e.': point estimator. 'c.i.': confidence interval. 100 samples. Simulation setup as in chapter 2.12.3. 'C.V.': coefficient of variation.

$n_1 = n_2$	Estimation	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	C.V. (%)
1000	Reg. (p.e.)	0.00	0.01	0.02	0.01	0.02	0.03	54
	PEL (p.e.)	0.11	0.13	0.14	0.14	0.14	0.33	20
	Comp. (p.e.)	0.47	0.54	0.58	0.57	0.61	0.69	8
	Comp. (c.i.)	0.02	0.03	0.04	0.04	0.05	0.08	24
	AEL (p.e.)	0.00	0.01	0.02	0.01	0.02	0.03	62
	AEL (c.i.)	0.76	1.07	1.24	1.38	1.41	7.64	58
2000	Reg. (p.e.)	0.03	0.05	0.06	0.05	0.06	0.10	25
	PEL (p.e.)	0.23	0.25	0.26	0.28	0.28	0.98	29
	Comp. (p.e.)	2.47	2.62	2.69	2.69	2.75	3.07	4
	Comp. (c.i.)	0.11	0.14	0.14	0.15	0.16	0.20	12
	AEL (p.e.)	0.01	0.02	0.03	0.03	0.03	0.06	43
	AEL (c.i.)	1.35	1.93	2.30	3.07	2.91	38.77	130
3000	Reg. (p.e.)	0.07	0.11	0.12	0.12	0.13	0.19	17
	PEL (p.e.)	0.35	0.37	0.39	0.40	0.39	0.73	13
	Comp. (p.e.)	5.86	6.11	6.25	6.43	6.53	8.11	8
	Comp. (c.i.)	0.25	0.30	0.32	0.32	0.33	0.51	13
	AEL (p.e.)	0.01	0.03	0.03	0.04	0.05	0.16	47
	AEL (c.i.)	1.56	2.48	3.15	4.65	4.16	34.06	110
4000	Reg. (p.e.)	0.15	0.20	0.21	0.21	0.22	0.28	11
	PEL (p.e.)	0.41	0.50	0.50	0.54	0.52	1.62	30
	Comp. (p.e.)	10.36	10.84	10.94	11.02	11.08	13.47	4
	Comp. (c.i.)	0.48	0.53	0.54	0.55	0.56	0.67	6
	AEL (p.e.)	0.03	0.04	0.05	0.06	0.06	0.25	66
	AEL (c.i.)	2.01	3.05	4.10	5.87	6.43	29.53	87

5.5 Conclusions

While empirical likelihood is less computationally intensive then the methods that involve resampling, its computational aspects are not trivial. The computational difficulty is driven by the dimension of the parameter of interest, the parameter space and the number and type of additional constraints.

The estimation of the adjusted weights $\widehat{m}_i(\boldsymbol{\varphi}_U)$ is a relatively straightforward task of optimisation under constraints. If a solution with an acceptable precision cannot be found, this is an indication that the constraints might contradict each other, or that they are linearly dependent.

Point estimates can be obtained by solving an estimating equation. This method is preferred to direct minimisation of the log-likelihood ratio function, as it is much less computationally demanding.

Estimation of the confidence interval bounds is the most computationally expensive part of empirical likelihood estimation, as it involves evaluating the log-likelihood ratio function for multiple candidate values of the parameter of interest. The complexity of this task increases as the dimension of the parameter of interest increases. Therefore, obtaining joint confidence regions for parameters of a high dimension might be computationally challenging. However, a separate confidence interval for each scalar parameter of interest can be obtained as long as this parameter is uniquely defined by a single estimating equation. It is unlikely that joint confidence regions for parameters of a large dimension will be required in practice.

Finding confidence intervals is considerably simpler than a general problem of finding a contour of a function, as we know that the log-likelihood ratio is a convex function taking the minimum value at the point estimate. We propose an algorithm which utilises this property and avoids evaluating the log-likelihood

ratio function for extreme values of the parameter.

Various algorithms are available to find the Lagrange's multipliers. The quasi-Newton algorithm with a correction applied to ensure that the adjusted weights are positive, has been used by several authors. Alternatively, several other quasi-Newton methods are available and implemented in most statistical packages.

It is also worth noting that the field of mathematical optimisation is intensively developing and various complex methods and heuristics for optimisation and root finding have been presented. We did not find the need to resort to any of these far more complex methods as the relatively simple and easily interpretable quasi-Newton approach was sufficient. Evaluating the relative convergence speed and precision of the achieved solution across a range of algorithms could be an interesting direction of future research.

Discussion

In this piece of work we extended the design-based empirical likelihood methodology to accommodate alignment constraints and inference from multiple frame surveys. This included defining the empirical likelihood methodology, specifying the relevant constraints and estimating equations for the considered problems, as well as showing that in these circumstances the empirical likelihood ratio function is still pivotal and that the empirical likelihood point estimator is $\sqrt{n}$ design-consistent. We also discuss how the proposed empirical likelihood estimators relate to other estimators available for each of the studied problems. We consider these theoretical results to be the main contribution of the presented work.

Following that, we applied empirical likelihood to estimation of census coverage. This leveraged the fact that empirical likelihood confidence intervals are asymmetric and range-preserving. We conclude that empirical likelihood indeed correctly accounted for the unequal variability of data on both sides of the point estimate and produced confidence intervals within the desired range and with good coverage. However, we notice that further developments would be needed if empirical likelihood was to be used as an alternative for the current methodology. In particular, finding a way of incorporating the uncertainty around the Dual System estimator and the adjustments applied on top of it into the empirical likelihood framework would be desirable. We also notice that empirical likelihood, as many other design-based methods, requires large samples to obtain confidence intervals with good coverage, which would make it unsuitable for the last part of census coverage estimation which involves small area estimation methods. However, em-

pirical likelihood could be used to produce confidence intervals for census coverage at the Estimation Area level, as well as a source of comparative information in the quality assurance process. Empirical likelihood could also be applied to estimation of other ratios, e.g. domain proportions, or other range-restricted parameters.

In the simulation studies performed, we confirmed that empirical likelihood deals well with skewed data and found that in such circumstances the empirical likelihood point estimator might be more precise than the regression-based estimators. In relation to aligning estimates from samples of different sizes, we found that empirical likelihood, even without introducing any adjustment factors, performs relatively well. The precision of estimates obtained from the larger sample is only mildly deteriorated, while the precision of estimates obtained from the small sample is hugely increased. This is due to the implicit sample size adjustment imposed by the design constraints.

We notice that empirical likelihood confidence intervals tend to show some under-coverage in small samples. We also notice that if the coverage of empirical likelihood confidence intervals differs from the nominal value, this is much more often due to under-coverage than over-coverage.

The possibility to calculate confidence interval bounds without the intermediate step of variance estimation is likely to be of practical benefit. Empirical likelihood is also likely to be less computationally demanding than bootstrap-based approaches, as it does not require resampling. However, the numerical operations necessary to obtain empirical likelihood point estimates and confidence intervals might not be trivial. For some problems obtaining precise solutions might be computationally difficult. In practical applications, it is therefore important to make sure that the numerical error in any empirical likelihood estimation is negligible.

The research described in the previous chapters allows us to conclude that it is possible to extend empirical likelihood beyond the basic single sample setup. This encourages further developments and new applications. The main challenge in

extending empirical likelihood to accommodate alignment of estimates and multiple frame surveys was showing that the empirical log-likelihood ratio function is still pivotal and follows a χ^2 distribution. This is indeed the case, which was shown both analytically and in simulation studies. We notice, however, that the result is based on the assumption of independence between samples. Extending the proposed approach to the case of dependent samples would be a desirable direction of future research. Note that the empirical likelihood point estimator is asymptotically $\sqrt{n}$ design-consistent whether or not the samples are independent. One possible way of dealing with dependent samples would be using the empirical likelihood point estimator and constructing symmetric confidence intervals based on the estimated variance of the asymptotically equivalent GREG estimator. A similar approach was proposed by Wu (2004*a*) for pseudoempirical likelihood. However, this would only have practical merits in situations when empirical likelihood is likely to be more precise than the composite regression estimator and when these variance estimates are easy to obtain.

Since the first results on design-based empirical likelihood were published (Berger and De La Riva Torres, 2011), the method has been extended to handle nuisance parameters, non-response (Berger, 2017) and cluster sampling designs (Oguz-Alper and Berger, 2016). Design-based empirical likelihood inference has been applied to several sampling designs, such as the Hartley-Rao-Cochran design (Berger, 2016) and the adaptive cluster sampling (Salehi et al., 2010). It has also been used in research on the EU-SILC data (Berger and Torres, 2014). Possible directions of future research could include extending the method to handle cluster sampling beyond the ultimate cluster approach. The low coverage of empirical likelihood confidence intervals in small samples indicates that an empirical likelihood methodology for small domain estimation might not be achievable.

Another challenge in applying empirical likelihood is the lack of a closed form for variance of the empirical likelihood point estimator. This variance can be approximated by the variance of the asymptotically equivalent regression estimator.

