Empirical likelihood approach for aligning information from multiple surveys

Yves G. Berger
University of Southampton,
SO17 1BJ, United Kingdom.
Y.G.Berger@soton.ac.uk

Ewa Kabzinska
University of Southampton,
SO17 1BJ, United Kingdom.
kabzinska.ewa@gmail.com

May 3, 2019

Abstract

When two surveys carried out separately in the same population have common variables, it might be desirable to adjust each survey's weights so that they give equal estimates for the common variables. This problem has been studied extensively and has often been referred to as '*alignment*' or '*numerical consistency*'. We develop a design-based empirical likelihood approach for alignment and estimation of complex parameters defined by estimating equations. We focus on a general case when a single set of adjusted weights, which can be applied to both common and non-common variables, is produced for each survey. The main contribution of the paper is to show that the empirical log-likelihood ratio statistic is pivotal in presence of alignment constraints. This pivotal statistic can be used to test hypotheses and derive confidence regions. Hence, the empirical likelihood approach proposed for alignment possesses the

self-normalization property, under a design-based approach. The proposed approach accommodates large sampling fractions, stratification and population level auxiliary information. It is particularly well suited for inference about small domains, when data are skewed. It includes implicit adjustments when the samples considerably differ in size. The confidence regions are constructed without the need for variance estimates, joint-inclusion probabilities, linearisation and re-sampling.

Key Words: Design-based approach, design-consistency, estimating equations, inclusion probabilities, population-level information, stratification.

1 Introduction

Suppose that we have two independent samples, $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$, selected from the same finite population $\mathcal{U}$ of size N. Let $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ be vectors of variables observed respectively for $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$. Let $\boldsymbol{x}$ denote a vector of '*common variables*' observed in both samples, which constitutes the key feature of the approach considered. A population parameter of $\boldsymbol{x}$, e.g. the mean of $\boldsymbol{x}$, can be estimated either from $\boldsymbol{S}_1$ or $\boldsymbol{S}_2$. It would be inconvenient to obtain different estimates for the same parameter, especially if other estimates are based on them. For example, suppose that $\boldsymbol{x}$ is the vector of age-sex categories measured in both samples. The two samples may not give the same estimates for the proportions within each categories. A similar situation occurs if totals of turnover for various industries are estimated from a sample $\boldsymbol{S}_1$, while a second 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 possible to obtain an estimate for the common parameter by taking a weighted average of the estimates obtained from two samples. The weight applied to each sur-

vey's estimate might be selected based on an efficiency argument, e.g. inversely proportional to the estimated variance or to the sample size. However, in practice it is desirable to have a single set of weights for each survey which can be applied to all survey variables. Furthermore, it is a common practice that weights are calibrated on known population parameters (e.g Deville & Särndal, 1992). The problem can therefore be summarised as follows: how to adjust the design weights of both surveys so that both calibration (i.e., benchmark) and alignment (i.e., numerical consistency) constraints 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, providing 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 small 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 are collected from another, smaller sample (e.g. Hidiroglou, 2001).

This problem has been studied by several authors and various estimators have been proposed. Zieschang (1990) and Renssen & Nieuwenbroek (1997) estimate the unknown population mean of $\boldsymbol{x}$ 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 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 $\boldsymbol{x}$. In Zieschang's (1990), Renssen & Nieuwenbroek's (1997) and Merkouris (2004) approaches, symmetric confidence intervals are constructed based on a suitably adjusted variance estimator.

Wu (2004) proposed an estimator for a mean of $\mathbf{y}$ based on '*aligned pseudoempirical likelihood*' weights. Symmetric confidence intervals are created using the variance estimate for the asymptotically-equivalent regression estimator. Chen & Kim (2014) developed an '*aligned population empirical likelihood*' approach, based on an empirical likelihood function defined at population level, and conjectured that the population empirical log-likelihood ratio statistic is pivotal under poisson sampling with negligible sampling fraction.

Methods outside of the design-based paradigm have also been proposed, see e.g. Kim & Rao (2012) for a model-assisted approach, Kim, Park & Kim (2015) for a model based small area application, Dong, Elliott & Raghunathan (2014) for a bayesian bootstrap approach.

We propose a new '*aligned design-based empirical likelihood*' approach, which is called '*empirical likelihood*' hereafter. The proposed approach is different from the Zieschang's (1990), Renssen & Nieuwenbroek's (1997), Merkouris's (2004) and Wu's (2004) methods, because we consider a general class of parameters which are defined by estimating equations, rather than means or totals, and allows for construction of Wilks type confidence intervals. It also differs from Chen & Kim's (2014) approach for Poisson sampling, in that it does not require N to be known, allows for large sampling fractions, and stratification and its properties are proven for a wider class of sampling designs than Poisson sampling. The empirical likelihood approach proposed for alignment possesses the self-normalization property, under a design-based approach. This property does not hold with the pseudoempirical likelihood approach.

Most of the literature on empirical likelihood assumes independent and identically distributed observations. We do not consider this assumption, because survey units are typically selected with unequal probabilities, fixed sample size, and sampling fractions that may be large. We consider Neyman's (1938) non-parametric design-

based approach, where the sampling distribution is specified by the sampling design.

The proposed approach treats the empirical likelihood function as a standard likelihood. Point estimates are obtained by maximising this function. Confidence regions are obtained from an empirical log-likelihood ratio statistic, rather than through variance estimation. The proposed method does not rely on N and on the estimation of the population mean of $\boldsymbol{x}$. It is valid under without-replacement stratified samples with small or large sampling fractions. The empirical likelihood weights are always positive.

The method proposed has some practical advantages. Confidence intervals are range-preserving and their construction does not require unknown population parameters such as variance estimates, unlike the pseudoempirical likelihood and the composite regression confidence intervals. Our simulation study shows that our confidence interval has good coverage and can be better than the symmetric confidence intervals, especially when the variables of interest are skewed or when samples of very different sizes are combined. The proposed approach also allows users to use 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 proposed approach is derived from the Berger & Torres's (2016a) empirical likelihood methodology for construction of confidence intervals in a single sample, in presence of benchmark constraints and under complex sampling designs. However, the core problem tackled here is different. Berger & Torres (2016a) deal with a traditional setup when there is a single sample and benchmark constraints only involve known population parameters. We focus on alignment of two samples and allow for constraints including unknown (yet not nuisance) parameters. The main contribution id to show that the proposed alignment-respecting empirical log-likelihood ratio

statistic is asymptotically pivotal, and can be used for tests and confidence regions. Theorem 1 shows that the proposed point estimator is $\sqrt{n}$-design-consistent.

In §2 and 3, we describe the sampling design, the variables and the parameters. The empirical likelihood function and the empirical likelihood estimator is introduced in §4. Its asymptotically design-consistency is shown in §4.1. In §5, we define the penalised empirical log-likelihood ratio statistic. In §6, we show that the empirical log-likelihood ratio statistic is asymptotically pivotal. In §7, we show how the precision is affected by alignment. A series of Monte-Carlo simulations can be found in §8.

2 Notation, sampling design, stratification, variables and sample data

Suppose that two independent stratified samples $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are selected. The sample $\boldsymbol{S}_1$ is a vector of n_1 labels from $U_1 = \{1, \ldots, N\}$. The sample $\boldsymbol{S}_2$ is a vector of n_2 labels from $U_2 = \{(N+1), \ldots, (2N)\}$. We make the convention that a label $i \in \boldsymbol{S}_2$ corresponds to the $(i-N)$-th unit of the population $\mathcal{U}$. For example, suppose $N = 10$, $\boldsymbol{S}_1 = (2, 5, 7)^\top$ and $\boldsymbol{S}_2 = (12, 16, 19)^\top$. The population units $\{12, 16, 19\} - N = \{2, 6, 9\}$ are selected in $\boldsymbol{S}_2$. The unit 2 is included in both samples. It is labelled 2 in $\boldsymbol{S}_1$ and $N + 2 = 12$ in $\boldsymbol{S}_2$. The samples may or may not overlap, because they are independent. Thus, this notation allows to distinguish if a unit has been selected in $\boldsymbol{S}_1$ and/or in $\boldsymbol{S}_2$. Consider the indicator variables

$$\delta_{1i} = \delta(i \leqslant N) \quad \text{and} \quad \delta_{2i} = \delta(i > N),$$

where $\delta(A) = 1$ if A is true and $\delta(A) = 0$ otherwise. Hence, $\delta_{ti} = 1$ for $i \in \boldsymbol{S}_t$, with $t = 1, 2$.

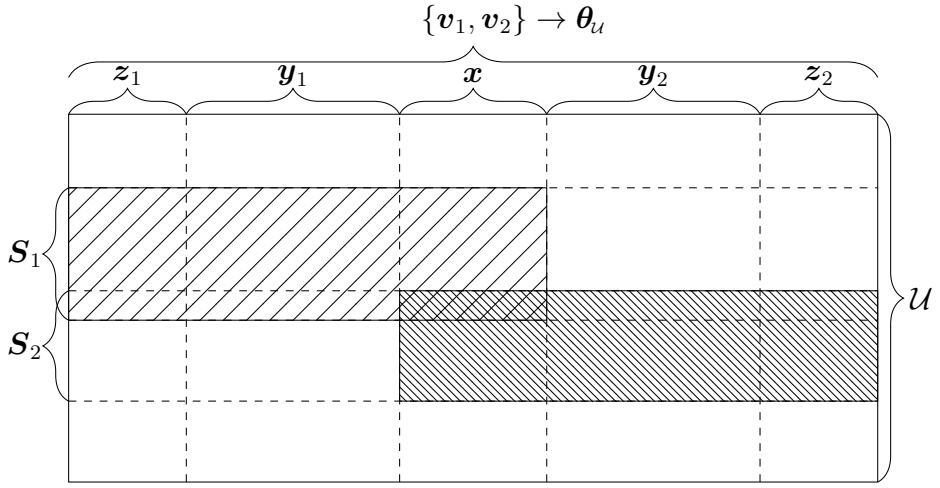

Figure 1: *The horizontal axis represents the variables:* $\boldsymbol{z}_t$, $\boldsymbol{y}_t$ *and* $\boldsymbol{x}$. *The vertical axis represents the labels of the population.* $\mathcal{D}_1$. $\mathcal{D}_2$.

The stratified samples $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are selected the following way. Consider that the population U_t is split into H_t disjoint strata $\{U_{t1}, \ldots, U_{th}, \ldots, U_{tH_t}\}$, such that $\cup_{h=1}^{H_t} U_{th} = U_t$, with $t = 1, 2$. The labels of $\boldsymbol{S}_{th}$ are selected without-replacement with unequal probabilities from U_{th}. We have $\boldsymbol{S}_t = \cup_{h=1}^{H_1} \boldsymbol{S}_{th}$ and $n_t = \sum_{h=1}^{H_t} n_{th}$. We allow for large sampling fractions; that is, n_{th}/N_{th} may be non-negligible, where N_{th} denotes the size of U_{th}. Let $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ denote the stratification variables. The values of $\boldsymbol{z}_t$ for unit i are given by the H_t-vector

$$\boldsymbol{z}_{ti} = (z_{t1i}, \ldots, z_{thi}, \ldots, z_{tH_ti})^\top, \tag{1}$$

with $z_{thi} = \pi_{ti}/n_t$ when $i \in U_{th}$ and $z_{thi} = 0$ otherwise, where π_{ti} denotes the inclusion probability of unit $i \in U_t$, with $t = 1, 2$.

Consider that the values of a vector of variables, denoted $\boldsymbol{v}_t$, are collected from the units in the sample $\boldsymbol{S}_t$. The set $\boldsymbol{v}_t$ is composed of three types of variables: $\boldsymbol{z}_t$, $\boldsymbol{y}_t$ and $\boldsymbol{x}$; that is, $\boldsymbol{v}_t \equiv \{\boldsymbol{z}_t, \boldsymbol{y}_t, \boldsymbol{x}\}$. The variables $\boldsymbol{z}_t$ are defined by (1). The variables $\boldsymbol{x}$ denote the '*common variables*' observed for both samples. The $\boldsymbol{y}_t$ represents the

remaining variables. The parameter of interest $\boldsymbol{\theta}_{\mathcal{U}}$ is a function of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. The overall sample data are $\mathcal{D} = \{\mathcal{D}_1, \mathcal{D}_2\}$, with $\mathcal{D}_t = \{\boldsymbol{v}_{ti} : \ i \in \boldsymbol{S}_t\}$, where $\boldsymbol{v}_{ti}$ denotes the vector of the values of $\boldsymbol{v}_t$ for a unit $i \in U_t$. Under the '*design-based approach*', the $\boldsymbol{v}_{ti}$ are fixed quantities and the samples are the only source of randomness (Neyman, 1938). Figure 1 is a '*spreadsheet*' visualisation of the variables and population units.

