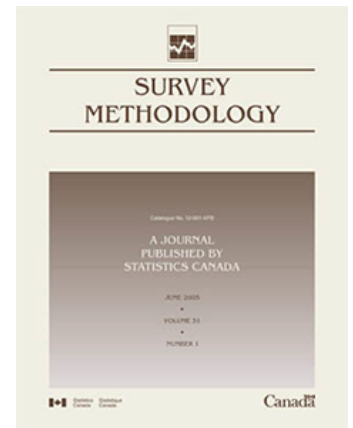## Survey Methodology

A method to find an efficient and robust sampling strategy under model uncertainty

by Edgar Bueno and Dan Hedlin

Statistics Canada   Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at STATCAN.infostats-infostats.STATCAN@canada.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

• Statistical Information Service	1-800-263-1136
• National telecommunications device for the hearing impaired	1-800-363-7629
• Fax line	1-514-283-9350

Depository Services Program

• Inquiries line	1-800-635-7943
• Fax line	1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

A method to find an efficient and robust sampling strategy under model uncertainty

Edgar Bueno and Dan Hedlin[1]

Abstract

We consider the problem of deciding on sampling strategy, in particular sampling design. We propose a risk measure, whose minimizing value guides the choice. The method makes use of a superpopulation model and takes into account uncertainty about its parameters through a prior distribution. The method is illustrated with a real dataset, yielding satisfactory results. As a baseline, we use the strategy that couples probability proportional-to-size sampling with the difference estimator, as it is known to be optimal when the superpopulation model is fully known. We show that, even under moderate misspecifications of the model, this strategy is not robust and can be outperformed by some alternatives.

Key Words: Sampling design; GREG estimator; Risk Measure.

1. Introduction

We consider the problem of choosing strategy, in particular the design, for the estimation of the total of a study variable in a finite population when a set of J auxiliary variables is available in a list sampling frame. We focus on the estimation of the total.

The decision about sampling strategy involves parameters which are unknown at the stage when the decision needs to be taken. After data collection the parameters can be estimated, although sometimes only under some assumptions. In practice, we often use data from previous waves of a repeated survey, frame variables or data from another survey that is similar to the one at planning stage. There is a risk that the available data do not give reliable information about relevant parameters. The method presented here involves a risk measure, which takes into account the possibility of being misled by inaccurate or incorrect beliefs about the values of the needed parameters. The risk measure is derived for the difference and the generalized regression estimators. Other than that, the measure is general. This measure and the discussion of its practical use are the main result of this paper.

One aim when selecting and devising the sampling strategy is efficiency in terms of small mean-squared error. The definition of "efficiency" is not unique, however, as it depends on the inference approach. Under the design-based approach, Godambe (1955), Lanke (1973) and Cassel, Särndal and Wretman (1977) show that there is no uniformly best linear estimator, in the sense of being best for all populations. There is no best design either. Therefore, a traditional approach for defining the strategy has been to assume that the finite population is a realization of some superpopulation model. The strategy is then defined in such a way that it minimizes the model expected value of the design mean-squared error, a parameter called anticipated mean-squared error. The adjective "anticipated" was first introduced by Isaki

and Fuller (1982) to emphasize the fact that this is a conceptual mean-squared error which is calculated in advance to sampling, based only on information available prior to sampling.

Assuming that a superpopulation model holds and its parameters are known, several authors have shown that the optimal strategy should make use of a probability proportional-to-size sampling design (e.g., Hájek, 1959; Cassel, Särndal and Wretman, 1976; Nedyalkova and Tillé, 2008). In practice, however, there is not even a consensus about the existence of a generating model, let alone what model to rely on. And even if there is a model, its parameters are unknown. There is evidence, rather empirical, that probability proportional-to-size sampling is not robust towards model misspecifications (e.g., Holmberg and Swensson, 2001). A second result of this paper is to provide some theoretical evidence of this fact.

Many articles discuss robustness in the survey sampling field. Beaumont, Haziza and Ruiz-Gazen (2013), for instance, propose a robust estimator that downweights influential observations; Royall and Herson (1973) consider robustness under polynomial models; Bramati (2012) and Zhai and Wiens (2015) propose robust stratification methods. We provide theoretical evidence of lack of robustness of proportional-to-size sampling and propose a method for assisting in the decision about the sampling design.