However, the lack of a closed form for the variance makes it difficult to define an optimal way of combining estimates or samples. This is possible for regression based estimators and optimal composite regression estimator for alignment and optimal adjustment for the generalized multiplicity adjusted Horwitz-Thompson estimator were proposed. While it is possible to incorporate adjustment factors based on an efficiency calculation into the empirical likelihood constraints, as it has been shown in chapter 3, defining the adjustments that would minimise the variance of the resulting empirical likelihood point estimator is not straightforward. The adjustment factors would need to be based on the variance of the asymptotically equivalent regression estimator rather than the empirical likelihood estimator. Future research aiming to derive variance of the empirical likelihood estimator in various settings would be useful. However, we should note that while the sub-optimal regression estimators are available for both estimates alignment and multiple frame surveys, they rely on variance estimates and require selecting variables with respect to which optimality will be achieved.

Appendix A

Proofs of the results

Below proofs of the results presented in earlier chapters are given. Chapter 2 extends the results of Berger and De La Riva Torres (2016) to a two sample case with alignment constraints. The proofs, except the proof of theorem 2, are therefore an adaptation of the proofs presented in the original paper. Chapter 3 applies a similar reasoning to multiple frame surveys, yet the results of chapter 2 need some adjustments to account for the differences in the sampling design. These are summarised in the second part of the appendix. Chapter 4 is an application of the design based empirical likelihood of Berger and De La Riva Torres (2016) and all the relevant proofs can be found in the original paper.

A.1 Proofs of the results of Chapter 2

Below we show how the proofs of lemma (1), lemma (2) and theorem (16) presented by Berger and De La Riva Torres (2016) can be adapted to derive the GREG estimator asymptotically equivalent to the aligned empirical likelihood estimator .

Lemma 1 (Adaptation of lemma 1 in (Berger and De La Riva Torres, 2016))**.** *Let N be the population size. Let $n = n_1 + n_2$, where n_t is the size of sample $\boldsymbol{S}_t$,*

$t = 1, 2$. *Let* $\boldsymbol{c_i}$ *and* $\boldsymbol{C}$ *be defined by (2.48) and (2.49) respectively. Let* $\boldsymbol{\eta}$ *be the vector of Lagrange multipliers in (2.61). Let us assume that the regularity conditions (2.70–2.75) hold. Then,*

$$\frac{N}{n}||\boldsymbol{\eta}|| = O_{\mathcal{P}}(n^{-\frac{1}{2}})\cdot \tag{A.1}$$

Proof. Let $\widehat{m}_i(\boldsymbol{\varphi}_U)$ be defined by (2.61). Berger and De La Riva Torres (2016) show that for any $\boldsymbol{L}$ such that

$$||\boldsymbol{\eta}||\boldsymbol{L} = \boldsymbol{\eta}, \tag{A.2a}$$

$$||\boldsymbol{L}|| = O_{\mathcal{P}}(1), \tag{A.2b}$$

$$||\boldsymbol{L}^{-1}|| = O_{\mathcal{P}}(1), \tag{A.2c}$$

we have that

$$||\boldsymbol{\eta}||\{-nN^{-2}\boldsymbol{L}^\top\hat{\boldsymbol{\Sigma}}\boldsymbol{L} - nN^{-1}MN^{-1}|\boldsymbol{L}^\top(\hat{\boldsymbol{C}}_{\boldsymbol{\pi}} - \boldsymbol{C})|\} \leq nN^{-1}N^{-1}|\boldsymbol{L}^\top(\hat{\boldsymbol{C}}_{\boldsymbol{\pi}} - \boldsymbol{C})|, \tag{A.3}$$

where $M = max|(\pi_i^{-1}||\boldsymbol{c_i}||)|$ and

$$\hat{\boldsymbol{\Sigma}} = \sum_{i\in\boldsymbol{S}} \frac{\boldsymbol{c_i}\boldsymbol{c_i}^\top}{\pi_i^2}\cdot \tag{A.4}$$

The term $-nN^{-2}\boldsymbol{L}^\top\hat{\boldsymbol{\Sigma}}\boldsymbol{L}-$ is $O_{\mathcal{P}}(1)$ due to (2.73) and (A.2b). Furthermore, $nN^{-1}M = o_{\mathcal{P}}(n^{\frac{1}{2}})$ because of (A.2c) and $N^{-1}|\boldsymbol{L}^\top(\hat{\boldsymbol{C}}_{\boldsymbol{\pi}} - \boldsymbol{C})| = O_{\mathcal{P}}(n^{-\frac{1}{2}})$ due to (2.71) and (A.2b), giving

$$nN^{-1}MN^{-1}|\boldsymbol{L}^\top(\hat{\boldsymbol{C}}_{\boldsymbol{\pi}} - \boldsymbol{C})| = o_{\mathcal{P}}(n^{\frac{1}{2}})O_{\mathcal{P}}(n^{-\frac{1}{2}}) = o_{\mathcal{P}}(n^{\frac{1}{2}}n^{-\frac{1}{2}}) = o_{\mathcal{P}}(1)\cdot \tag{A.5}$$

Therefore, using (A.3), we have that

$$Nn^{-1}||\boldsymbol{\eta}||\{O_{\mathcal{P}}(1) - o_{\mathcal{P}}(1)\} \leq O_{\mathcal{P}}(n^{-\frac{1}{2}}), \tag{A.6}$$

which gives

$$Nn^{-1}||\boldsymbol{\eta}|| = O_{\mathcal{P}}(n^{-\frac{1}{2}}). \tag{A.7}$$

□

Lemma 2 (Adaptation of lemma 2 in (Berger and De La Riva Torres, 2016))**.** *Let $\boldsymbol{\eta}$ be the vector of Lagrange multipliers in (2.61). Let $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{S}}$ be defined by (A.4) and (2.76) respectively. Under the regularity conditions (2.70–2.75), we have that:*

$$\boldsymbol{\eta} = \frac{n}{N^2}\hat{\boldsymbol{S}}^{-1}(\boldsymbol{C} - \hat{\boldsymbol{C}}_{\boldsymbol{\pi}}) + \frac{n}{N}O_{\mathcal{P}}(n^{-1})\cdot \tag{A.8}$$

Proof. Berger and De La Riva Torres (2016) show that, based on (2.76), we have that

$$\boldsymbol{\eta} = \frac{n}{N^2}\hat{\boldsymbol{S}}^{-1}(\boldsymbol{C} - \hat{\boldsymbol{C}}_{\boldsymbol{\pi}}) + \frac{n}{N}\hat{\boldsymbol{e}}, \tag{A.9}$$

where

$$\hat{\boldsymbol{e}} = -\frac{1}{N}\hat{\boldsymbol{S}}^{-1}\sum_{i\in\boldsymbol{S}}\frac{\boldsymbol{c_i}\theta_i}{\pi_i} \tag{A.10}$$

and $\theta_i = v_i(1 - \widehat{m}_i(\boldsymbol{\varphi}_U)\pi_i)$ with $v_i = \pi_i^{-1}\boldsymbol{\eta}^\top\boldsymbol{c}_i$.

This gives (Berger and De La Riva Torres, 2016)

$$\begin{aligned}
\|\hat{\boldsymbol{e}}\| \quad &\leq \frac{1}{N}||\hat{\boldsymbol{S}}^{-1}||\sum_{i\in\boldsymbol{S}}\left\{\frac{||\boldsymbol{c_i}||}{\pi_i}\frac{1}{\pi_i^2}||\boldsymbol{\eta}^\top||^2||\boldsymbol{c_i}||^2\right\} + \sum_{i\in\boldsymbol{S}}\left\{\frac{||\boldsymbol{c_i}||}{\pi_i}|\gamma_i|\right\} \\
&\leq ||\hat{\boldsymbol{S}}^{-1}||\,(\frac{N}{n}||\boldsymbol{\eta}^\top||)^2 n^3\frac{1}{nN^3}\sum_{i\in\boldsymbol{S}}\left\{\frac{||\boldsymbol{c_i}||^3}{\pi_i^3}\right\} + \frac{1}{N}||\hat{\boldsymbol{S}}^{-1}||\sum_{i\in\boldsymbol{S}}\left\{\frac{||\boldsymbol{c_i}||}{\pi_i}|\gamma_i|\right\}
\end{aligned} \tag{A.11}$$

where $|\gamma_i|$ is such that $Pr\left\{|\gamma_i| \leq k|v_i|^3, i \in \boldsymbol{S}\right\} \to 1$ with $k > 0$. According to (2.74), $||\hat{\boldsymbol{S}}^{-1}||$ is $O_{\mathcal{P}}(1)$. Using lemma (1), we have that

$$(\frac{N}{n}||\boldsymbol{\eta}^\top||)^2 = O_{\mathcal{P}}(n^{-\frac{1}{2}})^2 = O_{\mathcal{P}}(n^{-1})\cdot \tag{A.12}$$

Based on (2.75),

$$\frac{1}{nN^3}\sum_{i\in\boldsymbol{S}}\left\{\frac{||\boldsymbol{c_i}||^3}{\pi_i^3}\right\} = O_{\mathcal{P}}(n^{-3})\cdot \tag{A.13}$$