3 Parameters

We consider two types of population parameters: the '*unknown population parameter of interest*' $\boldsymbol{\theta}_{\mathcal{U}}$ and the '*known population parameter*' $\boldsymbol{\varphi}_{\mathcal{U}}$. Both parameters are vectors.

The parameter $\boldsymbol{\varphi}_{\mathcal{U}}$ is defined as the unique solution to

$$\sum_{i \in U_t} \mathbf{f}_t(\boldsymbol{v}_{ti}, \boldsymbol{\varphi}_{\mathcal{U}}) = \mathbf{0}. \tag{2}$$

The known parameter $\boldsymbol{\varphi}_{\mathcal{U}}$ will be treated as a vector of constants, which will not be estimated. Known population parameters are a common feature of sample survey data. For example, $\boldsymbol{\varphi}_{\mathcal{U}}$ can be a vector of population means, counts, proportions or ratios available from censuses or administrative records. Examples can be found in Chaudhuri, Handcock & Rendall (2008).

The unknown population parameters $\boldsymbol{\theta}_{\mathcal{U}}$ is the unique solution to p estimating equations,

$$\boldsymbol{G}(\boldsymbol{\theta}) = \{\boldsymbol{G}_1(\boldsymbol{\theta})^\top, \boldsymbol{G}_2(\boldsymbol{\theta})^\top\}^\top = \mathbf{0}, \quad \text{where} \quad \boldsymbol{G}_t(\boldsymbol{\theta}) = \sum_{i \in U} \boldsymbol{g}_t(\boldsymbol{v}_{ti}, \boldsymbol{\varphi}_{\mathcal{U}}, \boldsymbol{\theta}), \tag{3}$$

and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $\boldsymbol{\Theta}$ denotes the compact parameter space of $\boldsymbol{\theta}_{\mathcal{U}}$. Here, $U = U_1 \cup U_2$. For example, $\boldsymbol{\theta}_{\mathcal{U}}$ could be means, quantiles, ratios, generalised linear regression coefficients (Chen & Van Keilegom, 2009).

For simplicity, we replace $\mathbf{f}_t(\boldsymbol{v}_{ti}, \boldsymbol{\varphi}_u)$ and $\boldsymbol{g}_t(\boldsymbol{v}_{ti}, \boldsymbol{\varphi}_u, \boldsymbol{\theta})$ respectively by $\mathbf{f}_{ti}$ and $\boldsymbol{g}_{ti}(\boldsymbol{\theta})$. Hence, (2) and (3) reduces to

$$\sum_{i \in U_t} \mathbf{f}_{ti} = \mathbf{0} \qquad \text{and} \qquad \boldsymbol{G}_t(\boldsymbol{\theta}) = \sum_{i \in U_t} \boldsymbol{g}_{ti}(\boldsymbol{\theta}) = \mathbf{0}\cdot \tag{4}$$

We implicitly assume that $\mathbf{f}_{ti}$ and $\boldsymbol{g}_{ti}(\boldsymbol{\theta})$ are functions of only $\boldsymbol{v}_{ti}$, because $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are distinct samples. They can be non-smooth functions which are not differentiable. The $\mathbf{f}_{ti}$ are the "*auxiliary variables*".

4 Maximum empirical likelihood estimator

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

$$\ell(m) = \sum_{i \in \boldsymbol{S}} \log(m_i), \tag{5}$$

where $\boldsymbol{S} = (\boldsymbol{S}_1^\top, \boldsymbol{S}_2^\top)^\top$ denotes the pooled set of labels and

$$m_i = \delta_{1i}\ m_{1i} + \delta_{2i}\ m_{2i}\cdot$$

Here, $\log(\cdot)$ is the natural logarithm and p_{ti} are unknown positive quantities. The '*joint empirical log-likelihood function*' is defined by

$$\ell(\boldsymbol{\theta}) = \max\Big\{\ell(m)\ :\ p_i > 0,\ \textstyle\sum_{i \in \boldsymbol{S}} m_i\, \boldsymbol{c}_i^\star(\boldsymbol{\theta}) = \boldsymbol{C}^\star\Big\}\cdot \tag{6}$$

where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

$$\boldsymbol{c}_i^\star(\boldsymbol{\theta}) \ = \ \Big\{\ \boldsymbol{c}_i^\top\ ,\ \boldsymbol{g}_i(\boldsymbol{\theta})^\top\ \Big\}^\top, \quad \boldsymbol{C}^\star = \Big(\boldsymbol{C}^\top, \mathbf{0}^\top\Big)^\top,$$

$$
\begin{aligned}
\boldsymbol{c}_i &= \left(\boldsymbol{z}_i^\top \ , \ \mathbf{f}_i^\top \ , \ \boldsymbol{\xi}_i^{\circ\top}\right)^\top, && (7)\\
\boldsymbol{z}_i &= \left(\ \delta_{1i}\boldsymbol{z}_{1i}^\top \ , \ \delta_{2i}\boldsymbol{z}_{2i}^\top \ \right)^\top,\\
\boldsymbol{C} &= \left(\textstyle\sum_{i\in\boldsymbol{S}} \breve{\boldsymbol{z}}_i^\top \ , \ \mathbf{0}^\top \ , \ \mathbf{0}^\top\right)^\top, \quad \text{where } \breve{\boldsymbol{z}}_i = \boldsymbol{z}_i\pi_i^{-1}, && (8)\\
\boldsymbol{g}_i(\boldsymbol{\theta}) &= \left\{ \ \delta_{1i} \ \boldsymbol{g}_{1i}(\boldsymbol{\theta})^\top \ , \ \delta_{2i} \ \boldsymbol{g}_{2i}(\boldsymbol{\theta})^\top \ \right\}^\top,\\
\mathbf{f}_i &= \left(\ \delta_{1i}\mathbf{f}_{1i}^\top \ , \ \delta_{2i}\mathbf{f}_{2i}^\top \ \right)^\top,\\
\boldsymbol{\xi}_i^\circ &= (-1)^{\delta_{2i}} \ \boldsymbol{\xi}(\boldsymbol{x}_i) && (9)\\
\pi_i &= \delta_{1i} \ \pi_{1i} + \delta_{2i} \ \pi_{2i}.
\end{aligned}
$$

Here, $\boldsymbol{\xi}(\boldsymbol{x}_i)$ is a known function of the values $\boldsymbol{x}_i$ of $\boldsymbol{x}$. For example, if we have one common variable x, we may consider $\boldsymbol{\xi}(\boldsymbol{x}_i) = \{x_i, x_i^2, x_i^3, \delta(x_i \leqslant \alpha)\}^\top$, where x_i denotes the values of x and α is a known constant. Note that $\boldsymbol{\xi}(\boldsymbol{x}_i)$ and $\boldsymbol{x}_i$ may be of different dimensions. The $\boldsymbol{\xi}(\boldsymbol{x}_i)$ may improve the precision, when they are correlated with $\boldsymbol{g}_{ti}(\boldsymbol{\theta}_\mathcal{U})$. Various functions $\boldsymbol{\xi}(\boldsymbol{x}_i)$ can be used to improve this correlation. For example, when $\boldsymbol{g}_{ti}(\boldsymbol{\theta}_\mathcal{U})$ is the estimating function for an α-quantile of the y_i, it is recommended to use $\boldsymbol{\xi}(\boldsymbol{x}_i) = \delta(x_i \leqslant \alpha)$, when x_i and y_i are correlated.

The constraint in (6) implies $\sum_{i\in\boldsymbol{S}_1} m_{ti} \, \boldsymbol{g}_{ti}(\boldsymbol{\theta}) = \mathbf{0}$, which are the parameters' constraints. It also implies the design-constraint $\sum_{i\in\boldsymbol{S}_t} m_{ti} \ \boldsymbol{z}_{ti} = \boldsymbol{n}_t^{st} n_t^{-1}$, where $\boldsymbol{n}_t^{st} = (n_{t1}, \ldots, n_{tH_t})^\top$. This constraint includes information about the stratification and π_{ti}, which ensures design-consistency in Theorem 1. We also have the benchmark constraints

$$
\sum_{i\in\boldsymbol{S}} \widehat{m}_i \ \mathbf{f}_i = \mathbf{0}_q, \tag{10}
$$

related to $\boldsymbol{\varphi}_{\mathcal{U}}$. The constraint in (6) also implies the '*alignment constraint*',

$$\sum_{i \in \boldsymbol{S}_1} m_{1i}\ \boldsymbol{\xi}(\boldsymbol{x}_i) = \sum_{i \in \boldsymbol{S}_2} m_{2i}\ \boldsymbol{\xi}(\boldsymbol{x}_i) \cdot \tag{11}$$

The function (6) can be computed using

$$\ell(\boldsymbol{\theta}) = \sum_{i \in \boldsymbol{S}} \log\{\widehat{m}_i^{\star}(\boldsymbol{\theta})\}$$

where $\widehat{m}_i^{\star}(\boldsymbol{\theta})$ are the m_i that maximise (5), for a given $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. We assume that $\boldsymbol{\theta}$ is such that $\boldsymbol{C}^{\star}$ is an inner point of the convex conical hull formed by the sample observations $\{\boldsymbol{c}_i^{\star}(\boldsymbol{\theta}) : i \in s\}$, so that the solution $\{\widehat{m}_i^{\star}(\boldsymbol{\theta}) : i \in s\}$ exists. Using Lagrange multipliers, we have that

$$\widehat{m}_i^{\star}(\boldsymbol{\theta}) = \Big\{\pi_i + \boldsymbol{\eta}^{\star\top} \boldsymbol{c}_i^{\star}(\boldsymbol{\theta})\Big\}^{-1}, \tag{12}$$

and $\boldsymbol{\eta}^{\star}$ is such that the following constraint holds.

$$\sum_{i \in \boldsymbol{S}} m_i\ \boldsymbol{c}_i^{\star}(\boldsymbol{\theta}) = \boldsymbol{C}^{\star} \cdot \tag{13}$$

Mainstream empirical likelihood approach is based on quantities p_i which maximise $\sum_{i \in \boldsymbol{S}} \log(p_i)$ under constraints. In fact, if we substitute m_i by $(n_1 + n_2) p_i \pi_i^{-1}$, we obtain a dual representation of $\ell(\boldsymbol{\theta})$ which is similar to the mainstream empirical likelihood approach. The only difference is presence of a weight π_i^{-1} within the constraint. More details can be found in Berger (2018c,b). The quantities p_i are in fact a kind of g-weights or calibration weights. We prefer using m_i instead of p_i, because the m_i are related to traditional survey weights. It has also the merit of simplifying the notation.

The *maximum empirical likelihood point estimator* of $\boldsymbol{\theta}_{\mathcal{U}}$ is

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \ell(\boldsymbol{\theta}). \tag{14}$$

In the Appendix B of the Supplementary Material, we show that the estimator $\widehat{\boldsymbol{\theta}}$ is the solution to

$$\widehat{\boldsymbol{G}}(\boldsymbol{\theta}) = \sum_{i\in\boldsymbol{S}} \widehat{m}_i \; \boldsymbol{g}_i(\boldsymbol{\theta}) = \mathbf{0}_p, \tag{15}$$

if this solution is unique, where the $\widehat{m}_i = (\pi_i + \boldsymbol{\eta}^\top \boldsymbol{c}_i)^{-1}$ maximise (5) under $m_i > 0$ and

$$\sum_{i\in\boldsymbol{S}} m_i \; \boldsymbol{c}_i = \boldsymbol{C}. \tag{16}$$

Here, $\boldsymbol{\eta}$ is such that (16) holds. Equation (15) resembles the pseudo-likelihood estimating equations (Binder, 1983).

The $\widehat{m}_i$ are weights which depend on the known parameter $\boldsymbol{\varphi}_{\mathcal{U}}$, but do not depend on $\boldsymbol{\theta}_{\mathcal{U}}$. Note that the $\widehat{m}_i$ are always positive and satisfy the constraint (10), implying that the maximum empirical likelihood estimator $\widehat{\boldsymbol{\varphi}}$ of $\boldsymbol{\varphi}_{\mathcal{U}}$ is degenerated; that is, $\widehat{\boldsymbol{\varphi}} = \boldsymbol{\varphi}_{\mathcal{U}}$. The *calibration property* (10) is the consequence of the fact that $\boldsymbol{\varphi}_{\mathcal{U}}$ is constant within $\ell(\boldsymbol{\theta})$.