The contents of the paper are arranged as follows. The optimal strategy under the superpopulation model is defined in Section 2. The lack of robustness of this strategy when the model is misspecified is studied in Section 3. The method for assisting on the choice of the sampling design is presented in Section 4. In Section 5, the risk measure introduced in the previous section is extended to be used together with the GREG estimator. Section 6 presents numerical illustrations of the results in the paper. First, we illustrate the lack of robustness of probability proportional-to-size sampling and the flexibility of the GREG estimator with a small simulation study. Second, we illustrate the implementation of the risk measure with real survey data. Finally, Section 7 presents some conclusions.

2.   Optimal strategy under the superpopulation model

Let U be a finite population of size N with elements labeled $\{1, 2, \ldots, k, \ldots, N\}$. Let $x_k = (x_{1k}, x_{2k}, \ldots, x_{Jk})$ be a known vector of values of J auxiliary variables and y_k the unknown value of a study variable associated to unit $k \in U$. We are interested in the estimation of the total of y, $t_y = \sum_U y_k$.

Let Ω be the power set of U. A *sample* is any subset $s \in \Omega$ and a *sampling design* is a probability distribution on Ω, denoted by $P(S = s)$ or simply $p(s)$. Let $\pi_k = \sum_{s \ni k} p(s)$ be the *inclusion probability* of k and $\pi_{kl} = \sum_{s \supset \{k,l\}} p(s)$ the *joint inclusion probability* of k and l. A *probability sampling design* is a sampling design such that $\pi_k > 0$ for all $k \in U$.

An *estimator* is a real valued function of the sample, $\hat{t}_y = \hat{t}_y(S)$. By *strategy* we refer to the couple sampling design and estimator, $(p(\cdot), \hat{t}_y)$.

We consider only probability sampling designs with fixed sample size. As a convenient stepping stone we begin by considering unbiased linear estimators of the form

$$\hat{t}_y = \left(\sum_U z_k - \sum_s \frac{z_k}{\pi_k} \right) + \sum_s \frac{y_k}{\pi_k} = \sum_U z_k + \sum_s \frac{e_k}{\pi_k} \tag{2.1}$$

with z_k arbitrary known constants and $e_k = y_k - z_k$. This estimator is called the *difference estimator*. The estimator defined in this way is said to be *calibrated* on z as it satisfies $\hat{t}_z = \sum_U z_k$. Note that if $z_k = 0$ for all $k \in U$ the estimator reduces to $\hat{t}_y = \sum_s y_k / \pi_k$, that is, the Horvitz-Thompson estimator (Horvitz and Thompson, 1952). In later sections we focus on the generalized regression estimator (GREG).

The design Mean Squared Error (MSE) of the difference estimator is

$$\text{MSE}_p \left(\hat{t}_y \right) = \text{MSE}_p \left(\sum_s \frac{e_k}{\pi_k} \right) = \sum_U \sum_U (\pi_{kl} - \pi_k \pi_l) \frac{e_k}{\pi_k} \frac{e_l}{\pi_l}. \tag{2.2}$$

As mentioned in the introduction, due to the non-existence of an optimal strategy under the design-based approach, often a superpopulation model, ξ_0, is proposed and we search for an optimal strategy with respect to the *anticipated mean-squared error*,

$$\text{MSE}_{\xi_0 p} \left(\hat{t}_y \right) = \text{E}_{\xi_0} \text{MSE}_p \left(\hat{t}_y \right) = \text{E}_{\xi_0} \text{E}_p \left(\left(\hat{t}_y - t_y \right)^2 \right). \tag{2.3}$$