Therefore the first term of the right hand side of (A.11) is $n^3 O_{\mathcal{P}}(1)O_{\mathcal{P}}(n^{-3})O_{\mathcal{P}}(n^{-1}) = n^3 O_{\mathcal{P}}(n^{-4}) = O_{\mathcal{P}}(n^{-1})$. Omitting $||\hat{\boldsymbol{S}}^{-1}||$, which is $O_{\mathcal{P}}(1)$, we can write the second term of (A.11) as:

$$\begin{aligned}\frac{1}{N}\sum_{i\in\boldsymbol{S}}\left\{\frac{||\boldsymbol{c_i}||}{\pi_i}|\gamma_i|\right\} &\leq \frac{1}{N}\sum_{i\in\boldsymbol{S}}\left\{\frac{||\boldsymbol{c_i}||}{\pi_i}k|v_i|^3\right\} \\ &\leq k\frac{N^3}{n^3}||\boldsymbol{\eta}^\top||^3\frac{1}{nN^4}\sum_{i\in\boldsymbol{S}}\left\{\frac{||\boldsymbol{c_i}||^4}{\pi_i^4}\right\}n^4. \end{aligned}\tag{A.14}$$

Using Lemma (1), $\frac{N^3}{n^3}||\boldsymbol{\eta}^\top||^3$ is $O_{\mathcal{P}}(n^{-\frac{3}{2}})$. The term $\frac{1}{nN^4}\sum_{i\in\boldsymbol{S}}\left\{\frac{||\boldsymbol{c_i}||^4}{\pi_i^4}\right\}$ is $O_{\mathcal{P}}(n^{-4})$ given (2.75). Therefore the second term of (A.11) is $O_{\mathcal{P}}(n^{-\frac{3}{2}})$. This makes $||\hat{\boldsymbol{e}}|| = O_{\mathcal{P}}(n^{-1})$. The lemma follows.

□

Proof of Theorem 1

(Adaptation of proof of equation (16), page 2 of supplementary materials in (Berger and De La Riva Torres, 2016))

Let us define:

$$\hat{\boldsymbol{e}}_1 = \frac{n}{N}\left(\sum_{i\in\boldsymbol{S}}\frac{\boldsymbol{c}_i\boldsymbol{g}_i(\boldsymbol{\theta})^\top}{(1+v_i)\pi_i^2}\hat{\boldsymbol{e}}\right), \tag{A.15}$$

where $\hat{\boldsymbol{e}}$ is defined by (A.10),

$$\hat{\boldsymbol{e}}_2 = \frac{n}{N^2}\left(\sum_{i\in\boldsymbol{S}}\frac{\boldsymbol{c}_i\boldsymbol{g}_i(\boldsymbol{\theta})^\top}{\pi_i^2}\right)\hat{\boldsymbol{S}}^{-1}\tilde{\boldsymbol{C}}_{\boldsymbol{\pi}}\frac{v_i}{1+v_i}, \tag{A.16}$$

$$\widehat{\boldsymbol{G}}(\boldsymbol{\theta})_\pi = \sum_{i\in\boldsymbol{S}}\frac{\boldsymbol{g}_i(\boldsymbol{\theta})}{\pi_i} \tag{A.17}$$

and $v_i = \pi_i^{-1}\boldsymbol{\eta}^\top \boldsymbol{c}_i$. Considering Lemmas (1) and (2) and using analogous reasoning to that presented by Berger and De La Riva Torres (2016), we can express (2.58) as:

$$\widehat{\boldsymbol{G}}(\boldsymbol{\theta}) = \widehat{\boldsymbol{G}}(\boldsymbol{\theta})_\pi + \widehat{\boldsymbol{B}}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)^\top (\boldsymbol{C} - \hat{\boldsymbol{C}}_{\boldsymbol{\pi}}) - \hat{\boldsymbol{e}}_2 - \hat{\boldsymbol{e}}_1, \tag{A.18}$$

where

$$\widehat{\boldsymbol{B}}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U) = \left(\sum_{i \in \boldsymbol{S}_1, \boldsymbol{S}_2} \frac{\boldsymbol{c}_i \boldsymbol{c}_i^\top}{\pi_i^2} \right)^{-1} \left(\sum_{i \in \boldsymbol{S}} \frac{\boldsymbol{c}_i \boldsymbol{g}_i(\boldsymbol{\theta})^\top}{\pi_i^2} \right) = \hat{\boldsymbol{\Sigma}}^{-1} \left(\sum_{i \in \boldsymbol{S}} \frac{\boldsymbol{c}_i \boldsymbol{g}_i(\boldsymbol{\theta})^\top}{\pi_i^2} \right) \tag{A.19}$$

and $\hat{\boldsymbol{e}}_1$, $\hat{\boldsymbol{e}}_2$, $\widehat{\boldsymbol{G}}(\boldsymbol{\theta})_\pi$ and $\hat{\boldsymbol{\Sigma}}$ are defined as in (A.15), (A.16), (A.17) and (A.4) respectively.

Using Cauchy-Schwartz inequality as proposed by Berger and De La Riva Torres (2016), (2.75), (2.87) and the result $||\hat{\boldsymbol{e}}|| = O_{\mathcal{P}}(n^{-1})$ from the proof of Lemma (2), we can derive the asymptotic properties of $\hat{\boldsymbol{e}}_1$ as follows:

$$\begin{aligned} \|\hat{\boldsymbol{e}}_1\| &\leq \frac{n}{N} \|\hat{\boldsymbol{e}}\| \zeta \left(\sum_{i \in \boldsymbol{S}} \frac{||\boldsymbol{c}_i||\, |||\boldsymbol{g}_i(\boldsymbol{\theta})^\top||}{\pi_i^2} \right) \\ &\leq N n^2 \|\hat{\boldsymbol{e}}\| \zeta \left(\frac{1}{nN^2} \sum_{i \in \boldsymbol{S}} \frac{||\boldsymbol{c}_i^\top||^2}{\pi_i^2} \right)^{\frac{1}{2}} \left(\frac{1}{nN^2} \sum_{i \in \boldsymbol{S}} \frac{||\boldsymbol{g}_i(\boldsymbol{\theta})||^2}{\pi_i^2} \right)^{\frac{1}{2}}, \end{aligned} \tag{A.20}$$

where $\zeta = (min|1 + v_i| : i \in \boldsymbol{S})^{-1} = O_{\mathcal{P}}(1)$ because $|v_i| = o_{\mathcal{P}}(1)$. Therefore,

$$\|\hat{\boldsymbol{e}}_1\| \leq N n^2 O_{\mathcal{P}}(n^{-1}) O_{\mathcal{P}}(1) O_{\mathcal{P}}(n^{-2}) = O_{\mathcal{P}}(N n^{-1}) \cdot \tag{A.21}$$

This gives $\|\hat{\boldsymbol{e}}_1\| = o_{\mathcal{P}}(N n^{-1})$. Furthermore,

$$\|\hat{\boldsymbol{e}}_2\| \leq \frac{n}{N^2} \zeta \tau ||\hat{\boldsymbol{S}}^{-1}|| ||\tilde{\boldsymbol{C}}_{\boldsymbol{\pi}}|| \left(\sum_{i \in \boldsymbol{S}} \frac{||\boldsymbol{c}_i||\, ||\boldsymbol{g}_i(\boldsymbol{\theta})^\top||}{\pi_i^2} \right), \tag{A.22}$$

where $\tau = (max|v_i| : i \in \boldsymbol{S}) = o_{\mathcal{P}}(1)$. Following the same argument as in (A.20),

we have that

$$\left(\sum_{i\in\boldsymbol{S}}\frac{||\boldsymbol{c}_i||\,||\boldsymbol{g}_i(\boldsymbol{\theta})^\top||}{\pi_i^2}\right)=nN^2O_{\mathcal{P}}(n^{-2})=O_{\mathcal{P}}(N^2n^{-1})\cdot \tag{A.23}$$

According to (2.74), $||\hat{\boldsymbol{S}}^{-1}||=O_{\mathcal{P}}(1)$. Furthermore, $N^{-1}||\tilde{\boldsymbol{C}}_{\boldsymbol{\pi}}||=O_{\mathcal{P}}(n^{-\frac{1}{2}})$ because of (2.71). This gives

$$\|\hat{\boldsymbol{e}}_2\|\leq nN^{-1}O_{\mathcal{P}}(1)O_{\mathcal{P}}(n^{-\frac{1}{2}})O_{\mathcal{P}}(N^2n^{-1})=O_{\mathcal{P}}(Nn^{-\frac{1}{2}})\cdot \tag{A.24}$$

and hence $\|\hat{\boldsymbol{e}}_2\|=o_{\mathcal{P}}(Nn^{-\frac{1}{2}})$. This gives

$$\begin{aligned}\widehat{\boldsymbol{G}}(\boldsymbol{\theta}) &= \widehat{\boldsymbol{G}}(\boldsymbol{\theta})_\pi+\widehat{\boldsymbol{B}}(\boldsymbol{\theta},\boldsymbol{\varphi}_{U})^\top(\boldsymbol{C}-\hat{\boldsymbol{C}}_{\boldsymbol{\pi}})+O_{\mathcal{P}}(Nn^{-\frac{1}{2}})+O_{\mathcal{P}}(Nn^{-1}) &\text{(A.25)}\\ &= \widehat{\boldsymbol{G}}(\boldsymbol{\theta})_\pi+\widehat{\boldsymbol{B}}(\boldsymbol{\theta},\boldsymbol{\varphi}_{U})^\top(\boldsymbol{C}-\hat{\boldsymbol{C}}_{\boldsymbol{\pi}})+o_{\mathcal{P}}(Nn^{-\frac{1}{2}}) &\text{(A.26)}\end{aligned}$$

The theorem follows.