Note that also the alignment property also holds. For example, suppose that we wish to estimate the mean $\boldsymbol{\xi}_{\mathcal{U}} = N^{-1}\sum_{i\in U} \boldsymbol{\xi}(\boldsymbol{x}_i)$; that is, $\boldsymbol{\xi}_{\mathcal{U}} \subset \boldsymbol{\theta}_{\mathcal{U}}$. This can be achieved by incorporating $\boldsymbol{h}_{ti}(\boldsymbol{\xi}) = \boldsymbol{\xi}(\boldsymbol{x}_i) - Nn_t^{-1}\pi_{ti}\; \boldsymbol{\xi}$ within $\boldsymbol{g}_{ti}(\boldsymbol{\theta})$; that is, $\boldsymbol{g}_{ti}(\boldsymbol{\theta}) = \{\ldots,\; \boldsymbol{h}_{ti}(\boldsymbol{\xi})^\top,\; \ldots\}^\top$. The maximum empirical likelihood estimator $\widehat{\boldsymbol{\xi}}_t$ of $\boldsymbol{\xi}_{\mathcal{U}}$ based on the sample $\boldsymbol{S}_t$ is the solution to (15) or equivalently $\widehat{\boldsymbol{\xi}}_t = N^{-1}\sum_{i\in\boldsymbol{S}_t} \widehat{m}_i \; \boldsymbol{\xi}(\boldsymbol{x}_i)$. The constraint (11) implies $\widehat{\boldsymbol{\xi}}_1 = \widehat{\boldsymbol{\xi}}_2$. Hence, the alignment property holds, because

both samples give the same estimate for any parameters which are functions of the components of $\boldsymbol{\xi}_{\mathcal{U}}$.

4.1 Design-consistency

As $\widehat{\boldsymbol{\theta}}$ is a function of the sample data $\mathcal{D}$, the sampling distribution is specified by the sampling design, under the design-based framework. Consider a sequence of nested populations $U^{(\nu)}$ of size $N^{(\nu)}$, where $\nu = 1, 2, \ldots, \infty$ (Isaki & Fuller, 1982). Consider a sequence of samples $\boldsymbol{S}_t^{(\nu)}$ of size $n_t^{(\nu)} \leqslant N^{(\nu)}$ selected from $U^{(\nu)}$ We assume $n_t^{(\nu)} \to \infty$, as $\nu \to \infty$, with $n_1^{(\nu)}/n_2^{(\nu)}$ being a fixed constant that does not vary as $\nu \to \infty$. We assume that $n_t^{(\nu)}/N^{(\nu)} \leqslant \lambda_t < 1$, where λ_t is a fixed constant that does not vary as $\nu \to \infty$. This assumption allows for large sampling fractions. Let H_t, $N_{th}^{(\nu)}$ and $n_{th}^{(\nu)}$ denote respectively the number of strata, the population size and sample size of the stratum $U_{th}^{(\nu)}$. We assume that H_t, $N_{th}^{(\nu)}/N^{(\nu)}$ and $n_{th}^{(\nu)}/n_t^{(\nu)}$ are fixed constants that do not vary as $\nu \to \infty$. To simplify the notation, we shall drop the index ν in what follows. Let $o_{\mathcal{P}}(\cdot)$ and $O_{\mathcal{P}}(\cdot)$ be the orders of convergence in probability with respect to the sampling design.

Theorem 1. *Under the regularity conditions* (A.2)–(A.5), (A.7)–(A.12) *given in Appendix A, we have* $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_{\mathcal{U}}\| = O_{\mathcal{P}}(n^{-1/2})$, *where* $n = n_1 + n_2$ *and* $\|\cdot\|$ *denotes the Frobenius's norm.*

Theorem 1 holds whether or not $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are independent. However, independence will be necessary for self-normalising property (38).

5 Penalised empirical log-likelihood ratio function

In this §, we define a test statistic (23), and derive its asymptotic pivotal property. This statistic is based on the following *penalised empirical log-likelihood function* (Berger & Torres, 2016a),

$$\widetilde{\ell}(m) = \sum_{i\in\boldsymbol{S}}\Big\{\log(\pi_i m_i) + 1 - \pi_i m_i\Big\}. \tag{17}$$

We can interpret (17) the following way. Let $\ell(\pi) = \sum_{i\in\boldsymbol{S}} \log(\pi_i)$. We have $\ell(m) + \ell(\pi) \leqslant \sum_{i\in\boldsymbol{S}}(\pi_i m_i - 1)$, because $\log(x) \leqslant x - 1$. Hence, $\widetilde{\ell}(m)$ is the difference between $\ell(m) + \ell(\pi)$ and its upper bound $\sum_{i\in\boldsymbol{S}}(\pi_i m_i - 1)$. Thus, maximising the negative function $\widetilde{\ell}(m)$ is equivalent to minimising the difference between $\ell(m) + \ell(\pi)$ and its upper bound. Hence, maximising $\ell(m)$ is equivalent to maximising $\widetilde{\ell}(m)$.

Consider

$$\widetilde{\ell}_{\max} = \max\Big\{\widetilde{\ell}(m) \ : \ m_i > 0, \ \textstyle\sum_{i\in\boldsymbol{S}} m_i \ \widetilde{\boldsymbol{c}}_i = \widetilde{\boldsymbol{C}}\Big\}, \tag{18}$$

$$\widetilde{\ell}(\boldsymbol{\theta}) = \max\Big\{\widetilde{\ell}(m) \ : \ m_i > 0, \ \textstyle\sum_{i\in\boldsymbol{S}} m_i \ \widetilde{\boldsymbol{c}}_i^{\star}(\boldsymbol{\theta}) = \widetilde{\boldsymbol{C}}^{\star}(\boldsymbol{\theta})\Big\}, \tag{19}$$

where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

$$\widetilde{\boldsymbol{c}}_i = q_i \ \boldsymbol{c}_i, \qquad \widetilde{\boldsymbol{C}} = (\widetilde{\boldsymbol{z}}^{\top} \ , \ \widetilde{\mathbf{f}}^{\top} \ , \ \widetilde{\boldsymbol{\xi}}^{\circ\top})^{\top}, \quad q_i = (1-\pi_i)^{\frac{1}{2}},$$

$$\widetilde{\boldsymbol{z}} = \textstyle\sum_{i\in\boldsymbol{S}} q_i \ \breve{\boldsymbol{z}}_i, \qquad \widetilde{\mathbf{f}} = \textstyle\sum_{i\in\boldsymbol{S}}(q_i - 1) \ \breve{\mathbf{f}}_i, \quad \widetilde{\boldsymbol{\xi}}^{\circ} = \textstyle\sum_{i\in\boldsymbol{S}}(q_i - 1) \ \breve{\boldsymbol{\xi}}_i^{\circ}, \tag{20}$$

$$\widetilde{\boldsymbol{c}}_i^{\star}(\boldsymbol{\theta}) = \{\widetilde{\boldsymbol{c}}_i^{\top}, \widetilde{\boldsymbol{g}}_i(\boldsymbol{\theta})^{\top} \ \}^{\top}, \ \widetilde{\boldsymbol{C}}^{\star}(\boldsymbol{\theta}) = \{\widetilde{\boldsymbol{C}}^{\top}, \widetilde{\boldsymbol{g}}(\boldsymbol{\theta})^{\top}\}^{\top}, \ \widetilde{\boldsymbol{g}}(\boldsymbol{\theta}) = \textstyle\sum_{i\in\boldsymbol{S}}(q_i - 1)\breve{\boldsymbol{g}}_i(\boldsymbol{\theta}). \tag{21}$$

Here, $\breve{\boldsymbol{z}}_i = \boldsymbol{z}_i\pi_i^{-1}$, $\breve{\mathbf{f}}_i = \mathbf{f}_i\pi_i^{-1}$, $\breve{\boldsymbol{\xi}}_i^{\circ} = \boldsymbol{\xi}_i^{\circ}\pi_i^{-1}$, $\widetilde{\boldsymbol{g}}_i(\boldsymbol{\theta}) = q_i \ \boldsymbol{g}_i(\boldsymbol{\theta})$ and $\breve{\boldsymbol{g}}_i(\boldsymbol{\theta}) = \boldsymbol{g}_i(\boldsymbol{\theta})\pi_i^{-1}$.

It can be shown that

$$\begin{aligned}\widetilde{\ell}_{\max} &= \sum_{i\in\boldsymbol{S}}\Big\{\log(\pi_i\widetilde{m}_i)+1-\pi_i\widetilde{m}_i\Big\},\\ \widetilde{\ell}(\boldsymbol{\theta}) &= \sum_{i\in\boldsymbol{S}}\Big[\log\{\pi_i\widetilde{m}_i^{\star}(\boldsymbol{\theta})\}+1-\pi_i\widetilde{m}_i^{\star}(\boldsymbol{\theta})\Big],\end{aligned} \tag{22}$$

where $\widetilde{m}_i = (\pi_i + \widetilde{\boldsymbol{\eta}}^{\top}\widetilde{\boldsymbol{c}}_i)^{-1}$ and $\widetilde{m}_i^{\star}(\boldsymbol{\theta}) = \{\pi_i + \widetilde{\boldsymbol{\eta}}^{\star\top}\widetilde{\boldsymbol{c}}_i^{\star}(\boldsymbol{\theta})\}^{-1}$. Here, $\widetilde{\boldsymbol{\eta}}$ and $\widetilde{\boldsymbol{\eta}}^{\star}$ are such that the respective constraints within (18) and (19) hold.

The '*penalised empirical log-likelihood ratio statistics*' is defined by the positive function

$$\widehat{r}(\boldsymbol{\theta}) = 2\Big\{\widetilde{\ell}_{\max} - \widetilde{\ell}(\boldsymbol{\theta})\Big\}. \tag{23}$$

The advantage of empirical likelihood is the fact that (23) can be treated as a traditional likelihood ratio function to make inference about $\boldsymbol{\theta}_{\mathcal{U}}$.

The q_i within (20) and (21) are Hájek's (1964) '*finite population correction factors*' which reduces the effect of units with large π_i within $\widehat{r}(\boldsymbol{\theta})$. For large sampling fractions and moderate sample sizes, we recommend computing q_i from the recursive formula (3.25) in Hájek (1981). For negligible sampling fractions, $q_i \doteq 1$ and can be replaced by 1.

6 Asymptotic properties of the penalised empirical log-likelihood ratio function

In this §, we show that $\widehat{r}(\boldsymbol{\theta}_{\mathcal{U}})$ converges to a χ^2-distribution. Theorem 2 shows that $\widehat{r}(\boldsymbol{\theta}_{\mathcal{U}})$ can be approximated by a quadratic form. Then, using linearisation (Lemma 1), we show that the covariance matrix within the quadratic form is consistent (Lemma

2 and 3). Thus, $\widehat{r}(\boldsymbol{\theta}_{\mathcal{U}})$ converges to a χ^2-distribution because of condition (A.15) in Appendix A.

Theorem 2. *Under the conditions* (A.2)–(A.8) *given in Appendix A, we have that*

$$\widehat{r}(\boldsymbol{\theta}_{\mathcal{U}}) = \widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}})^\top \ \widehat{\boldsymbol{V}}_{\mathcal{P}}\Big\{\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}})\Big\}^{-1} \ \widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}}) + O_{\mathcal{P}}(n^{-1/2}), \tag{24}$$

where

$$\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}}) = \sum_{i\in\boldsymbol{S}} \breve{\boldsymbol{g}}_i(\boldsymbol{\theta}_{\mathcal{U}}) - \widetilde{\boldsymbol{B}}(\boldsymbol{\theta}_{\mathcal{U}})^\top \Big(\mathbf{0}_H^\top, \textstyle\sum_{i\in\boldsymbol{S}} \breve{\mathbf{f}}_i^\top, \sum_{i\in\boldsymbol{S}} \breve{\boldsymbol{\xi}}_i^{\circ\top}\Big)^\top, \tag{25}$$

$$\widehat{\boldsymbol{V}}_{\mathcal{P}}\Big\{\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}})\Big\} = \sum_{i\in\boldsymbol{S}} \frac{q_i^2}{\pi_i^2} \ \boldsymbol{\mathcal{E}}_i(\boldsymbol{\theta}_{\mathcal{U}}) \ \boldsymbol{\mathcal{E}}_i(\boldsymbol{\theta}_{\mathcal{U}})^\top. \tag{26}$$

Here, $\mathbf{0}_H$ *denotes the* $(H_1 + H_2)$*-vector of zeros and*

$$\widetilde{\boldsymbol{B}}(\boldsymbol{\theta}_{\mathcal{U}}) = \widetilde{\boldsymbol{\Sigma}}_{cc}^{-1} \ \widetilde{\boldsymbol{\Sigma}}_{cg}, \quad \widetilde{\boldsymbol{\Sigma}}_{cc} = \sum_{i\in\boldsymbol{S}} \frac{1}{\pi_i^2} \ \widetilde{\boldsymbol{c}}_i \ \widetilde{\boldsymbol{c}}_i^\top, \quad \widetilde{\boldsymbol{\Sigma}}_{cg} = \sum_{i\in\boldsymbol{S}} \frac{1}{\pi_i^2} \ \widetilde{\boldsymbol{c}}_i \ \widetilde{\boldsymbol{g}}_i(\boldsymbol{\theta}_{\mathcal{U}})^\top \tag{27}$$

$$\boldsymbol{\mathcal{E}}_i(\boldsymbol{\theta}_{\mathcal{U}}) = \boldsymbol{g}_i(\boldsymbol{\theta}_{\mathcal{U}}) - q_i^{-1}\widetilde{\boldsymbol{B}}(\boldsymbol{\theta}_{\mathcal{U}})^\top \widetilde{\boldsymbol{c}}_i. \tag{28}$$

The proof of Theorem 2 can be found in Appendix B of the supplementary materials. In this proof, we show that by using standard matrix algebra Berger & Torres's (2016b) Lemma 3 implies (24).