We may assume that the y-values are realizations of the following model, denoted ξ_0,

$$Y_k = f \left(x_k \mid \delta_1 \right) + \varepsilon_k$$

with

$$\text{E}_{\xi_0} \left(\varepsilon_k \right) = 0, \quad \text{V}_{\xi_0} \left(\varepsilon_k \right) = \sigma_0^2 g \left(x_k \mid \delta_2 \right)^2 \quad \text{and} \quad \text{E}_{\xi_0} \left(\varepsilon_k \varepsilon_l \right) = 0 \ \forall \ k \neq l \tag{2.4}$$

where $\delta = (\delta_1, \delta_2)$ is a vector of parameters, $f: \mathcal{R}^J \rightarrow \mathcal{R}$ and $g: \mathcal{R}^J \rightarrow \mathcal{R}^+$. The random sample s and the errors ε_k are assumed to be independent. Following Rosén (2000), the terms $f \left(x_k \mid \delta_1 \right)$ and $g \left(x_k \mid \delta_2 \right) > 0$ will be called *trend* and *spread*, respectively. The term trend should not in general be understood in a temporal sense, rather it refers to the development of y-values with x.

Note that under ξ_0, e_k in the difference estimator (2.1) is a random variable that represents the distance between the value of the study variable and z_k, i.e., $e_k = f \left(x_k \mid \delta_1 \right) + \varepsilon_k - z_k$. Therefore $\text{E}_{\xi_0} e_k = f \left(x_k \mid \delta_1 \right) - z_k$ and $\text{E}_{\xi_0} e_k^2 = \left(f \left(x_k \mid \delta_1 \right) - z_k \right)^2 + \sigma_0^2 g \left(x_k \mid \delta_2 \right)^2$. With some algebra, it can be seen from (2.2) and (2.3) that the anticipated MSE of the difference estimator becomes

$$\text{MSE}_{\xi_0 p} \left(\hat{t}_y \right) = \text{MSE}_p \left(\sum_s \frac{f \left(x_k \mid \delta_1 \right) - z_k}{\pi_k} \right) + \sigma_0^2 \sum_U \left(\frac{1}{\pi_k} - 1 \right) g \left(x_k \mid \delta_2 \right)^2 \tag{2.5}$$

Nedyalkova and Tillé (2008) derive the anticipated MSE in a more general case.

Tillé and Wilhelm (2017) give the anticipated MSE of the Horvitz-Thompson estimator. The second term in (2.5) is the Godambe-Joshi lower bound (e.g., Särndal, Swensson and Wretman, 1992, page 453).

The anticipated MSE in (2.5) is the sum of two positive terms. It is easy to see that if

1. the estimator is calibrated on $z_k = f(x_k | \delta_1)$ the first term vanishes and the anticipated MSE equals the Godambe-Joshi lower bound

$$\text{MSE}_{\xi_0 \text{P}}(\hat{t}_y) = \sigma_0^2 \sum_U \left(\frac{1}{\pi_k} - 1 \right) g(x_k | \delta_2)^2. \tag{2.6}$$

Furthermore, after imposing the fixed sample size restriction $\sum_U \pi_k = n$, if

2. the design is such that $\pi_k \propto g(x_k | \delta_2)$, denoted $\pi\text{ps}(\delta_2)$, the second term is minimized and we obtain

$$\text{MSE}_{\xi_0 \text{P}}^{\text{opt}}(\hat{t}_y) = \sigma_0^2 \left(\frac{1}{n} \left(\sum_U g(x_k | \delta_2) \right)^2 - \sum_U g(x_k | \delta_2)^2 \right).$$

Conditions 1 and 2 suggest the specific roles of the design and the estimator in the sampling strategy. The estimator should "explain" the trend in the calibration sense of condition 1. The design should "explain" the spread. A strategy that satisfies conditions 1 and 2 simultaneously will be called optimal. In the same sense, any estimator and any design satisfying, respectively, condition 1 and 2, will be called optimal. As this strategy plays an important role in this paper, we will denote it by $\pi\text{ps}(\delta_2) - \text{diff}(\delta_1)$.

3. Robustness under a misspecified model

If the finite population is a realization of the superpopulation model (2.4), and if f, g and δ were known, then an optimal strategy could be defined. In this section we study the robustness of this strategy when the model is misspecified.