Proof of theorem 2

Let

$$\widehat{\boldsymbol{B}}(\boldsymbol{\theta}_{U},\boldsymbol{\varphi}_{U})^\top:=\left(\sum_{i\in\boldsymbol{S}}\frac{1}{\pi_i^2}\boldsymbol{c}_i\boldsymbol{c}_i^\top\right)^{-1}\sum_{i\in\boldsymbol{S}}\frac{1}{\pi_i^2}\boldsymbol{c}_i\boldsymbol{g}_i(\boldsymbol{\theta}_{U})^\top\cdot \tag{A.27}$$

Let us assume that the constraints (2.70)-(2.75), (2.79), (2.80) (with $\tau=2$), (2.81)-(2.84) hold. Following theorem 1, we notice that:

$$N^{-1}||\widehat{\boldsymbol{G}}(\boldsymbol{\theta}_{U})||\leq N^{-1}\left(||\widehat{\boldsymbol{G}}(\boldsymbol{\theta}_{U})_\pi||+||\widehat{\boldsymbol{B}}(\boldsymbol{\theta}_{U},\boldsymbol{\varphi}_{U})^\top||\,||(\boldsymbol{C}-\hat{\boldsymbol{C}}_{\boldsymbol{\pi}})||+o_{\mathcal{P}}(Nn^{-\frac{1}{2}})\right). \tag{A.28}$$

According to (2.79) $N^{-1}||\widehat{\boldsymbol{G}}(\boldsymbol{\theta}_{U})_\pi||=O_{\mathcal{P}}(n^{-\frac{1}{2}})$. In (A.20) it has been shown that

$$\left(\sum_{i\in\boldsymbol{S}}\frac{||\boldsymbol{g}_i(\boldsymbol{\theta}_{U})||||\boldsymbol{c}_i{}^\top||}{\pi_i^2}\right)=nN^2O_{\mathcal{P}}(n^{-2})=O_{\mathcal{P}}(N^2n^{-1})\cdot \tag{A.29}$$

Using this property and (2.70), we conclude that

$$||\widehat{\boldsymbol{B}}(\boldsymbol{\theta}_U, \boldsymbol{\varphi}_U)|| = O_{\mathcal{P}}(1)\cdot \tag{A.30}$$

Therefore, using (2.72),

$$N^{-1}||\widehat{\boldsymbol{G}}(\boldsymbol{\theta}_U)|| = O_{\mathcal{P}}(n^{-\frac{1}{2}}). \tag{A.31}$$

Using Taylor expansion, we can write:

$$N^{-1}\widehat{\boldsymbol{G}}(\widehat{\boldsymbol{\theta}}) = N^{-1}\widehat{\boldsymbol{G}}(\boldsymbol{\theta}_U) + \frac{\partial \widehat{\boldsymbol{G}}(\boldsymbol{\theta}_U)}{\partial \boldsymbol{\theta}_U}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_U) + O(||\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_U||^2)\cdot \tag{A.32}$$

Suppose that $\widehat{\boldsymbol{\theta}}$ satisfies $N^{-1}\widehat{\boldsymbol{G}}(\widehat{\boldsymbol{\theta}}) = \mathbf{0}$. Then, using (A.31), (2.83) and (2.84), we have that

$$O_{\mathcal{P}}(||\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_U||) = O_{\mathcal{P}}(n^{-\frac{1}{2}}). \tag{A.33}$$

Proof of Theorem 3

(Adaptation of proof of *Theorem 1*, page 4 of supplementary materials in (Berger and De La Riva Torres, 2016))

Let $\boldsymbol{c}_i^\star(\boldsymbol{\theta})$, $\boldsymbol{C}^\star$, $\boldsymbol{c}_i$ and $\boldsymbol{C}$ be defined by (2.46), (2.47), (2.48) and (2.49) respectively. Consider

$$\sum_{i\in\boldsymbol{S}} \frac{1}{\pi_i^2}\boldsymbol{c}_i^\star(\boldsymbol{\theta})\boldsymbol{c}_i^\star(\boldsymbol{\theta})^T = \begin{bmatrix} \hat{\boldsymbol{\Sigma}} & \hat{\boldsymbol{\Sigma}}_{\boldsymbol{cg}}(\boldsymbol{\theta}) \\ \hat{\boldsymbol{\Sigma}}_{\boldsymbol{cg}}(\boldsymbol{\theta})^\top & \hat{\boldsymbol{\sigma}}_{\boldsymbol{gg}}(\boldsymbol{\theta}) \end{bmatrix}, \tag{A.34}$$

where $\hat{\boldsymbol{\Sigma}}$ is defined by (A.4), i.e., $\hat{\boldsymbol{\Sigma}} = \sum_{i\in\boldsymbol{S}} \frac{\boldsymbol{c}_i\boldsymbol{c}_i^\top}{\pi_i^2}$,

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{cg}}(\boldsymbol{\theta}) = \sum_{i\in\boldsymbol{S}} \frac{1}{\pi_i^2}\boldsymbol{c}_i\boldsymbol{g}_i(\boldsymbol{\theta})^\top \tag{A.35}$$

and

$$\hat{\boldsymbol{\sigma}}_{\boldsymbol{gg}}(\boldsymbol{\theta}) = \sum_{i \in \boldsymbol{S}} \frac{1}{\pi_i^2} \boldsymbol{g}_i(\boldsymbol{\theta}) \boldsymbol{g}_i(\boldsymbol{\theta})^\top. \tag{A.36}$$

Let

$$\widetilde{\boldsymbol{g}}_i(\boldsymbol{\theta}) = \boldsymbol{g}_i(\boldsymbol{\theta}) - \widehat{\boldsymbol{B}}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{U})^\top \boldsymbol{c}_i, \tag{A.37}$$

with $\widehat{\boldsymbol{B}}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{U})$ defined by (2.90).

Consider

$$\tilde{\boldsymbol{c}}_{\boldsymbol{i}}^{*}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{c}_i \\ \widetilde{\boldsymbol{g}}_i(\boldsymbol{\theta}) \end{pmatrix} \tag{A.38}$$

and

$$\tilde{\boldsymbol{C}}^{*} = \begin{pmatrix} \boldsymbol{C} \\ -\widehat{\boldsymbol{B}}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{U})^\top \boldsymbol{C} \end{pmatrix}. \tag{A.39}$$

We have that

$$\sum_{i \in \boldsymbol{S}} \frac{1}{\pi_i^2} \tilde{\boldsymbol{c}}_{\boldsymbol{i}}^{*}(\boldsymbol{\theta}) \tilde{\boldsymbol{c}}_{\boldsymbol{i}}^{*\boldsymbol{T}}(\boldsymbol{\theta}) = \begin{bmatrix} \hat{\boldsymbol{\Sigma}} & \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{cg}}(\boldsymbol{\theta}) \\ \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{cg}}(\boldsymbol{\theta})^\top & \widehat{\boldsymbol{V}}_{\mathcal{P}}\big\{\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{U})\big\} \end{bmatrix}, \tag{A.40}$$

where

$$\widehat{\boldsymbol{V}}_{\mathcal{P}}\big\{\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{U})\big\} = \sum_{i \in \boldsymbol{S}} \frac{1}{\pi_i^2} \widetilde{\boldsymbol{g}}_i(\boldsymbol{\theta}) \widetilde{\boldsymbol{g}}_i(\boldsymbol{\theta})^\top \tag{A.41}$$

and

$$\begin{aligned} \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{cg}}(\boldsymbol{\theta}) &= \sum_{i \in \boldsymbol{S}} \frac{\boldsymbol{c}_i \widetilde{\boldsymbol{g}}_i(\boldsymbol{\theta})^\top}{\pi_i^2} = \sum_{i \in \boldsymbol{S}} \frac{\boldsymbol{c}_i (\boldsymbol{g}_i(\boldsymbol{\theta}) - \widehat{\boldsymbol{B}}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{U})^\top \boldsymbol{c}_i)^\top}{\pi_i^2} \\ &= \sum_{i \in \boldsymbol{S}} \frac{\boldsymbol{c}_i \boldsymbol{g}_i(\boldsymbol{\theta})^\top}{\pi_i^2} - \sum_{i \in \boldsymbol{S}} \frac{\boldsymbol{c}_i \boldsymbol{c}_i^\top}{\pi_i^2} \widehat{\boldsymbol{B}}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{U}) \\ &= \sum_{i \in \boldsymbol{S}} \frac{\boldsymbol{c}_i \boldsymbol{g}_i(\boldsymbol{\theta})^\top}{\pi_i^2} - \left(\sum_{i \in \boldsymbol{S}} \frac{\boldsymbol{c}_i \boldsymbol{c}_i^\top}{\pi_i^2} \right) \left(\sum_{i \in \boldsymbol{S}} \frac{\boldsymbol{c}_i \boldsymbol{c}_i^\top}{\pi_i^2} \right)^{-1} \sum_{i \in \boldsymbol{S}} \frac{\boldsymbol{c}_i \boldsymbol{g}_i(\boldsymbol{\theta})^\top}{\pi_i^2} = \boldsymbol{0}. \end{aligned} \tag{A.42}$$