The main purpose of Theorem 2 is to show that the empirical log-likelihood ratio statistic $\widehat{r}(\boldsymbol{\theta})$ can be approximated by a quadratic form when $\boldsymbol{\theta} = \boldsymbol{\theta}_{\mathcal{U}}$. This is a property which constitutes the core of any proper "*likelihood-type*" approach. The merit of Theorem 2 is to show that this property is also met with aligned survey data, which is a non-standard situation that differs from the mainstream empirical likelihood approach based on independent and identically distributed observations.

In order for the right hand side of (24) to converge to a χ^2-distribution under

(A.15), we need (26) to be a consistent estimator of the variance of (25). Note that (25) is a regression estimator and (28) are the associated residuals. Furthermore, (26) takes the form of a residual variance, since the weighted mean of the residuals is zero. It remains to show that (26) is indeed consistent for the variance of (25). First, we will derive a linearised approximation of (25) in Lemma 1, and its linearised variance in Lemma 2. Finally, we show that (26) is a consistent estimator of this linearised variance.

Lemma 1. *Under conditions* (A.3) *and* (A.7)*, we have*

$$\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}}) = \widetilde{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}}) + \boldsymbol{\beta}_U^\top \boldsymbol{C}_N + \boldsymbol{\mathcal{O}}_{\mathcal{P}}(N/n) \tag{29}$$

where

$$\widetilde{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}}) = \widehat{\boldsymbol{G}}_N - \boldsymbol{\beta}_U^\top \widehat{\boldsymbol{C}}_N \cdot \tag{30}$$

$$\boldsymbol{\beta}_U = (\boldsymbol{\Sigma}_{cc}^U)^{-1}\ \boldsymbol{\Sigma}_{cg}^U, \quad \boldsymbol{\Sigma}_{cc}^U = \sum_{i\in U} \frac{1}{\pi_i}\ \widetilde{\boldsymbol{c}}_i\ \widetilde{\boldsymbol{c}}_i^\top, \quad \boldsymbol{\Sigma}_{cg}^U = \sum_{i\in U} \frac{1}{\pi_i}\ \widetilde{\boldsymbol{c}}_i\ \widetilde{\boldsymbol{g}}_i(\boldsymbol{\theta}_{\mathcal{U}})^\top \cdot \tag{31}$$

Here, $\widehat{\boldsymbol{G}}_N = \sum_{i\in \boldsymbol{S}} \breve{\boldsymbol{g}}_i(\boldsymbol{\theta}_{\mathcal{U}})$*,* $\widehat{\boldsymbol{C}}_N = \sum_{i\in \boldsymbol{S}} \breve{\boldsymbol{c}}_i$*,* $\breve{\boldsymbol{c}}_i = (\breve{\boldsymbol{z}}_i^\top, \breve{\mathbf{f}}_i^\top, \breve{\boldsymbol{\xi}}_i^{\circ\top})^\top$*,* $\boldsymbol{C}_N = (\boldsymbol{z}_N^\top, \mathbf{0}_q^\top, \mathbf{0}_r^\top)^\top$ *and* $\boldsymbol{z}_N = \sum_{i\in U} \boldsymbol{z}_i$*.*

The proof of Lemma 1 can be found in Appendix B of the supplementary materials. From (29), we see that the random vector $\widetilde{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}})$ is a linearised approximation of $\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}})$, because $\boldsymbol{\beta}_U$ is a population quantity. We shall show that $\widehat{\boldsymbol{V}}_{\mathcal{P}}\Big\{\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}})\Big\}$ is a consistent estimator of the variance of $\widetilde{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}})$. This implies that $\widehat{\boldsymbol{V}}_{\mathcal{P}}\Big\{\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}})\Big\}$ is also a consistent estimator of the variance of $\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}})$ because of Lemma 1 (e.g. Fuller, 1996; Wolter, 2007, Ch.6).

Since $\widetilde{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}})$ is a linear combination of $\widehat{\boldsymbol{G}}_N$ and $\widehat{\boldsymbol{C}}_N$, the variance of $\widetilde{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}})$ is given by (33) in the following Lemma.

Lemma 2. *Let $\boldsymbol{I}_U$ denote the $(2N) \times (2N)$ identity matrix,*

$$\begin{aligned}
\breve{\boldsymbol{\Omega}}_U &= \left\{q_i^{-1}\widetilde{\boldsymbol{c}}_i^{\top}\pi_i^{-1} : i \in U\right\}, \qquad a\ (2N) \times (H+q+r)\ \textit{matrix},\\
\breve{\boldsymbol{\Gamma}}_U &= \left\{\boldsymbol{g}_i(\boldsymbol{\theta}_{\mathcal{U}})^{\top}\pi_i^{-1} : i \in U\right\}, \qquad a\ (2N) \times p\ \textit{matrix},\\
\boldsymbol{W}_U &= \operatorname{diag}\left\{\pi_i q_i^2\ : i \in U\right\}, \qquad a\ (2N) \times (2N)\ \textit{diagonal matrix},\\
\boldsymbol{\phi}_U &= \left[\{\phi_{11},\ldots,\phi_{1N}\}^{\top}, \{\phi_{2(N+1)},\ldots,\phi_{2(2N)}\}^{\top}\right]^{\top}, \quad a\ (2N)\textit{-vector.}
\end{aligned} \tag{32}$$

where q, r and p are respectively the dimension of $\mathbf{f}_i$, $\boldsymbol{\xi}_i^{\circ}$ and $\boldsymbol{g}_i(\boldsymbol{\theta}_{\mathcal{U}})$. The ϕ_{1i} and ϕ_{2i} are the indicators of the sample $\boldsymbol{S}$ defined by $\phi_{1i} = \delta(i \in \boldsymbol{S}_1)$ and $\phi_{2i} = \delta(i \in \boldsymbol{S}_2)$. We have that

$$\boldsymbol{V}_{\mathcal{P}}\left\{\widetilde{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}})\right\} = \breve{\boldsymbol{\Gamma}}_U^{\top}\left(\boldsymbol{I}_U - \boldsymbol{P}_{\Omega}\right)^{\top}\boldsymbol{V}_{\mathcal{P}}(\boldsymbol{\phi}_U)\left(\boldsymbol{I}_U - \boldsymbol{P}_{\Omega}\right)\breve{\boldsymbol{\Gamma}}_U, \tag{33}$$

where $\boldsymbol{V}_{\mathcal{P}}(\boldsymbol{\phi}_U)$ denotes the variance of $\boldsymbol{\phi}_U$ and

$$\boldsymbol{P}_{\Omega} = \breve{\boldsymbol{\Omega}}_U\left(\breve{\boldsymbol{\Omega}}_U^{\top}\boldsymbol{W}_U\breve{\boldsymbol{\Omega}}_U\right)^{-1}\breve{\boldsymbol{\Omega}}_U^{\top}\boldsymbol{W}_U. \tag{34}$$

The variance (33) has three components: $\breve{\boldsymbol{\Gamma}}_U$, related to the estimating function; $\boldsymbol{I}_U - \boldsymbol{P}_{\Omega}$, a projection matrix into the orthogonal linear space spanned by the column of $\breve{\boldsymbol{\Omega}}_U$; and $\boldsymbol{V}_{\mathcal{P}}(\boldsymbol{\phi}_U)$, the variance between sample indicators which involves first and second order inclusion probabilities. The product $(\boldsymbol{I}_U - \boldsymbol{P}_{\Omega})\breve{\boldsymbol{\Gamma}}_U$ gives the residuals that are found in linearised variance of regression estimators.

The independence between $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ implies that $\boldsymbol{V}_{\mathcal{P}}(\boldsymbol{\phi}_U)$ is the block diagonal matrix

$$\boldsymbol{V}_{\mathcal{P}}(\boldsymbol{\phi}_U) = \begin{pmatrix} \boldsymbol{\Delta}_1 & \mathbf{0} \\ \mathbf{0}^{\top} & \boldsymbol{\Delta}_2 \end{pmatrix}, \tag{35}$$

where $\boldsymbol{\Delta}_1$ contains the covariance between the ϕ_{1i} and $\boldsymbol{\Delta}_2$ contains the covariances between the ϕ_{2i}. Hájek (1981, Ch. 14) proposed an asymptotic expression of these covariances (see also Hájek, 1964) given by

$$\widetilde{\boldsymbol{\Delta}}_t = \Big\{ \mathbb{1}_{ij}\ d_t^{-1}\pi_{ti}(\pi_{ti}-1)\pi_{tj}(1-\pi_{tj})\{1+o_{\mathcal{P}}(d_t^{-1})\} :\ i,\, j \in \mathcal{U} \Big\}, \tag{36}$$

where $\mathbb{1}_{ij} = 1$ if i and j belong to the same stratum and $\mathbb{1}_{ij} = 0$ otherwise. Here, $d_t = n_t \max\{\sum_{i\in U}\ \boldsymbol{z}_{ti}(1-\pi_{ti})\} \to \infty$.

Lemma 3. *By substituting* $\boldsymbol{\Delta}_1$ *and* $\boldsymbol{\Delta}_2$ *respectively by* $\widetilde{\boldsymbol{\Delta}}_1$ *and* $\widetilde{\boldsymbol{\Delta}}_2$ *given by* (36)*, we have that* (33) *reduces to*

$$\boldsymbol{V}_{\mathcal{P}}\Big\{\widetilde{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}})\Big\} = \sum_{i\in U}\frac{1}{\pi_i}\Big\{q_i\boldsymbol{g}_i(\boldsymbol{\theta}_{\mathcal{U}}) - \boldsymbol{\beta}_U^{\top}\widetilde{\boldsymbol{c}}_i\Big\}\Big\{q_i\boldsymbol{g}_i(\boldsymbol{\theta}_{\mathcal{U}}) - \boldsymbol{\beta}_U^{\top}\widetilde{\boldsymbol{c}}_i\Big\}^{\top}\{\boldsymbol{I}_U + \boldsymbol{\mathcal{O}}_{\mathcal{P}}(d^{-1})\}, \tag{37}$$

where $d = \max\{d_1, d_2\} \to \infty$.

A design consistent predictor of the variance (37) is obtained by substituting $\boldsymbol{\beta}_U$ in (37) by its consistent estimator $\widetilde{\boldsymbol{B}}(\boldsymbol{\theta}_{\mathcal{U}})$ and the sum in (37) by its Horvitz & Thompson (1952) estimator. This gives $\widehat{\boldsymbol{V}}_{\mathcal{P}}\Big\{\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}})\Big\}$ defined by (26). More rigorously, it can be shown that $\widehat{\boldsymbol{V}}_{\mathcal{P}}\Big\{\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}})\Big\}$ is consistent under (A.3), (A.5), (A.6), (A.13) and (A.14) (e.g. Berger, 2018d). It can be also shown that (26) exists (see Lemma 6 in supplementary materials).