We begin by defining how "misspecification" shall be understood in this paper. The *working model* ξ_0 reflects the beliefs the statistician has about the relation between the auxiliary variables x and the study variable y at the design stage. We shall assume that a true, unknown model ξ exists. Any deviation of ξ_0 with respect to ξ is a misspecification of the model. In order to keep the analysis tractable, we limit ourselves to the situation when the working model is of the form (2.4) and the true model, ξ, is

$$Y_k = f(x_k | \beta_1) + \varepsilon_k$$

with

$$\text{E}_\xi(\varepsilon_k) = 0, \quad \text{V}_\xi(\varepsilon_k) = \sigma^2 g(x_k | \beta_2)^2 \quad \text{and} \quad \text{E}_\xi(\varepsilon_k \varepsilon_l) = 0 \; \forall \; k \neq l \tag{3.1}$$

where $\beta = (\beta_1, \beta_2)$ is a vector of parameters, f and g as in (2.4) and $\beta \neq \delta$. The random sample s and the errors ε_k are assumed to be independent.

Result 1. If ξ_0 is assumed when ξ is the true superpopulation model, the model expected value of the design MSE in (2.2), under the difference estimator satisfying condition 1 above, becomes

$$\text{MSE}_{\xi\text{p}}\left(\hat{t}_y\right) \;=\; \text{MSE}_{\text{p}}\left(\sum_s \frac{f\left(x_k \mid \beta_1\right) - f\left(x_k \mid \delta_1\right)}{\pi_k}\right) + \sigma^2 \sum_U \left(\frac{1}{\pi_k} - 1\right) g\left(x_k \mid \beta_2\right)^2. \qquad (3.2)$$

The result is proven by noting that $f\left(x_k \mid \delta_1\right)$ takes the role of z_k in (2.5) and by taking into account that $e_k = f\left(x_k \mid \beta_1\right) - f\left(x_k \mid \delta_1\right) + \varepsilon_k$, therefore $\text{E}_\xi e_k = f\left(x_k \mid \beta_1\right) - f\left(x_k \mid \delta_1\right)$ and $\text{E}_\xi e_k^2 = \left(f\left(x_k \mid \beta_1\right) - f\left(x_k \mid \delta_1\right)\right)^2 + \sigma^2 g\left(x_k \mid \beta_2\right)^2$. As the model is misspecified, we have deliberately avoided the use of the adjective "anticipated" in Result 1.

Using Result 1, it can be seen that for a design that satisfies condition 2 we obtain

$$\text{MSE}_{\xi,\pi\text{ps}}\left(\hat{t}_y\right) = \left(\frac{\sum_U g\left(x_k \mid \delta_2\right)}{n}\right)^2 \text{MSE}_{\pi\text{ps}}\left(\sum_s \frac{f\left(x_k \mid \beta_1\right) - f\left(x_k \mid \delta_1\right)}{g\left(x_k \mid \delta_2\right)}\right)$$

$$+ \; \sigma^2 \sum_U \left(\frac{n \sum_U g\left(x_k \mid \delta_2\right)}{g\left(x_k \mid \delta_2\right)} - 1\right) g\left(x_k \mid \beta_2\right)^2. \qquad (3.3)$$

It is now possible to see that, even under a mild misspecification as the one considered here, the strategy $\pi\text{ps}\left(\delta_2\right) - \text{diff}\left(\delta_1\right)$ is not optimal anymore, as its MSE (3.3) can be greater than the MSE obtained under other designs (3.2). In particular, the strategy using the correct model, i.e., $z_k = f\left(x_k \mid \beta_1\right)$ into the estimator and a design such that $\pi_k \propto g\left(x_k \mid \beta_2\right)$, would be more efficient than $\pi\text{ps}\left(\delta_2\right) - \text{diff}\left(\delta_1\right)$.

4. Guiding the choice of sampling design with the help of a risk measure

We have seen in Section 3 that even a simple misspecification of the working model might result in the strategy $\pi\text{ps}\left(\delta_2\right) - \text{diff}\left(\delta_1\right)$ not being optimal. It is therefore risky to accept a given model as correct without any type of assessment. While most of the information needed for an "objective" evaluation of the model is not available at the design stage, it is possible to reach some degree of confidence about the parameters in the working model that allows for comparing a set of designs and make the decision about which one to implement. In this section we propose a method to assist in the choice of the sampling design.