Let us define

$$\tilde{\boldsymbol{C}}_{\boldsymbol{\pi}}^{*} = \sum_{i \in \boldsymbol{S}} \frac{1}{\pi_i} \tilde{\boldsymbol{c}}_{\boldsymbol{i}}^{*} = \begin{bmatrix} \hat{\boldsymbol{C}}_{\boldsymbol{\pi}} \\ \hat{\boldsymbol{G}}_{\boldsymbol{\pi}}(\boldsymbol{\theta}) - \widehat{\boldsymbol{B}}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{U})^\top \hat{\boldsymbol{C}}_{\boldsymbol{\pi}} \end{bmatrix}. \tag{A.43}$$

We have that

$$\tilde{\boldsymbol{C}}^*_{\boldsymbol{\pi}} - \tilde{\boldsymbol{C}}^* = \begin{bmatrix} \hat{\boldsymbol{C}}_{\boldsymbol{\pi}} - \boldsymbol{C} \\ \widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}) \end{bmatrix} \tag{A.44}$$

where

$$\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}) = \widehat{\boldsymbol{G}}_\pi(\boldsymbol{\theta}) + \widehat{\boldsymbol{B}}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)^\top (\boldsymbol{C} - \hat{\boldsymbol{C}}_{\boldsymbol{\pi}}). \tag{A.45}$$

Finally,

$$\begin{aligned} &(\tilde{\boldsymbol{C}}^*_{\boldsymbol{\pi}} - \tilde{\boldsymbol{C}}^*)^\top \hat{\boldsymbol{\Sigma}}^{-1} (\tilde{\boldsymbol{C}}^*_{\boldsymbol{\pi}} - \tilde{\boldsymbol{C}}^*) \qquad (A.46) \\ &\quad = \left[(\hat{\boldsymbol{C}}_{\boldsymbol{\pi}} - \boldsymbol{C})^\top, \widehat{\boldsymbol{G}}_r(\boldsymbol{\theta})^\top \right] \begin{bmatrix} \hat{\boldsymbol{\Sigma}}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \widehat{\boldsymbol{V}}_{\mathcal{P}}\{\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_U)\}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{C}}_{\boldsymbol{\pi}} - \boldsymbol{C} \\ \widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}) \end{bmatrix} \\ &\quad = (\hat{\boldsymbol{C}}_{\boldsymbol{\pi}} - \boldsymbol{C})^\top \hat{\boldsymbol{\Sigma}}^{-1} (\hat{\boldsymbol{C}}_{\boldsymbol{\pi}} - \boldsymbol{C}) + \widehat{\boldsymbol{G}}_r(\boldsymbol{\theta})^\top \widehat{\boldsymbol{V}}_{\mathcal{P}}\{\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_U)\}^{-1} \widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}) \end{aligned}$$

Let $\ell(\widehat{\boldsymbol{m}}) = \sum_{i \in \boldsymbol{S}} \log(\widehat{m}_i(\boldsymbol{\varphi}_U))$, $\ell(\pi) = \sum_{i \in \boldsymbol{S}} \log(\pi_i)$ and $\ell(\boldsymbol{\theta}_U | \boldsymbol{\varphi}_U) = \sum_{i \in \boldsymbol{S}} \log(\widehat{m}^\star_i(\boldsymbol{\theta}_U, \boldsymbol{\varphi}_U))$. Berger and De La Riva Torres (2016) showed that

$$-2\left\{\ell(\widehat{\boldsymbol{m}}) + \ell(\pi)\right\} = (\hat{\boldsymbol{C}}_{\boldsymbol{\pi}} - \boldsymbol{C})^\top \hat{\boldsymbol{\Sigma}}^{-1} (\hat{\boldsymbol{C}}_{\boldsymbol{\pi}} - \boldsymbol{C}) + O_{\mathcal{P}}(n^{-1/2}) \tag{A.47}$$

and

$$-2\left\{\ell(\boldsymbol{\theta}_U | \boldsymbol{\varphi}_U) + \ell(\pi)\right\} = (\tilde{\boldsymbol{C}}^*_{\boldsymbol{\pi}} - \tilde{\boldsymbol{C}}^*)^\top \hat{\boldsymbol{\Sigma}}^{-1} (\tilde{\boldsymbol{C}}^*_{\boldsymbol{\pi}} - \tilde{\boldsymbol{C}}^*) + O_{\mathcal{P}}(n^{-1/2}) \cdot \tag{A.48}$$

Therefore,

$$\begin{aligned} \widehat{r}(\boldsymbol{\theta}_U | \boldsymbol{\varphi}_U) &= 2\left\{\ell(\boldsymbol{m}) - \ell(\boldsymbol{\theta}_U | \boldsymbol{\varphi}_U)\right\} \\ &= 2\ell(\boldsymbol{m}) + 2\ell(\pi) - 2\ell(\boldsymbol{\theta}_U | \boldsymbol{\varphi}_U) - 2\ell(\pi) \\ &= (\hat{\boldsymbol{C}}_{\boldsymbol{\pi}} - \boldsymbol{C})^\top \hat{\boldsymbol{\Sigma}}^{-1} (\hat{\boldsymbol{C}}_{\boldsymbol{\pi}} - \boldsymbol{C}) + \widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_U)^\top \widehat{\boldsymbol{V}}_{\mathcal{P}}\{\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_U)\}^{-1} \widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_U) - \\ &\quad - (\hat{\boldsymbol{C}}_{\boldsymbol{\pi}} - \boldsymbol{C})^\top \hat{\boldsymbol{\Sigma}}^{-1} (\hat{\boldsymbol{C}}_{\boldsymbol{\pi}} - \boldsymbol{C}) + O_{\mathcal{P}}(n^{-1/2}) \\ &= \widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_U)^\top \widehat{\boldsymbol{V}}_{\mathcal{P}}\{\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_U)\}^{-1} \widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_U) + O_{\mathcal{P}}(n^{-1/2}) \cdot \end{aligned} \tag{A.49}$$

The theorem follows.

A.2 Proofs of the results of Chapter 3

Proof of theorem 4

Let N be the population size and T be the number of sampling frames used. Let $n = \sum_{t=1}^{T} n_t$, where n_t is the size of sample $\boldsymbol{S}_t$ selected from frame Q_t. Let $\boldsymbol{c_i}$ and $\boldsymbol{C}$ be defined by (3.37) and (3.38) respectively.

Let $\boldsymbol{\eta}$ be the vector of Lagrange's multipliers in (3.43). Let ρ_i be the multiplicity-adjusted selection probability for unit $i \in \boldsymbol{S}$ defined by (3.18).

It can be show that with π_i substituted by ρ_i, $\widehat{\boldsymbol{C}}_\pi$ defined by (3.57), $\widehat{m}_i(\boldsymbol{\varphi}_U)$ defined by (3.43),

$$\hat{\boldsymbol{\Sigma}} := \sum_{i \in \boldsymbol{S}} \frac{\boldsymbol{c_i}\boldsymbol{c_i}^\top}{\rho_i^2}, \tag{A.50}$$

$\hat{\boldsymbol{S}}$ defined by (3.50) and assuming conditions (3.44–3.49), lemma (1) and lemma (2) still hold.

Using lemma (1) and lemma (2) with the adjustments listed above, it can be shown that

$$\widehat{\boldsymbol{G}}(\boldsymbol{\theta}) = \widehat{\boldsymbol{G}}(\boldsymbol{\theta})_\pi + \widehat{\boldsymbol{B}}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)^\top (\boldsymbol{C} - \hat{\boldsymbol{C}}_{\boldsymbol{\pi}}) - \hat{\boldsymbol{e}}_2 - \hat{\boldsymbol{e}}_1, \tag{A.51}$$

where

$$\widehat{\boldsymbol{B}}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U) = \left(\sum_{i \in \boldsymbol{S}_1, \boldsymbol{S}_2} \frac{\boldsymbol{c}_i \boldsymbol{c}_i^\top}{\rho_i^2} \right)^{-1} \left(\sum_{i \in \boldsymbol{S}} \frac{\boldsymbol{c}_i \boldsymbol{g}_i(\boldsymbol{\theta})^\top}{\rho_i^2} \right) = \hat{\boldsymbol{\Sigma}}^{-1} \left(\sum_{i \in \boldsymbol{S}} \frac{\boldsymbol{c}_i \boldsymbol{g}_i(\boldsymbol{\theta})^\top}{\rho_i^2} \right) \tag{A.52}$$

and $\hat{\boldsymbol{e}}_1$, $\hat{\boldsymbol{e}}_2$, $\widehat{\boldsymbol{G}}(\boldsymbol{\theta})_\pi$ are defined as in (A.15), (A.16) and (A.17) with π_i substituted

by ρ_i, respectively.