Finally, if $\widetilde{\boldsymbol{\Delta}}_t$ is a "*suitable approximation*" of $\boldsymbol{\Delta}_t$ for $t = 1$ and 2, we have that $\widehat{\boldsymbol{V}}_{\mathcal{P}}\Big\{\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}})\Big\}$ is a consistent estimator of the variance of $\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}})$ (see Lemma 1 and 3). "*Suitable approximation*" means that $\boldsymbol{\Delta}_t$ equals $\widetilde{\boldsymbol{\Delta}}_t$ times a term of order $\{1 + o_{\mathcal{P}}(d_t^{-1})\}$. Thus, when $\widetilde{\boldsymbol{\Delta}}_t$ is a "*suitable approximation*", Theorem 2 implies that $\widehat{r}(\boldsymbol{\theta}_{\mathcal{U}})$

is asymptotically pivotal; that is,

$$\widehat{r}(\boldsymbol{\theta}_{\mathcal{U}}) \longrightarrow \chi^2_{df=p}, \tag{38}$$

in distribution. Here, $\chi^2_{df=p}$ denotes a χ^2-distribution with p degrees of freedom, where p is the dimension of $\boldsymbol{g}_i(\boldsymbol{\theta})$. The self-normalizing property (38) is well-known with standard empirical likelihood approach. Note that this property does not hold with aligned pseudoempirical likelihood (Wu, 2004).

An asymptotically consistent confidence region is

$$\alpha\text{-level confidence region of } \boldsymbol{\theta}_{\mathcal{U}} = \left\{\boldsymbol{\theta} : \widehat{r}(\boldsymbol{\theta}) \leqslant \chi^2_{df=p}(\alpha)\right\}, \tag{39}$$

where $\chi^2_{df=p}(\alpha)$ denotes the upper α-quantile of the χ^2-distribution. The region (39) reduces to a confidence interval, when we have a single scalar parameter defined by a single estimating equation, which does not involve any unknown parameters. The property (38) can be also used to test the statistical significance of parameters of interest.

The self-normalizing property (38) requires the independence between $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$, otherwise (35) is not a block diagonal matrix and $\widehat{\boldsymbol{V}}_{\mathcal{P}}\left\{\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}})\right\}$ would not be consistent. The property (38) relies on $\widetilde{\boldsymbol{\Delta}}_t$ being a "*suitable approximation*" of $\boldsymbol{\Delta}_t$ to ensure consistency. We now need to investigate when this would be the case. If the sampling fraction is negligible by replacing $\boldsymbol{\Delta}_t$ by $\widetilde{\boldsymbol{\Delta}}_t$, we have that $\widehat{\boldsymbol{V}}_{\mathcal{P}}\left\{\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}})\right\}$ reduces to the usual Hansen & Hurwitz's (1943) consistent variance estimator, and (38) holds. With non-negligible sampling fraction, Hájek (1981) showed that $\widetilde{\boldsymbol{\Delta}}_t$ is the matrix obtained under rejective stratified sampling. Thus, (38) also holds in this case. Hartley & Rao (1962) proposed an approximation for the joint-inclusion

probabilities of the randomized systematic design which leads to the approximation (36). This implies that (38) holds in this case. Berger (1998) showed that under large entropy designs $\widetilde{\boldsymbol{\Delta}}_t$ is also suitable. The definition of large entropy designs can be found in Berger (1998, 2011). It is based on a Kullback-Leibler divergence that tends to zero. Berger (1998) showed that Rao (1965) & Sampford's (1967) design and Chao's (1982) design are also large entropy designs (Berger, 2005).

Finally, (38) holds with negligible sampling fractions or under large entropy designs with large sampling fractions. In practice, negligible sampling fractions are often used in social surveys. The large entropy assumption is only necessary for large sampling fraction. This assumption should not be viewed as a restriction, because most designs are large entropy designs (e.g. Matei & Tillé, 2005), except of the non-randomized systematic design, for which unbiased variance estimation is impossible. We only need an asymptotically consistent variance for (38) to hold. A large entropy assumption is often used to ensure this consistency. Under a non-asymptotic setup n is fixed and $\boldsymbol{\Delta}_t$ involves joint-inclusion probabilities. Even if we knew these probabilities, they would be of little use to obtain the distribution of $\widehat{r}(\boldsymbol{\theta}_U)$. This distribution can only be obtained asymptotically, by using asymptotic joint-inclusion probabilities, implicitly given within (36). Note that $\widehat{r}(\boldsymbol{\theta}_U)$ has the advantage of recovering these asymptotic joint-inclusion probabilities.

7 Effect of alignment on precision

We have that $N^{-1}\|\widehat{\boldsymbol{G}}(\boldsymbol{\theta}_U) - \widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_U)\| = o_{\mathcal{P}}(n^{-1/2})$ (See (B.12) in Appendix B of the supplementary material). Thus, standard linearisation theory of estimating equations (e.g. Godambe & Thompson, 2009), we can show that the efficiency of $\widehat{\boldsymbol{\theta}}$ is driven by the variance of $\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_U)$. In the rest of this §, we shall examine how alignment may

reduce this variance.

It can be shown that $\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}}) = \{\widehat{\boldsymbol{G}}_{r1}(\boldsymbol{\theta}_{\mathcal{U}})^\top, \widehat{\boldsymbol{G}}_{r2}(\boldsymbol{\theta}_{\mathcal{U}})^\top\}^\top$, where

$$\widehat{\boldsymbol{G}}_{r1}(\boldsymbol{\theta}_{\mathcal{U}}) = \widehat{\boldsymbol{G}}_{1\pi}(\boldsymbol{\theta}_{\mathcal{U}}) - \widetilde{\boldsymbol{B}}_{1\mathbf{f}1}^\top \, \widehat{\mathbf{f}}_{1\pi} - \widetilde{\boldsymbol{B}}_{1\mathbf{f}2}^\top \, \widehat{\mathbf{f}}_{2\pi} + \widetilde{\boldsymbol{B}}_{1\boldsymbol{\xi}}^\top \big(\widehat{\boldsymbol{\xi}}_{2\pi} - \widehat{\boldsymbol{\xi}}_{1\pi}\big), \tag{40}$$

$$\widehat{\boldsymbol{G}}_{r2}(\boldsymbol{\theta}_{\mathcal{U}}) = \widehat{\boldsymbol{G}}_{2\pi}(\boldsymbol{\theta}_{\mathcal{U}}) - \widetilde{\boldsymbol{B}}_{2\mathbf{f}1}^\top \, \widehat{\mathbf{f}}_{1\pi} - \widetilde{\boldsymbol{B}}_{2\mathbf{f}2}^\top \, \widehat{\mathbf{f}}_{2\pi} + \widetilde{\boldsymbol{B}}_{2\boldsymbol{\xi}}^\top \big(\widehat{\boldsymbol{\xi}}_{1\pi} - \widehat{\boldsymbol{\xi}}_{2\pi}\big). \tag{41}$$

Here $\widehat{\boldsymbol{G}}_{t\pi}(\boldsymbol{\theta}_{\mathcal{U}}) = \sum_{i\in\boldsymbol{S}_t} \pi_{ti}^{-1}\boldsymbol{g}_{ti}(\boldsymbol{\theta}_{\mathcal{U}})$, $\widehat{\mathbf{f}}_{t\pi} = \sum_{i\in\boldsymbol{S}_t} \pi_{ti}^{-1}\mathbf{f}_{ti}$ and $\widehat{\boldsymbol{\xi}}_{t\pi} = \sum_{i\in\boldsymbol{S}_t} \pi_{ti}^{-1}\boldsymbol{\xi}(\boldsymbol{x}_i)$ are Horvitz & Thompson's (1952) estimators. The coefficients $\widetilde{\boldsymbol{B}}_{1\mathbf{f}1}$, $\widetilde{\boldsymbol{B}}_{1\mathbf{f}2}$, $\widetilde{\boldsymbol{B}}_{1\boldsymbol{\xi}}$, $\widetilde{\boldsymbol{B}}_{2\mathbf{f}1}$, $\widetilde{\boldsymbol{B}}_{2\mathbf{f}2}$, $\widetilde{\boldsymbol{B}}_{2\boldsymbol{\xi}}$ are sub-matrices of (27). The random variables (40) and (41) are extended regression estimators (Renssen & Nieuwenbroek, 1997; Merkouris, 2004; Berger, Muñoz & Rancourt, 2009). Based on the standard regression estimator theory, we expect a reduction in the variance when $\mathbf{f}_{ti}$ and $\boldsymbol{\xi}(\boldsymbol{x}_i)$ are correlated with $\boldsymbol{g}_i(\boldsymbol{\theta}_{\mathcal{U}})$. Note that this reduction is reflected by the residuals $\boldsymbol{\mathcal{E}}_i(\boldsymbol{\theta}_{\mathcal{U}})$ within the Hájek's (1964) variance (26).

In (40), the term $\widetilde{\boldsymbol{B}}_{1\mathbf{f}2}^\top\widehat{\mathbf{f}}_{2\pi}$ is an adjustment based on the possible correlation between $\boldsymbol{g}_{1i}(\boldsymbol{\theta}_{\mathcal{U}})$ and $\mathbf{f}_{2i}$. As the $\mathbf{f}_{2i}$ are not observed in $\boldsymbol{S}_1$, the regression coefficient $\widetilde{\boldsymbol{B}}_{1\mathbf{f}2}$ is usually negligible, unless we have a strong correlation between $\boldsymbol{\xi}(\boldsymbol{x}_i)$ and $\mathbf{f}_{2i}$. The same comment applies to $\widetilde{\boldsymbol{B}}_{2\mathbf{f}1}^\top\widehat{\mathbf{f}}_{1\pi}$ in (41).

The term $\widetilde{\boldsymbol{B}}_{1\boldsymbol{\xi}}^\top(\widehat{\boldsymbol{\xi}}_{2\pi}-\widehat{\boldsymbol{\xi}}_{1\pi})$ in (40) is the adjustment factor of the extended regression estimator, where the totals of $\boldsymbol{\xi}(\boldsymbol{x}_i)$ are estimated by $\widehat{\boldsymbol{\xi}}_{2\pi}$ instead of being known. The term $\widetilde{\boldsymbol{B}}_{2\boldsymbol{\xi}}^\top(\widehat{\boldsymbol{\xi}}_{1\pi}-\widehat{\boldsymbol{\xi}}_{2\pi})$ in (41) plays the same role. The variance (26) is inflated to take account of the fact that the totals of $\boldsymbol{\xi}(\boldsymbol{x}_i)$ are estimated. However, the $\boldsymbol{\xi}(\boldsymbol{x}_i)$ also reduces the variance. When the reduction of the variance dominates its inflation, we have a beneficial effect of $\boldsymbol{\xi}(\boldsymbol{x}_i)$. This is the case when $\boldsymbol{\xi}(\boldsymbol{x}_i)$ are strongly correlated with $\boldsymbol{g}_{ti}(\boldsymbol{\theta}_{\mathcal{U}})$. The advantage of the approach proposed is that various functions $\boldsymbol{\xi}(\boldsymbol{x}_i)$ can be used to improve this correlation. For example, suppose that x_i is correlated

with a variable y_i. When $\boldsymbol{g}_{ti}(\boldsymbol{\theta}_\iota)$ is the estimating function for an α-quantile of the y_i, it is recommended to use $\boldsymbol{\xi}(\boldsymbol{x}_i) = \delta(x_i \leqslant \alpha)$. If $\boldsymbol{g}_{ti}(\boldsymbol{\theta}_\iota)$ is the estimating function for the variance, we should use $\boldsymbol{\xi}(\boldsymbol{x}_i) = (x_i, x_i^2)^\top$.

7.1 Effect of the relative sample size

In practical applications samples $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ might have different sizes and utilise different designs. Efficient ways of using alignment constraints when samples considerably differ in size, have been investigated by both Renssen & Nieuwenbroek (1997) and Merkouris (2004, 2010, 2015). These involves introducing adjustment factors. We shall see that the aligned empirical likelihood estimator have implicit adjustments that takes into account of the relative sample size, because of the design constraint $\sum_{i\in\boldsymbol{S}} m_i \boldsymbol{z}_i = \sum_{i\in\boldsymbol{S}} \breve{\boldsymbol{z}}_i$ within (16). In the rest of this §, we show why we have implicit adjustments and why the empirical likelihood estimator is not affected by the relative sample size.