Using analogous reasoning to Appendix A.1, conditions (3.49), (3.52), (3.45) and (3.48), it can be shown that

$$\|\hat{\boldsymbol{e}}_1\| = o_{\mathcal{P}}(Nn^{-1})\cdot \tag{A.53}$$

and

$$\|\hat{\boldsymbol{e}}_2\| = o_{\mathcal{P}}(Nn^{-\frac{1}{2}})\cdot \tag{A.54}$$

The theorem follows.

Proof of theorem 5

Let $\widehat{\boldsymbol{G}}(\boldsymbol{\theta}_U)$ be defined by (3.41) with $\boldsymbol{\theta} = \boldsymbol{\theta}_U$. Assuming conditions (3.44)-(3.49), (3.58), (3.59) (with $\tau = 2$), (3.60)-(3.63), (3.58) and theorem 4, using analogous argument as in Appendix A.1, it can be shown that

$$N^{-1}||\widehat{\boldsymbol{G}}(\boldsymbol{\theta}_U)|| = O_{\mathcal{P}}(n^{-\frac{1}{2}}). \tag{A.55}$$

Hence, assuming (3.62) and (3.63) and using Taylor expansion and the reasoning presented in Appendix A.1, we have that

$$O_{\mathcal{P}}(||\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_U||) = O_{\mathcal{P}}(n^{-\frac{1}{2}}). \tag{A.56}$$

The theorem follows.

Proof of Theorem 6

Let $\boldsymbol{c}_i^{\star}(\boldsymbol{\theta})$, $\boldsymbol{C}^{\star}$, $\boldsymbol{c}_i$ and $\boldsymbol{C}$ be defined by (3.35), (3.36), (3.37) and (3.38) respectively. Let

$$\widetilde{\boldsymbol{g}}_i(\boldsymbol{\theta}) = \boldsymbol{g}_i(\boldsymbol{\theta}) - \widehat{\boldsymbol{B}}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)^{\top} \boldsymbol{c}_i, \tag{A.57}$$

with $\widehat{\boldsymbol{B}}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)^{\top}$ defined by (3.56),

$$\tilde{\boldsymbol{c}}_{\boldsymbol{i}}^{*}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{c}_i \\ \widetilde{\boldsymbol{g}}_i(\boldsymbol{\theta}) \end{pmatrix} \tag{A.58}$$

and

$$\tilde{\boldsymbol{C}}^{*} = \begin{pmatrix} \boldsymbol{C} \\ -\widehat{\boldsymbol{B}}(\boldsymbol{\theta}, \boldsymbol{\varphi}_U)^{\top} \boldsymbol{C} \end{pmatrix}. \tag{A.59}$$

Following an argument presented in Appendix A.1,it can be shown that

$$\sum_{i \in \boldsymbol{S}} \frac{1}{\rho_i^2} \tilde{\boldsymbol{c}}_{\boldsymbol{i}}^{*}(\boldsymbol{\theta}) \tilde{\boldsymbol{c}}_{\boldsymbol{i}}^{*\boldsymbol{T}}(\boldsymbol{\theta}) = \begin{bmatrix} \hat{\boldsymbol{\Sigma}} & \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{cg}}(\boldsymbol{\theta}) \\ \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{cg}}(\boldsymbol{\theta})^{\top} & \widehat{\boldsymbol{V}}_{\mathcal{P}}\{\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_U)\} \end{bmatrix}, \tag{A.60}$$

where $\hat{\boldsymbol{\Sigma}}$ is defined by (A.50),

$$\widehat{\boldsymbol{V}}_{\mathcal{P}}\{\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_U)\} = \sum_{i \in \boldsymbol{S}} \frac{1}{\rho_i^2} \widetilde{\boldsymbol{g}}_i(\boldsymbol{\theta}) \widetilde{\boldsymbol{g}}_i(\boldsymbol{\theta})^{\top} \tag{A.61}$$

and

$$\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{cg}}(\boldsymbol{\theta}) = \mathbf{0}\cdot \tag{A.62}$$

This leads to

$$\begin{aligned} (\tilde{\boldsymbol{C}}_{\boldsymbol{\pi}}^{*} - \tilde{\boldsymbol{C}}^{*})^{\top} \quad & \hat{\boldsymbol{\Sigma}}^{-1} (\tilde{\boldsymbol{C}}_{\boldsymbol{\pi}}^{*} - \tilde{\boldsymbol{C}}^{*}) = \\ & = (\hat{\boldsymbol{C}}_{\boldsymbol{\pi}} - \boldsymbol{C})^{\top} \hat{\boldsymbol{\Sigma}}^{-1} (\hat{\boldsymbol{C}}_{\boldsymbol{\pi}} - \boldsymbol{C}) + \widehat{\boldsymbol{G}}_r(\boldsymbol{\theta})^{\top} \widehat{\boldsymbol{V}}_{\mathcal{P}}\{\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_U)\}^{-1} \widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}) \end{aligned} \tag{A.63}$$

where

$$\tilde{\boldsymbol{C}}^*_{\boldsymbol{\pi}} = \sum_{i \in \boldsymbol{S}} \frac{1}{\rho_i} \tilde{\boldsymbol{c}}^*_{\boldsymbol{i}}(\boldsymbol{\theta}) \cdot \tag{A.64}$$

Using (A.63) and analogous arguments as presented in Appendix A.1, with $\ell(\widehat{\boldsymbol{m}}) = \sum_{i \in \boldsymbol{S}} \log(\widehat{m}_i(\boldsymbol{\varphi}_U))$, $\ell(\pi) = \sum_{i \in \boldsymbol{S}} \log(\rho_i)$ and $\ell(\boldsymbol{\theta}_U | \boldsymbol{\varphi}_U) = \sum_{i \in \boldsymbol{S}} \log(\widehat{m}^\star_i(\boldsymbol{\theta}_U, \boldsymbol{\varphi}_U))$, it can be shown that

$$\widehat{r}(\boldsymbol{\theta}_U | \boldsymbol{\varphi}_U) = \widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_U)^\top \widehat{\boldsymbol{V}}_{\mathcal{P}} \big\{ \widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_U) \big\}^{-1} \widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_U) + O_{\mathcal{P}}(n^{-1/2}) \cdot \tag{A.65}$$

The theorem follows.

Bibliography

Abbott, O. (2009), "2011 UK Census Coverage Assessment and Adjustment Methodology," *Population trends*, (137), 25.

Abbott, O. (2011), "Advisory Group paper AG (08) 05-2011 UK Census Coverage Assessment and Adjustment Methodology,".
URL: *https://www.ons.gov.uk/ons/guide-method/census/2011/the-2011-census/census-consultations/uag/census-advisory-groups/statistical-development/2011-uk-census-coverage-assessment.pdf (accessed January 2017).*

Alfons, A. (2013), "simFrame: Simulation framework," *R package version 0.5*, 1.

Alfons, A., Filzmoser, P., Hulliger, B., Kolb, J., Kraft, S., and Münnich, R. (2011), Synthetic Data Generation of SILC Data,, Research Project Report WP6 – D6.2, University of Trier. Report WP6 D6.2, FP7-SSH-2007-217322 AMELI, `http://ameli.surveystatistics.net`.

Alfons, A., Templ, M., and Filzmoser, P. (2010), Simulation of EU-SILC population data: Using the R package simPopulation,, Technical report, Research Report CS-2010-5, Department of Statistics and Probability Theory, Vienna University of Technology.

Arcos, A., Molina, D., Rueda, M., and Ranalli, M. (2015), "Frames2: a package for estimation in dual frame surveys," *The R Journal*, 7(1), 52–72.

Baffour, B., Brown, J. J., and Smith, P. W. (2013), "An investigation of triple system estimators in censuses," *Statistical Journal of the IAOS*, 29(1), 53–68.

Baillie, M., Brown, J., Taylor, A., and Abbott, O. (2011), "Variance Estimation," *Office for National Statistics*, Technical report.

URL: *http://www.ons.gov.uk/ons/guide-method/census/2011/the-2011-census/processing-the-information/statistical-methodology/variance-estimation-in-the-2011-census.pdf (accessesd April 2017)*

Bankier, M. D. (1986), "Estimators based on several stratified samples with applications to multiple frame surveys," *Journal of the American Statistical Association*, 81(396), 1074–1079.

Barr, M. L., Van Ritten, J. J., Steel, D. G., and Thackway, S. V. (2012), "Inclusion of mobile phone numbers into an ongoing population health survey in New South Wales, Australia: design, methods, call outcomes, costs and sample representativeness," *BMC medical research methodology*, 12(1), 177.

Bates, D. (2018), *Package 'Matrix'.*

URL: *https://cran.r-project.org/web/packages/Matrix/index.html*

Berger, Y. (2018), "Empirical likelihood approaches under complex sampling designs," in *Wiley StatsRef,* Wiley.

URL: *https://doi.org/10.1002/9781118445112.stat08066*

Berger, Y., and De La Riva Torres, O. (2016), "Empirical likelihood confidence intervals for complex sampling designs," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78, 319–341.