Consider a simple situation when there is no stratification and there is a single common variable ξ_i° and two scalar parameters of interest $\theta_{\iota;1}$ and $\theta_{\iota;2}$. Suppose that $\pi_{ti} = n_t/N$, and the sampling fractions are negligible; that is, we may consider $q_i = 1$. Known population parameters are not considered. In Appendix B of the supplementary material, we show that (40) and (41) reduce to

$$\widehat{G}_{r1}(\theta_{\iota;1}) = \widehat{G}_{1\pi}(\theta_{\iota;1}) - \widehat{b}_{1\xi}\big(\widehat{\xi}_{2\pi} - \widehat{\xi}_{1\pi}\big)\widehat{\zeta}_1, \tag{42}$$

$$\widehat{G}_{r2}(\theta_{\iota;2}) = \widehat{G}_{2\pi}(\theta_{\iota;2}) - \widehat{b}_{2\xi}\big(\widehat{\xi}_{1\pi} - \widehat{\xi}_{2\pi}\big)\widehat{\zeta}_2, \tag{43}$$

where $\widehat{b}_{t\xi} = \widehat{var}_{\xi t}^{-1}\widehat{\text{cov}}_t$ is the usual regression coefficient, with

$$\widehat{var}_{\xi t} = \frac{1}{n_t}\sum_{i\in\boldsymbol{S}_t}\xi_i^{\circ 2} - \Big(\frac{1}{n_t}\sum_{i\in\boldsymbol{S}_t}\xi_i^{\circ}\Big)^2, \tag{44}$$

$$\widehat{\text{cov}}_t = \frac{1}{n_t}\sum_{i\in\boldsymbol{S}_t}\xi_i^{\circ} g_{ti}(\theta) - \frac{1}{n_t^2}\sum_{i\in\boldsymbol{S}_t}\xi_i^{\circ}\sum_{i\in\boldsymbol{S}_t} g_{ti}(\theta)\cdot \tag{45}$$

The coefficients $\widehat{\zeta}_2$ and $\widehat{\zeta}_1$ are implicit adjustments that take into account of the relative sample size. They are given by

$$\widehat{\zeta}_1 = \widehat{var}_{\xi 1}\Big(\widehat{var}_{\xi 1} + \frac{n_1}{n_2}\widehat{var}_{\xi 2}\Big)^{-1}, \qquad \widehat{\zeta}_2 = \widehat{var}_{\xi 2}\Big(\frac{n_2}{n_1}\widehat{var}_{\xi 1} + \widehat{var}_{\xi 2}\Big)^{-1}\cdot$$

When $n_2/n_1 \to 0$, we have that $\widehat{\zeta}_2 \to 1$ and $\widehat{\zeta}_1 \to 0$. Thus,

$$\begin{aligned}
\widehat{G}_{r1}(\theta_{\mathcal{U};1}) &\approx \widehat{G}_{1\pi}(\theta_{\mathcal{U};1}) && \text{as } n_2/n_1 \to 0,\\
\widehat{G}_{r2}(\theta_{\mathcal{U};2}) &\approx \widehat{G}_{2\pi}(\theta_{\mathcal{U};2}) - \widehat{b}_{2\xi}\big(\widehat{\xi}_{1\pi} - \widehat{\xi}_{2\pi}\big) && \text{as } n_2/n_1 \to 0\cdot
\end{aligned}$$

This implies that when n_1 is very large compare to n_2, the estimator of the large sample $\widehat{G}_{r1}(\theta_{\mathcal{U};1})$, is approximately equal to $\widehat{G}_{1\pi}(\theta_{\mathcal{U};1})$ which is unaffected by the small sample $\boldsymbol{S}_2$. On the other hand, $\widehat{G}_{r2}(\theta_{\mathcal{U};2})$ converges to the usual regression estimator $\widehat{G}_{2\pi}(\theta_{\mathcal{U};2}) - \widehat{b}_{2\xi}(\widehat{\xi}_{1\pi} - \widehat{\xi}_{2\pi})$, which could be more efficient than $\widehat{G}_{2\pi}(\theta_{\mathcal{U};2})$. If known population parameters were used, we would have regression estimators instead of $\widehat{G}_{1\pi}(\theta_{\mathcal{U};1})$ and $\widehat{G}_{2\pi}(\theta_{\mathcal{U};2})$. In §8.2, we have a simulation study demonstrating that the empirical likelihood estimator of the large sample is not affected by alignment constraints based on a small sample.

8 Simulation studies

In §8.1 and 8.2, we compare the precision of the empirical likelihood estimator proposed with the combined pseudoempirical likelihood estimator (Wu, 2004) and the composite regression estimator (Merkouris, 2004). In §8.3, we compare the coverage of their respective confidence intervals. In Tables 1-3, 'EL' refers to the empirical likelihood approach proposed. 'PEL' refers to the combined pseudoempirical likelihood approach (Wu, 2004). 'Comp.' refers to the composite regression estimator (Merkouris, 2004). 'Adj.' refers to the adjusted composite regression estimator (Merkouris, 2004, 2010). 'Reg.' refers to standard (non-aligned) regression estimator (e.g. Särndal, Swensson & Wretman, 1992, Ch.6). All the samples are drawn using the customary randomised systematic sampling design (e.g. Hartley & Rao, 1962). The means are estimated using $\boldsymbol{g}_{ti}(\boldsymbol{\theta}) = y_{ti} - \theta_t \pi_{ti} n_t^{-1} N$, where y_{ti} denotes the variables of interest and $\boldsymbol{\theta} = (\theta_1, \theta_2)^\top$.

8.1 Point estimation

First, we consider an 'artificial population' of N values y_{ti} generated from the following model: $y_{1i} = 3 + a_{1i} + \widetilde{v}_{1i} + x_i + 0.3e_{1i}$ and $y_{2i} = 12 - a_{2i} - \widetilde{v}_{2i} - 0.5x_i - 0.2e_{2i}$, where a_{ti}, $\widetilde{v}_{ti}$ and x_i are generated independently from exponential distributions exp(1) and $e_{ti} \sim \chi^2_{df=1} - 1$. A similar artificial population was proposed by Wu & Rao (2006). The values generated are treated as fixed. We consider $N = 100\,000$, $N = 10\,000$ and $N = 2500$. The correlation between y_{1i} and y_{2i} is -0.96. These variables are skewed in opposite directions. The inclusion probabilities π_{ti} are proportional to $a_{ti} + 2$. The parameters of interest are the means of y_{1i} and y_{2i}. The common variable is $\boldsymbol{\xi}(\boldsymbol{x}_i) = x_i$. The known parameter $\boldsymbol{\varphi}_{\mathcal{U}}$ is the population mean of $\widetilde{v}_{ti}$.

[TABLE 1]

The second population data is based on the 2006 '*British Expenditure and Food Survey*' household dataset (Office for National Statistics, 2009), treated as a population of size $N = 6\,645$ households. The known parameter $\boldsymbol{\varphi}_{\mathcal{U}}$ is mean number of people living in households. The common variable $\boldsymbol{\xi}(\boldsymbol{x}_i)$ is the '*gross weekly income*'. The mean expenditure on clothing is estimated from $\boldsymbol{S}_1$. The mean expenditure on food is estimated from $\boldsymbol{S}_2$. The correlation between the variables of interest is 0.36. The inclusion probabilities are proportional to the total household expenditure and adjusted to ensure that $0.8 < \pi_{ti}\ N/n_t < 1.2$.

The third population data are the synthetic dataset AMELIA (Alfons, Filzmoser, Hulliger, Kolb, Kraft & Münnich, 2011) of $N = 3\,781\,289$ households, with variables simulated from the '*European Union statistics on income and living conditions survey*' (Eurostat, 2012). The inclusion probabilities are proportional to the tax on income and social insurance contributions and adjusted to ensure that $0.8 < \pi_{ti}\ N/n_t < 1.2$. The known parameter $\boldsymbol{\varphi}_{\mathcal{U}}$ is the mean number of households in each region. Mean 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 respectively 26%, 28%, 22% and 24% of N. The domains' means of gross household income are the parameters of interest.

For the first series of simulation, we consider $n = n_1 = n_2$. The results are given in Table 1. The *observed relative root mean square errors* (RRMSE) of the point estimator range between 0.5% and 6.5%. In all the cases, the RRMSE of the proposed estimator are lower than those of the other estimators, even with small sample size ($n = 80$, 160 and 200).

8.2 Samples of different sizes

Renssen & Nieuwenbroek (1997) pointed out that estimators based on a large sample $\boldsymbol{S}_2$ can be deteriorated by a small sample $\boldsymbol{S}_1$. In order to accommodate differential sample sizes, Merkouris (2004, 2010) considered an *adjusted composite estimator* ('Adj.'), based on adjustment factors. In this simulation study, we use the adjustment factor $n_t(1 - n_t N^{-1})^{-1}$ (Merkouris, 2010, §3.1). In §7.1, we show that the proposed empirical likelihood approach should not be affected by small sample sizes. For the next series of simulations, we show that the proposed empirical likelihood estimator performs well compared to its competitors, when n_1 is small compared to n_2. It is possible to use a better adjustment factor in the *adjusted composite estimator*. In particular, a choice that minimises the estimated variance yields and optimal estimator (e.g. Merkouris, 2004, 2010). This, however, requires variance estimation and may be difficult for complex sampling designs.

We use different artificial population data of size $N = 100\,000$ generated from the model $y_t = 3 + a + 2\widetilde{v}_{ti} + 2x_i + 0.3e_{ti}$ with $a_{ti} \sim \exp(1)$ and $e_{ti} \sim \chi^2_{df=1} - 1$. In order to investigate the effect of the skewness of $\widetilde{v}_{ti}$ and x_i, we consider $\widetilde{v}_{ti} \sim N(0,1)$ or $\widetilde{v}_{ti} \sim exp(1)$ and $x_i \sim N(5,1)$ or $x_i \sim exp(1)$ (see the columns $\widetilde{v}_{ti}$ and x_i of Table 2). We consider $\pi_{ti} \propto a_{ti} + 2$. The parameters of interest are the population means of y_{1i} and y_{2i}. The known parameter $\boldsymbol{\varphi}_{\mathcal{U}}$ is the population means of $\widetilde{v}_{ti}$. The common variable is $\boldsymbol{\xi}(\boldsymbol{x}_i) = x_i$. The aim of this simulation is to investigate the effect of a small n_1 compared to n_2. We consider $n_1 = 100$, 300, 600, 1000 and $n_2 = 1000$.

[TABLE 2]

The results are given in Table 2. We notice that the RRMSE of the estimator proposed based on $\boldsymbol{S}_1$ is smaller than those of the standard regression estimator, especially when n_1 is small. In the extreme case, when $n_1 = 100$, the RRMSE of

the proposed estimator based on $\boldsymbol{S}_1$ is between 23% and 35% of the corresponding regression estimator. The RRMSE for the proposed estimator based on $\boldsymbol{S}_2$ is only slightly larger as n_1 decreases. With the standard regression, the composite, the adjusted composite regression and the pseudoempirical likelihood estimator, we observe a much larger inflation of the RRMSE based on $\boldsymbol{S}_2$, as n_1 decreases. This deterioration is more pronounced when the common variable follows a skewed exp(1) distribution.

We observe a large RRMSE for the composite regression estimator based on $\boldsymbol{S}_2$, when n_1 is small, especially when $\widetilde{v}_t$ and x are skewed. The adjusted composite estimator has a smaller RRMSE. However, its RRMSE is still large with the small sample $\boldsymbol{S}_1$. This is more pronounced when $\widetilde{v}_t$ and x are skewed. The empirical likelihood approach does not have this drawback, because we observe small RRMSE for both samples.

8.3 Confidence intervals: British Labour Force Survey

The population data considered is a sub-sample of $N = 6\,645$ individuals selected from the October-December 2013 'British Quarterly Labour Force Survey' (Office for National Statistics, 2015). The parameter of interest is the vector of mean hourly pay for domains defined by the following industry sectors:

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

This parameter is estimated from $\boldsymbol{S}_1$. We generate 5% of outliers independently from an uniform distribution $U(y_{\max}, 3 \times y_{\max})$, where $y_{\max}$ is the maximum value of the variable of interest. We consider $n = n_1 = n_2 = 3\,000$. The known pa-

rameter $\boldsymbol{\varphi}_u$ is population proportions of the domains (i)-(viii). We use $\mathbf{f}_{1i} = [\delta\{i \in \text{Domain (i)}\}, \ldots, \delta\{i \in \text{Domain (viii)}\}]^\top - \boldsymbol{\varphi}_u$ and $\mathbf{f}_{2i} = \mathbf{0}$. The common variable $\boldsymbol{\xi}(\boldsymbol{x}_i)$ is gross weekly pay. The inclusion probabilities are proportional to the standardised net weekly income and adjusted so that $0.9 < \pi_{ti}N/n_t < 1.1$.

[TABLE 3]

Empirical likelihood confidence interval is given by (39). Composite regression confidence interval is based on variances estimates, as in Merkouris (2004). The pseudoempirical likelihood confidence interval is based on variance estimates of the generalised regression estimates, as in Wu (2004). The Hartley & Rao's (1962) estimator is used for both variance estimates. In Table 3, N_d denotes the population sizes of the d-th domain, with $d = (\text{i}), \ldots, (\text{viii})$, and $\bar{n}_d$ denotes the average sample size of the d-th domain.

In Table 3, we have the overall coverages and the tail error rates. For the domains (i)-(vi), we observe similar overall coverages. For the domain (vii), the coverage of (39) is not significantly different from 95%. The other confidence intervals have lower coverages. For the smallest domain (viii), we observe low coverages. The tail error rates are significantly different from 2.5% and unbalanced for the pseudoempirical likelihood and composite confidence intervals. The large right error rates are explained by the positive skewness of the data and the presence of outliers. The tail error rates of (39) are more balanced, because (39) is determined by the skewness of the data.

[TABLE 4]

8.4 Confidence intervals: quantiles

In this §, we consider the population data described in §8.1 and $n = n_1 = n_2$. The parameter of interest is the quantile of the variable of interest estimated from the first

sample. We use the estimating function of quantiles proposed by Berger & Torres (2016a). The composite regression and the pseudoempirical likelihood estimators are not considered in this §, because these approaches have not been developed for quantiles.

Coverages and tail error rates of (39) are given in Table 4. The overall coverages and tail error rates are of an acceptable order even with large sampling fractions, except for the smallest sample size ($n = 80$). The tail error rates are unbalanced. Further analyses, not presented here, showed that this effect is caused by skewness of the inclusion probabilities. With normally distributed inclusion probabilities, both error rates are approximately equal to 2.5%.

9 Discussion

The empirical likelihood confidence region does not rely directly on the normality of $\widehat{\boldsymbol{\theta}}$. However, it relies on the normality of $\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_\mathcal{U})$. That does not mean that the asymptotic normality of $\widehat{\boldsymbol{\theta}}$ does not hold. It can be shown that (A.15) implies that $\widehat{\boldsymbol{\theta}}$ is asymptotically normal. Standard confidence intervals based on normality of the point estimator rely on asymptotically unbiased variance estimator. This is not necessary for empirical likelihood confidence intervals, because they do not rely on variance estimates of $\widehat{\boldsymbol{\theta}}$. The confidence region (39) may be computationally demanding to construct for moderately large dimensions of the parameter. It would be useful to extend the empirical likelihood approach proposed to profiling, in order to reduce the dimensionality as in Oğuz-Alper & Berger (2016). The approach proposed is less computationally demanding, when each parameter is independently defined by a single estimating equation which does not involve unknown parameters. We considered uni-stage designs. However, the approach proposed can be extended to

multi-stage designs using an ultimate cluster approach, as in Berger (2018a).

The possibility of obtaining negative or extreme weights with large number of constraints is a common problem for calibration-type approaches. While the proposed approach gives non-negative weights, it does not protect from extreme weights, when the number of constraints is large. Iterative relaxation of calibration constraint to obtain weights range restriction as in Chen, Sitter & Wu (2002), could be extended for alignment.

Appendix A (regularity conditions)

Without loss of generality, in this Appendix, we shall substitute $\boldsymbol{c}_i$ and $\boldsymbol{C}$ by

$$\boldsymbol{c}_i = \left\{N\boldsymbol{z}_i^\top \ , \ \mathbf{f}_i^\top \ , \ \boldsymbol{\xi}_i^{\circ\top}\right\}^\top, \quad \boldsymbol{C} = \left(N\textstyle\sum_{i\in\boldsymbol{S}}\breve{\boldsymbol{z}}_i^\top \ , \ \mathbf{0}^\top \ , \ \mathbf{0}^\top\right)^\top, \tag{A.1}$$

By comparing with (7) and (8) with (A.1), we see that the only difference is within the first components, which are multiplied by N. This does not affect the constraints (13) and (16), because N appears on both sides of the constraints. Expressions (A.1) ensure that the components of $\boldsymbol{c}_i$ are bounded in probability, which is a necessary requirement for the regularity conditions to hold.

We assume that the sampling design is such that

$$\max_{i\in\boldsymbol{S}_t}\left\{\frac{N}{n_t}\pi_{ti}\right\} = O_{\mathcal{P}}(1) \text{ and } \max_{i\in\boldsymbol{S}_t}\left\{\frac{n_t}{N}\pi_{ti}^{-1}\right\} = O_{\mathcal{P}}(1) \text{ for } t = 1,\ 2, \tag{A.2}$$

$$N^{-1}(\widehat{\boldsymbol{C}}_\pi - \boldsymbol{C}) = \boldsymbol{\mathcal{O}}_{\mathcal{P}}(n^{-1/2}), \tag{A.3}$$

$$\max\{\|\boldsymbol{c}_i\| : i \in \boldsymbol{S}\} = o_{\mathcal{P}}(n^{1/2}), \tag{A.4}$$

$$n^{\tau-1}N^{-\tau}\sum_{i\in\boldsymbol{S}}\pi_i^{-\tau}\|\boldsymbol{c}_i\|^\tau = O_{\mathcal{P}}(1), \quad \tau = 2,\ 3,\ 4, \tag{A.5}$$

$$\widehat{\boldsymbol{V}} - \boldsymbol{V} = \boldsymbol{o}_{\mathcal{P}}(1) \quad \text{and} \quad \boldsymbol{V} = \boldsymbol{O}(1), \tag{A.6}$$

where $\boldsymbol{V}$ is a positive definite matrix of constants. Here, $n = n_1 + n_2$, $\widehat{\boldsymbol{V}} = nN^{-2}\sum_{i\in\boldsymbol{S}} \boldsymbol{c}_i\boldsymbol{c}_i^\top \pi_i^{-2}$ and $\widehat{\boldsymbol{C}}_\pi = \sum_{i\in\boldsymbol{S}} \boldsymbol{c}_i\pi_i^{-1}$. 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 $\|\boldsymbol{o}_{\mathcal{P}}(a)\| = o_{\mathcal{P}}(a)$.

Condition (A.2) is the key condition which ensures that the π_{ti} are of the same order as n_t/N (Isaki & Fuller, 1982; Krewski & Rao, 1981, p.1014). The condition (A.3) is a standard law of large numbers condition (e.g. Isaki & Fuller, 1982; Krewski & Rao, 1981, p. 1014). Condition (A.4) is a standard requirement (e.g. Chen & Sitter, 1999, Appendix 2). It is the consequence of Borel-Cantelli Lemma (see Owen, 2001, Lemma 11.2). Condition (A.6) means that the matrix $\widehat{\boldsymbol{V}}$ is consistent. The condition (A.5) is a Lyapunov-type condition for the existence of moments (e.g. Krewski & Rao, 1981, p. 1014, Deville & Särndal, 1992, p. 381). For conditions (A.3)-(A.5) to hold, we implicitly assume that $\boldsymbol{\xi}(\boldsymbol{x}_i) = \boldsymbol{\mathcal{O}}_{\mathcal{P}}(1)$ for all $i \in \boldsymbol{S}$, which is achieved when the components of $\boldsymbol{\xi}(\boldsymbol{x}_i)$ are bounded.

We also consider the following regularity conditions related to the parameter.

$$N^{-1}\widehat{\boldsymbol{G}}_\pi(\boldsymbol{\theta}_{\mathcal{U}}) = N^{-1}\sum_{i\in\boldsymbol{S}} \frac{1}{\pi_i}\boldsymbol{g}_i(\boldsymbol{\theta}_{\mathcal{U}}) = \boldsymbol{\mathcal{O}}_{\mathcal{P}}(n^{-1/2}), \tag{A.7}$$

$$\frac{n^{\tau-1}}{N^\tau}\sum_{i\in\boldsymbol{S}} \frac{\|\boldsymbol{g}_i(\boldsymbol{\theta}_{\mathcal{U}})\|^\tau}{\pi_i^\tau} = O_{\mathcal{P}}(1) \quad (\tau = 2,\ 3,\ 4), \tag{A.8}$$

$$\widehat{\boldsymbol{\nabla}}(\boldsymbol{\theta}) = N^{-1}\frac{\partial\widehat{\boldsymbol{G}}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \quad \text{is continuous within } \boldsymbol{\theta} \in \boldsymbol{\Theta}_{\mathcal{U}}, \tag{A.9}$$

$$N^{-1}\frac{\partial^2\widehat{\boldsymbol{G}}(\boldsymbol{\theta})}{\partial^2\boldsymbol{\theta}} = \boldsymbol{\mathcal{O}}_{\mathcal{P}}(1) \quad \text{uniformly for all } \boldsymbol{\theta} \in \boldsymbol{\Theta}_{\mathcal{U}}, \tag{A.10}$$

$$\exists\ a_1 \text{ and } a_2 :\ \mathbb{P}\Big\{0 < a_1 \leqslant \|\widehat{\boldsymbol{\nabla}}(\boldsymbol{\theta}_{\mathcal{U}})\| \leqslant a_2 < \infty\Big\} \to 1, \tag{A.11}$$

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_{\mathcal{U}} = \boldsymbol{o}_{\mathcal{P}}(1), \tag{A.12}$$

where $\boldsymbol{\Theta}_{\mathcal{U}}$ denote a compact neighbourhood containing $\boldsymbol{\theta}_{\mathcal{U}}$. Condition (A.7) assumes that the law of large numbers holds. Necessary conditions for (A.7) to hold can be

found in Isaki & Fuller (1982). Condition (A.8) is a Lyapunov-type condition for the existence of sample moments. Conditions (A.9)–(A.11) ensure that Taylor series expansion of $\widehat{\boldsymbol{G}}(\boldsymbol{\theta})$ exists. Justification for (A.12) can be found in Van Der Vaart (1998, Ch. 5). Berger & Patilea (2018) showed that (A.12) holds when we have a single sample. This proof can be extended for alignment constraints.

We assume that there exists positive random variables $\mathcal{G}_k$, $\mathcal{F}_\ell$ not depending on n and N, such that

$$\boldsymbol{E}_{\mathcal{P}}(\mathcal{G}_k) < \infty, \qquad nN^{-2} \sum_{i \in \boldsymbol{S}} \pi_i^{-2} g_{ik}(\boldsymbol{\theta}_{\mathcal{U}})^2 \leqslant \mathcal{G}_k, \quad \text{for all}\, k, \tag{A.13}$$

$$\boldsymbol{E}_{\mathcal{P}}(\mathcal{F}_\ell) < \infty, \qquad nN^{-2} \sum_{i \in \boldsymbol{S}} \pi_i^{-2} \mathrm{f}_{i\ell} \qquad \leqslant \mathcal{F}_\ell, \quad \text{for all}\, \ell, \tag{A.14}$$

where $g_{ik}(\boldsymbol{\theta}_{\mathcal{U}})$ is the k-th component of $\boldsymbol{g}_i(\boldsymbol{\theta}_{\mathcal{U}})$ and $\mathrm{f}_{i\ell}$ is the ℓ-th component of $\mathbf{f}_i$. Conditions (A.13) and (A.14) are weak conditions on $\boldsymbol{g}_i(\boldsymbol{\theta}_{\mathcal{U}})$ and $\mathbf{f}_i$, which mean that the weighted sample moments are dominated by $\mathcal{G}_k$ and $\mathcal{F}_\ell$. Theses conditions are required for the consistency of $\widehat{\boldsymbol{V}}_{\mathcal{P}}\Big\{\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}})\Big\}$.

We assume that the central limit theorem holds for $\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}})$; that is,

$$\boldsymbol{V}_{\mathcal{P}}\Big\{\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}})\Big\}^{-1/2}\, \widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}}) \longrightarrow \boldsymbol{\mathcal{N}}(\mathbf{0}_p, \boldsymbol{I}_p), \tag{A.15}$$

where $\boldsymbol{I}_p$ denotes the $p \times p$ identity matrix. Here, $\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}})$ is defined by (25), and $\boldsymbol{V}_{\mathcal{P}}\Big\{\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}})\Big\}$ is the variance of $\widehat{\boldsymbol{G}}_r(\boldsymbol{\theta}_{\mathcal{U}})$. Justification for (A.15) can be found in Fuller (2009, Ch.2) and in Scott & Wu (1981).

Acknowledgements

We wish to thank Paul Smith and Professor Li-Chun Zhang for helpful comments. Ewa Kabzinska was supported by the Economic and Social Research Council, United Kingdom.