Berger, Y. G. (2005), "Variance Estimation with Highly Stratified Sampling Designs with Unequal Probabilities," *Australian and New Zealand Journal of Statistics*, 47(3), 365–373.

Berger, Y. G. (2011), "Asymptotic consistency under large entropy sampling designs with unequal probabilities," *Pakistan Journal of Statistics, Festschrift to honour Ken Brewer's 80th birthday*, 27(4), 407–426.

Berger, Y. G. (2015), A consistent estimation approach which does not rely on confidential information specified by auxiliary survey variables,, Technical report, Southampton Statisticsl Sciences Research Institute, Southampton, UK.

Berger, Y. G. (2016), "Empirical Likelihood Inference for the

Rao-Hartley-Cochran Sampling Design," *to appear in the Scandinavian Journal of Statistics*, .

Berger, Y. G. (2017), "An empirical likelihood approach under cluster sampling with missing observations (submitted manuscript),", Southampton Statistical Sciences Research Institute.

Berger, Y. G., and De La Riva Torres, O. (2011), "Empirical likelihood ratio confidence intervals for unequal probability sampling," *Proceedings of the 58th World Statistical Congress, International Statistical Institute*, .

Berger, Y. G., and Kabzinska, E. (2017), "Empirical likelihood approach for aligning information from multiple surveys," *submitted manuscript*, .

Berger, Y. G., Tirari, M. E. H., and Tillé, Y. (2003), "Towards Optimal Regression Estimation in Sample Surveys," *Australian and New Zealand Journal of Statistics*, 45, 319–329.

Berger, Y. G., and Torres, O. D. L. R. (2014), "Empirical Likelihood Confidence Intervals: An Application to the EU-SILC Household Surveys," in *Contributions to Sampling Statistics* Springer, pp. 65–84.

Berger, Y., and Tillé, Y. (2009), "Sampling with unequal probabilities," *Handbook of Statistics: Design, Method and Applications: D. Pfeffermann and C.R. Rao.(editors). Elsevier*, 29A, 39–54.

Binder, D. A., and Patak, Z. (1994), "Use of estimating functions for estimation from complex surveys," *Journal of the American Statisticsl Association*, 89(427), 1035–1043.

Birnbaum, Z. W., and Sirken, M. G. (1965), "Design of sample surveys to estimate the prevalence of rare diseases: Three unbiased estimates," *Vital and Health Statistics*, .

URL: *https://stacks.cdc.gov/view/cdc/13016 (accessed November 2017)*

Brown, J., Abbott, O., and Diamond, I. (2006), "Dependence in the 2001 one-number census project," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169(4), 883–902.

Brown, J., Abbott, O., and Smith, P. A. (2011), "Design of the 2001 and 2011

census coverage Surveys for England and Wales," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(4), 881–906.

Brown, J., Diamond, I., Chambers, R., Buckner, L., and Teague, A. (1999), "A methodological strategy for a one-number census in the UK," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 162(2), 247–267.

Canty, A. J., and Davison, A. C. (1999), "Resampling-based variance estimation for labour force surveys," *Journal of the Royal Statistical Society: Series D (The Statistician)*, 48(3), 379–391.

Chaudhuri, S., Handcock, M. S., and Rendall, M. S. (2008), "Generalized Linear Models Incorporating Population Level Information: An Empirical-Likelihood-Based Approach," *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 70(2), 311–328.

Chen, J., and Qin, J. (1993), "Empirical likelihood estimation for finite populations and the effective usage of auxiliary information," *Biometrika*, 80(1), 107–116.

Chen, J., and Sitter, R. R. (1999), "A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys," *Statist. Sinica*, 9, 385–406.

Chen, J., Sitter, R. R., and Wu, C. (2002), "Using Empirical Likelihood Methods to Obtain Range Restricted Weights in Regression Estimators for Surveys," *Biometrika*, 89(1), 230–237.

Chen, S., and Kim, J. K. (2014), "Population empirical likelihood for nonparametric inference in survey sampling," *Statist. Sinica*, 24, 335–355.

Chipperfield, J., Brown, J., and Bell, P. (2017), "Estimating the Count Error in the Australian Census," *Journal of Official Statistics*, 33(1), 43–59.

Chipperfield, J. O., and Steel, D. G. (2009), "Design and estimation for split questionnaire surveys," , .

da Silva, A. D., de Freitas, M. P. S., and Pessoa, D. G. C. (2015), "Assessing coverage of the 2010 Brazilian Census," *Statistical Journal of the IAOS*, 31(2), 215–225.

Deville, J. C., and Särndal, C. E. (1992*a*), "Calibration Estimators in Survey Sampling," *Journal of the American Statistical Association*, 87(418), 376–382.

Deville, J.-C., and Särndal, C.-E. (1992*b*), "Calibration estimators in survey sampling," *Journal of the American statistical Association*, 87(418), 376–382.

Dong, Q., Elliott, M. R., and Raghunathan, T. E. (2014), "Combining Information from Multiple Complex Surveys.," *Survey Methodology*, 40(2), 347–354.

Efron, B. (1987), "Better bootstrap confidence intervals," *Journal of the American statistical Association*, 82(397), 171–185.

Eurostat (2012), "European Union Statistics on Income and Living Conditions (EU-SILC),", `http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/eu_silc`.

Fuller, W. A. (2009), "Some design properties of a rejective sampling procedure," *Biometrika*, 96, 933–944.

Fuller, W. A., and Burmeister, L. F. (1972), "Estimators for samples selected from two overlapping frames," *Proceedings of social science section of The American Statistical Asociation*, .

Godambe, V. P., and Thompson, M. . (2009), "Estimating functions and survey sampling," *Handbook of Statistics: Design, Method and Applications: D. Pfeffermann and C.R. Rao.(editors). Elsevier*, 29B, 83–101.

Griffin, R. A. (2014), "Potential uses of administrative records for triple system modeling for estimation of census coverage error in 2020," *Journal of Official Statistics*, 30(2), 177–189.

Hájek, J. (1964), "Asymptotic Theory of Rejective Sampling with Varying Probabilities from a Finite Population," *The Annals of Mathematical Statistics*, 35(4), 1491–1523.

Hájek, J. (1981), *Sampling from a Finite Population*, New York: Marcel Dekker.

Hansen, L. P. (1982), "Large sample properties of generalized method of moments estimators," *Econometrica: Journal of the Econometric Society*, pp. 1029–1054.

Hansen, M., Hurwitz, W., and Madow, W. (1953), *Sample Survey Methods and Theory, volume I*, New York: John Wiley and Sons.

Hartley, H. O. (1962), Multiple frame surveys,, in *Proceedings of the Social Statistics Section, American Statistical Association*, Vol. 19, Washington, DC, p. 2.

Hartley, H. O., and Rao, J. N. K. (1962), "Sampling with unequal probabilities without replacement," *Ann. math. Statist. Assoc.*, 33, 350–374.

Hartley, H. O., and Rao, J. N. K. (1968), "A new estimation theory for sample surveys," *Biometrika*, 55(3), 547–557.

Hartley, H. O., and Rao, J. N. K. (1969), *A new estimation theory for sample surveys, II*, A Symposium on the Foundations of Survey Sampling held at the University of North Carolina, Chapel Hill, North Carolina: Wiley-Interscience, New York.

Hidiroglou, M. (2001), "Double sampling," *Survey methodology*, 27(2), 143–154.

Ho, Y.-C., and Pepyne, D. L. (2001), Simple explanation of the no free lunch theorem of optimization,, in *Decision and Control, 2001. Proceedings of the 40th IEEE Conference on*, Vol. 5, IEEE, pp. 4409–4414.

Horvitz, D. G., and Thompson, D. J. (1952), "A Generalization of Sampling Without Replacement From a Finite Universe," *Journal of the American Statistical Association*, 47(260), 663–685.

Isaki, C. T., and Fuller, W. A. (1982), "Survey design under the regression super-population model," *Journal of the American Statistical Association*, 77, 89–96.

Kabzinska, E., Smith, P. A., and Berger, Y. G. (2017), "Using empirical likelihood to assess the accuracy of census coverage estimates," *submitted manuscript*, .

Kalton, G. (2009), "Methods for oversampling rare subpopulations in social surveys," *Survey Methodology*, 35(2), 125–141.

Kalton, G., and Anderson, D. W. (1986), "Sampling rare populations," *Journal of the royal statistical society. Series A (general)*, pp. 65–82.

Karlberg, M., Reis, F., Calizzani, C., and Gras, F. (2015), "A toolbox for a modular design and pooled analysis of sample survey programmes," *Statistical Journal of the IAOS*, 31(3), 447–462.

Kim, J.-k., Park, S., and Kim, S.-y. (2015), "Small area estimation combining information from several sources," *Survey Methodology*, 41(1), 21–36.

Kim, J. K., and Rao, J. N. (2012), "Combining data from two independent surveys: a model-assisted approach," *Biometrika*, 99(1), 85–100.

Krewski, D., and Rao, J. N. K. (1981), "Inference from stratified sample: properties of linearization jackknife, and balanced repeated replication methods," *The Annals of Statistics*, 9, 1010–1019.