References

Alfons, A., Filzmoser, P., Hulliger, B., Kolb, J., Kraft, S. & Münnich, R. (2011). Synthetic Data Generation of SILC Data. Research Project Report WP6 – D6.2, University of Trier. `http://ameli.surveystatistics.net`.

Berger, Y. (2005). Variance estimation with chao's sampling scheme. *Journal of Statistical Planning and Inference* **127**, 253 – 277.

Berger, Y. G. (1998). Rate of convergence to asymptotic variance for the Horvitz-Thompson estimator. *Journal of Statistical Planning and Inference* **74**, 149–168.

Berger, Y. G. (2011). Asymptotic consistency under large entropy sampling designs with unequal probabilities. *Pakistan Journal of Statistics* **27**, 407–426.

Berger, Y. G. (2018a). An empirical likelihood approach under cluster sampling with missing observations. *Annals of the Institute of Statistical Mathematics, doi:10.1007/s10463-018-0681-x* .

Berger, Y. G. (2018b). Empirical likelihood approaches in survey sampling. *The survey statistician* **78**, 22–31.

Berger, Y. G. (2018c). Empirical likelihood approaches under complex sampling designs. *Wiley StatsRef: Statistics Reference Online* , 20pp.

Berger, Y. G. (2018d). Online supplementary materials of Berger (2018a). *Annals of the Institute of Statistical Mathematics* , 13pp.

Berger, Y. G., Muñoz, J. F. & Rancourt, E. (2009). Variance estimation of survey estimates calibrated on estimated control totals - an application to the extended regression estimator and the regression composite estimator. *Computational Statistics and Data Analysis* **53**, 2596–2604.

Berger, Y. G. & Patilea, V. (2018). A semi-parametric empirical likelihood approach for conditional estimating equations under endogenous selection. *Submitted manuscript* .

Berger, Y. G., Tirari, M. E. H. & Tillé, Y. (2003). Towards optimal regression estimation in sample surveys. *Australian and New Zealand Journal of Statistics* **45**,

319–329.

Berger, Y. G. & Torres, O. D. L. R. (2016a). An empirical likelihood approach for inference under complex sampling design. *Journal of the Royal Statistical Society Series B* **78**, 319–341.

Berger, Y. G. & Torres, O. D. L. R. (2016b). Online supplementary materials of Berger & Torres (2016a). *Journal of the Royal Statistical Society Series B, http://onlinelibrary.wiley.com/doi/10.1111/rssb.12115/suppinfo* , 9pp.

Binder, D. A. (1983). On the variance of asymptotically normal estimators from complex surveys. *Int. Stat. Rev.* **51**, 279–292.

Chao, M. T. (1982). A general purpose unequal probability sampling plan. *Biometrika* **69**, 653–656.

Chaudhuri, S., Handcock, M. S. & Rendall, M. S. (2008). Generalized linear models incorporating population level information: An empirical-likelihood-based approach. *Journal of the Royal Statistical Society Series B* **70**, 311–328.

Chen, J. & 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. & Wu, C. (2002). Using empirical likelihood methods to obtain range restricted weights in regression estimators for surveys. *Biometrika* **89**, 230–237.

Chen, S. & Kim, J. K. (2014). Population empirical likelihood for nonparametric inference in survey sampling. *Statist. Sinica* **24**, 335–355.

Chen, S. & Van Keilegom, I. (2009). A review on empirical likelihood methods for regression. *Test* **18**, 415–447.

Deville, J. C. & Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association* **87**, 376–382.

Dong, Q., Elliott, M. R. & Raghunathan, T. E. (2014). Combining information from multiple complex surveys. *Survey Methodology* **40**, 347–354.

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. (1996). *Introduction to time series (2nd ed.)*. New York: John Wiley.

Fuller, W. A. (2009). Some design properties of a rejective sampling procedure. *Biometrika* **96**, 933–944.

Godambe, V. P. & Thompson, M. (2009). Estimating functions and survey sampling. In *Sample Surveys: Inference and Analysis*, D. Pfeffermann & C. Rao, eds.,

Handbook of Statistics. Amsterdam: Elsevier, pp. 83–101.

Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *The Annals of Mathematical Statistics* **35**, 1491–1523.

Hájek, J. (1981). *Sampling from a Finite Population*. New York: Marcel Dekker.

Hansen, M. H. & Hurwitz, W. N. (1943). On the theory of sampling from finite populations. *The Annals of Mathematical Statistics* **14**, pp. 333–362.

Hartley, H. O. & Rao, J. N. K. (1962). Sampling with unequal probabilities without replacement. *The Annals of Mathematical Statistics* **33**, 350–374.

Harville, D. A. (1997). *Matrix Algebra From a Statistician's Perspective*. Springer.

Hidiroglou, M. (2001). Double sampling. *Survey methodology* **27**, 143–154.

Horvitz, D. G. & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* **47**, 663–685.

Isaki, C. T. & Fuller, W. A. (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association* **77**, 89–96.

Kim, J.-k., Park, S. & Kim, S.-y. (2015). Small area estimation combining information from several sources. *Survey Methodology* **41**, 21–36.

Kim, J. K. & Rao, J. N. (2012). Combining data from two independent surveys: a model-assisted approach. *Biometrika* **99**, 85–100.

Krewski, D. & Rao, J. N. K. (1981). Inference from stratified sample: properties of linearization jackknife, and balanced repeated replication methods. *Annals of Statistics* **9**, 1010–1019.

Matei, A. & Tillé, Y. (2005). Evaluation of variance approximations and estimators in maximum entropy sampling with unequal probability and fixed sample size. *Journal of Official Statistics* **21**, 543–570.

Merkouris, T. (2004). Combining independent regression estimators from multiple surveys. *Journal of the American Statistical Association* **99**, 1131–1139.

Merkouris, T. (2010). Combining information from multiple surveys by using regression for efficient small domain estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72**, 27–48.

Merkouris, T. (2015). An efficient estimation method for matrix survey sampling. *Survey Methodology* **4**, 237–262.

Neyman, J. (1938). 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**, 558–625.

Office for National Statistics (2009). Living costs and food survey [computer file], department for environment, food and rural affairs. http://dx.doi.org/10.4225/13/511C71F8612C3.

Office for National Statistics (2015). Quarterly labour force survey household dataset, october - december, 2013 [computer file]. social survey division. http://dx.doi.org/10.5255/UKDA-SN-7497-3.

Oğuz-Alper, M. & Berger, Y. G. (2016). Empirical likelihood approach for modelling survey data. *Biometrika* **103**, 447–459.

Owen, A. B. (2001). *Empirical Likelihood*. New York: Chapman & Hall.

Rao, J. N. K. (1965). On two simple schemes of unequal probability sampling without replacement. *Journal of the Indian Statistical Association* **3**, 173–180.

Renssen, R. H. & Nieuwenbroek, N. J. (1997). Aligning estimates for common variables in two or more sample surveys. *Journal of the American Statistical Association* **92**, pp. 368–374.

Sampford, M. R. (1967). On sampling without replacement with unequal probabilities of selection. *Biometrika* **54**, 499–513.

Särndal, C.-E., Swensson, B. & Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Scott, A. & Wu, C. (1981). On the asymptotic distribution of ratio and regression estimators. *Journal of the American Statistical Association* **76**, 98–102.

Van Der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.

Wolter, K. M. (2007). *Introduction to Variance Estimation*. New York: Springer, 2nd ed.

Wu, C. (2004). Combining information from multiple surveys through the empirical likelihood method. *Canadian Journal of Statistics* **32**, 15–26.

Wu, C. & Rao, J. N. K. (2006). Pseudo-empirical likelihood ratio confidence intervals for complex surveys. *Canad. J. Statist.* **34**, 359–375.

Zieschang, K. D. (1990). Sample weighting methods and estimation of totals in the consumer expenditure survey. *Journal of the American Statistical Association* **85**, pp. 986–1001.

Table 1: Relative root mean squared errors (%) for estimators of means. Artificial and Expenditure & Food: 10 000 samples. AMELIA: 3 000 iterations. $n = n_1 = n_2$

Populations	n	Sample 1			Sample 2		
		Prop.	Pseudo	Comp.	Prop.	Pseudo	Comp.
Artificial ($N = 100\,000$)	1000	0.5	0.9	2.3	1.3	1.4	2.8
	200	1.1	2.1	5.0	2.8	3.1	6.4
Artificial ($N = 10\,000$)	500	0.7	1.3	3.1	1.7	1.9	4.0
Artificial ($N = 2500$)	250	1.0	1.8	4.5	2.4	2.6	5.5
	160	1.2	2.3	5.4	3.1	3.3	6.7
	80	1.9	3.5	8.0	4.6	5.0	10.0
Expenditure & Food ($N = 6\,645$)	500	6.4	6.5	6.5	3.0	3.1	3.4
	1000	4.3	4.4	4.4	2.0	2.1	2.3
AMELIA ($N = 3\,781\,289$)							
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

Table 2: Relative root mean squared errors (%) of estimator of means based on 10 000 samples.

Populations			Sample 1					Sample 2 ($n_2 = 1000$)				
n_1	$\widetilde{v}_{ti}$	x_i	Prop.	Pseudo	Comp.	Adjusted	Reg.	Prop.	Pseudo	Comp.	Adjusted	Reg.
100	$N(0,1)$	$N(5,1)$	0.9	0.9	2.9	1.5	3.3	0.8	1.1	3.6	2.1	0.9
	exp(1)	$N(5,1)$	1.5	1.7	3.4	2.9	4.3	1.2	3.6	10.2	5.7	1.2
	$N(0,1)$	exp(1)	1.4	1.7	3.3	2.7	4.1	1.1	3.5	9.8	5.5	1.2
	exp(1)	exp(1)	1.1	1.6	4.0	3.6	4.9	0.9	2.4	11.0	6.0	1.5
300	$N(0,1)$	$N(5,1)$	0.8	0.6	1.2	0.9	1.7	0.8	0.7	1.7	1.1	0.9
	exp(1)	$N(5,1)$	1.2	1.3	1.5	2.1	2.4	1.1	2.1	5.0	3.2	1.2
	$N(0,1)$	exp(1)	1.2	1.3	1.3	2.1	2.3	1.1	2.1	5.0	3.2	1.2
	exp(1)	exp(1)	0.9	1.1	1.8	2.6	2.8	0.8	1.4	5.5	3.5	1.5
600	$N(0,1)$	$N(5,1)$	0.7	0.5	0.7	0.7	1.1	0.7	0.5	0.9	0.8	0.9
	exp(1)	$N(5,1)$	1.0	1.2	1.4	1.9	1.6	1.0	1.5	2.9	2.3	1.2
	$N(0,1)$	exp(1)	1.0	1.2	1.3	1.9	1.6	1.0	1.5	2.9	2.3	1.2
	exp(1)	exp(1)	0.7	0.9	1.7	2.2	2.0	0.7	1.1	3.3	2.6	1.5
1000	$N(0,1)$	$N(5,1)$	0.6	0.4	0.7	0.7	0.9	0.6	0.4	0.7	0.7	0.9
	exp(1)	$N(5,1)$	0.9	1.2	1.8	1.8	1.2	0.9	1.2	1.8	1.8	1.2
	$N(0,1)$	exp(1)	0.9	1.2	1.8	1.8	1.2	0.9	1.2	1.8	1.8	1.2
	exp(1)	exp(1)	0.7	0.9	2.1	2.1	1.5	0.7	0.9	2.1	2.1	1.5

Table 3: British Quarterly Labour Force Survey. Coverages and tail error rates of confidence intervals. Mean number of hours worked per week per domains. 10 000 samples. The values reported in this Table do not reflect estimates of the British Labour Force Survey

Domains	Coverages			Left tail err. rates			Right tail err. rates				
(Sectors)	Prop.	Pseudo	Comp.	Prop.	Pseudo	Comp.	Prop.	Pseudo	Comp.	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

†: Coverages (or tail error rates) significantly different from 95% (or 2.5%). p-value $\leqslant 0.05$.

Table 4: Confidence intervals' coverages. Left and right tail error rates. 80% and 90% Quantiles. Artificial and Expenditure & Food: 10 000 samples. AMELIA: 3 000 samples. The values reported do not reflect the estimates of the British Expenditure and Food Survey

Population	n	80% Quantiles			90% Quantiles		
		Coverages	Left	Right	Coverages	Left	Right
Artificial ($N = 100\,000$)	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 ($N = 10\,000$)	500	95.3	2.9†	1.9†	94.7	2.9†	2.4
Artificial ($N = 2\,500$)	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†

†: Coverages (or tail error rates) significantly different from 95% (or 2.5%). p-value $\leqslant 0.05$.