Lavallée, P. (2009), *Indirect sampling*, Vol. 7397 Springer Science & Business Media.

Lesage, E. (2011), "The use of estimating equations to perform a calibration on complex parameters," *Survey Methodology*, 37(1), 103–108.

Liu, Y. K., and Kott, P. S. (2009), "Evaluating alternative one-sided coverage intervals for a proportion," *Journal of Official Statistics*, 25(4), 569.

Lohr, S. (2007), "Recent developments in multiple frame surveys," *cell*, 46(42.2), 6.

Lumley, T. (2016), *Package 'survey'*.

URL: *https://cran.r-project.org/web/packages/survey/survey.pdf*

Lumley, T., and Lumley, M. T. (2018), "Package survey,".

Lumley, T. et al. (2004), "Analysis of complex survey samples," *Journal of Statistical Software*, 9(1), 1–19.

Mecatti, F. (2005), A Single-Frame Multiplicity Estimator for Multiple Frame Survey,, Technical report, Universitá degli Studi di Milano.

Mecatti, F., and Singh, A. C. (2014), "Estimation in Multiple Frame Surveys: A Simplified and Unified Review using the Multiplicity Approach," *Journal de la Société Française de Statistique*, 155(4), 51–69.

Merkouris, T. (2004), "Combining independent regression estimators from multiple surveys," *Journal of the American Statistical Association*,

99(468), 1131–1139.

Merkouris, T. (2010*a*), "Combining information from multiple surveys by using regression for efficient small domain estimation," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(1), 27–48.

Merkouris, T. (2010*b*), "An Estimation Method for Matrix Survey Sampling," *Section on Survey Research Method - JSM*, .

Merkouris, T. (2015), "An efficient estimation method for matrix survey sampling," *Survey Methodology*, 41(1), 237–262.

Montanari, G. (1987), "Post sampling efficient QR-prediction in large sample survey," *International Statistical Review*, 55, 191–202.

Neyman, J. (1934), "On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection," *Journal of the Royal Statistical Society*, 97(4), 558–625.

Office for National Statistics (2012), "The 2011 Census Coverage Assessment and Adjustment Process," *Technical report*, .

Office for National Statistics (2017), "2011 Census General Report for England and Wales," *Technical report*, .

URL: *http://www.ons.gov.uk/ons/guide-method/census/2011/how-our-censusworks/how-did-we-do-in-2011-/2011-census-general-report/index.html (Accessed December 2017)*

Office for National Statistics and Department for Environment, Food and Rural Affairs (2009), "Living Costs and Food Survey [computer file],", `http://dx.doi.org/10.4225/13/511C71F8612C3`.

Office for National Statistics. Social Survey Division (2015), "Quarterly Labour Force Survey Household Dataset, October - December, 2013 [computer file],", `http://dx.doi.org/10.5255/UKDA-SN-7497-3`.

Oguz-Alper, M., and Berger, Y. G. (2016), "Modelling complex survey data with population level information: an empirical likelihood approach," *Biometrika*, 103(2), 447–459.

Owen, A. B. (1988), "Empirical Likelihood Ratio Confidence Intervals for a

Single Functional," *Biometrika*, 75(2), 237–249.

Owen, A. B. (1991), "Empirical Likelihood for Linear Models," *The Annals of Statistics*, 19(4), 1725–1747.

Owen, A. B. (2001), *Empirical Likelihood*, New York: Chapman & Hall.

Polyak, B. T. (1987), *Introduction to Optimization*, New York: Optimization Software, Inc., Publications Division.

Preston, J. (2009), "Rescaled bootstrap for stratified multistage sampling," *Survey Methodology*, 35(2), 227–234.

Qin, J., and Lawless, J. (1994), "Empirical Likelihood and General Estimating Equations," *The Annals of Statistics*, 22(1), pp. 300–325.

R Core Team (2015), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. **URL:** *https://www.R-project.org*

Ranalli, M. G., Arcos, A., del Mar Rueda, M., and Teodoro, A. (2016), "Calibration estimation in dual-frame surveys," *Statistical Methods & Applications*, 25(3), 321–349.

Rao, J. (2006), "Empirical likelihood methods for sample survey data: An overview," *Austrian Journal of Statistics*, 35(2&3), 191–196.

Rao, J. N. (2015), *Small-Area Estimation* Wiley Online Library.

Rao, J. N. K. (1965), "On two simple schemes of unequal probability sampling without replacement," *Journal of the Indian Statistical Association*, 3, 173–180.

Rao, J. N. K., and Wu, W. (2009*a*), "Empirical Likelihood Methods," *Handbook of statistics: Sample Surveys: Inference and Analysis, D. Pfeffermann and C. R. Rao eds. The Netherlands (North-Holland)*, 29B, 189–207.

Rao, J., and Wu, C. (2009*b*), Pseudo Empirical Likelihood Inference for Multiple Frame Surveys,, Technical Report 05, University of Waterloo Department ff Statistics And Actuarial Science.

Renssen, R. H., and Nieuwenbroek, N. J. (1997), "Aligning Estimates for Common Variables in Two or More Sample Surveys," *Journal of the American*

Statistical Association, 92(437), pp. 368–374.

URL: *http://www.jstor.org/stable/2291482*

Salehi, M., Mohammadi, M., Rao, J., and Berger, Y. G. (2010), "Empirical likelihood confidence intervals for adaptive cluster sampling," *Environmental and Ecological Statistics*, 17(1), 111–123.

Sampford, M. R. (1967), "On Sampling Without Replacement with Unequal Probabilities of Selection," *Biometrika*, 54(3/4), 499–513.

Särndal, C.-E. (2007), "The calibration approach in survey theory and practice," *Survey Methodology*, 33(2), 99–119.

Scott, A., and Wu, C. F. (1981), "On the asymptotic distribution of ratio and regression estimators," *Journal of the American Statistical Association*, 76, 98–102.

Singh, A. C., and Mecatti, F. (2011), "Generalized multiplicity-adjusted Horvitz-Thompson estimation as a unified approach to multiple frame surveys," *Journal of official statistics*, 27(4), 633.

Singh, A., Ganesh, N., and Lin, Y. (2013), Improved sampling weight calibration by generalized raking with optimal unbiased modification,, in *Proceedings of the American Statistical Association, Survey Research Methods Section*, pp. 3572–3583.

Singh, A., and Mecatti, F. (2014), "Use of zero functions for combining information from multiple frames," in *Contributions to Sampling Statistics* Springer, pp. 35–51.

Singh, A., and Wu, S. (2003), An extension of generalized regression estimator to dual frame surveys,, in *Proceedings of the Joint Statistical Meeting-Section on Survey Research Methods*, pp. 3911–3918.

Skinner, C. J. (1991), "On the efficiency of raking ratio estimation for multiple frame surveys," *Journal of the American Statistical Association*, 86(415), 779–784.

Skinner, C. J., and Rao, J. N. (1996), "Estimation in dual frame surveys with complex designs," *Journal of the American Statistical Association*,

91(433), 349–356.

Statistics New Zealand (2014), "Coverage in the 2013 Census based on the New Zealand 2013 Post-enumeration Survey," *Wellington: Statistics New Zealand*, .

Tillé, Y. (2006), *Sampling Algorithms*, Springer Series in Statistics, New York: Springer.

Wilks, S. S. (1938), "Shortest Average Confidence Intervals from Large Samples," *The Annals of Mathematical Statistics*, 9(3), 166–175.

Wolpert, D. H., and Macready, W. G. (1997), "No free lunch theorems for optimization," *IEEE transactions on evolutionary computation*, 1(1), 67–82.

Wu, C. (2004*a*), "Combining information from multiple surveys through the empirical likelihood method," *Canadian Journal of Statistics*, 32(1), 15–26.
URL: *http://dx.doi.org/10.2307/3315996*

Wu, C. (2004*b*), "Some algorithmic aspects of the empirical likelihood method in survey sampling," *Statistica Sinica*, 14, 1057–1067.

Wu, C. (2005), "Algorithms and R Codes for the Pseudo Empirical Likelihood Method in Survey Sampling," *Survey Methodology*, 31(2), 239–243.

Wu, C., and Rao, J. (2006), "Pseudo-empirical likelihood ratio confidence intervals for complex surveys," *Canadian Journal of Statistics*, 34(3), 359–375.

Zhang, L.-C. (2015), "On modelling register coverage errors," *Journal of Official Statistics*, 31(3), 381–396.

Zhong, B., and Rao, J. N. K. (1996), "Empirical Likelihood Inference Under Stratified Random Sampling Using Auxiliary Information," *ASA Proceedings of the Section on Survey Research Methods*, 87, 798–803.

Zhong, B., and Rao, J. N. K. (2000), "Empirical Likelihood Inference under Stratified Random Sampling Using Auxiliary Population information," *Biometrika*, 87(4), 929–938.

Zieschang, K. D. (1990), "Sample Weighting Methods and Estimation of Totals in the Consumer Expenditure Survey," *Journal of the American Statistical Association*, 85(412), pp. 986–1001.
URL: *http://www.jstor.org/stable/2289595*