The model expected MSE (3.2) in Result 1 can be viewed as a function of β and σ^2, as everything else is available at the design stage. To begin with, let us assume that σ^2 is also known. Then we can write

$$L_{\text{p}}\left(\beta\right) \;=\; \text{MSE}_{\xi\text{p}}\left(\beta \mid x, \delta, \sigma\right) = \text{MSE}_{\text{p}}\left(\sum_s \frac{f\left(x_k \mid \beta_1\right) - f\left(x_k \mid \delta_1\right)}{\pi_k}\right) + \sigma^2 \sum_U \left(\frac{1}{\pi_k} - 1\right) g\left(x_k \mid \beta_2\right)^2.$$

For any design, $p(\cdot)$, this function can be evaluated at any β and it indicates the loss incurred by assuming that δ is the true parameter when it is, in fact, β. We can assume a prior distribution on β, $h(\beta)$, and calculate the risk under h,

$$R(\mathrm{p}) \;=\; \mathrm{E}_h\left(\mathrm{MSE}_{\xi\mathrm{p}}\left(\beta \,|\, x,\, \delta,\, \sigma\right)\right) = \int_\Theta h(\beta) \;\cdot\; \mathrm{MSE}_{\xi\mathrm{p}}\left(\beta \,|\, x,\, \delta,\, \sigma\right) d\beta, \tag{4.1}$$

where Θ is the sample space of β. The design that yields the smallest risk should be chosen. Note that numerical integration methods (e.g., Monte Carlo simulation methods) may be needed to evaluate the risk (4.1).

In practice, σ^2 is unknown. We propose two ways for dealing with it. The first one is to see now the loss as a function of β and σ and calculate the risk as above, assuming a prior on the pair β and σ. The second one is to provide some "guess" about its value. This approach can use the fact that (Proof in the Appendix)

$$\sigma^2 \;\approx\; \frac{S_{f,f}}{\overline{g}^{\,2}}\left(\frac{1}{R^2_{f,y}} - 1\right) \tag{4.2}$$

where $S_{f,f} = \sum_U \left(f\left(x_k \,|\beta_1\right) - \overline{f}\right)^2 \big/ N$, $\overline{f} = \sum_U f\left(x_k \,|\beta_1\right)\big/N$, $\overline{g}^{\,2} = \sum_U g\left(x_k \,|\beta_2\right)^2 \big/ N$ and $R_{f,y}$ is the correlation between $f\left(x\,|\beta_1\right)$ and y. (In Example 3 below, we give a more convenient expression in a special case.) Although $R_{f,y}$ is unknown, for repeated surveys we do have some previous knowledge about it. In other cases it is often possible to have some reasonable "guess" about it.

It remains to comment on the choice of the prior distribution $h(\beta)$. The choice of the distribution and its parameters is subjective and defined by the statistician. Nevertheless, it should reflect the available knowledge about the model parameter β. In particular, $h(\beta)$ should be centered around $\beta = \delta$. Its variance should reflect how confident we are about the working model. Note that a full confidence on the working model would be a density with all its mass at $\beta = \delta$, in which case the risk (4.1) would be minimized by the $\pi\mathrm{ps}$ design given by condition 2 in Section 2.

It might be argued that by introducing $h(\beta)$ an additional source of subjectivity has been added to the choice of the sampling design. The prior may add a certain Bayesian flavor to the process, but note that $h(\beta)$ is only needed for choosing the design. Hence, the inference is still design-based. Furthermore, relying on an assumed model is also subjective in choice of assumption and it does involve a risk. The risk measure in (4.1) allows for quantification of that risk.

5. The risk measure under the Generalized Regression Estimator

The difference estimator (2.1) requires that δ_1 is fully specified in order to calculate $f\left(x_k \,|\delta_1\right)$, which is undesirable from a practical standpoint. The generalized regression (GREG) estimator is an alternative that allows for the estimation of all or some of the components of δ_1 at the cost of introducing a small bias. In this section we adapt the material in Sections 2 to 4 to strategies using the GREG estimator.

We define the generalized regression estimator in a more general way than in Särndal et al. (1992) as follows. Let a_k $(k = 1, \ldots, N)$ be a weight defined by the statistician and $\delta_1 = (\delta_1^*, \delta_1^{**})$ where δ_1^* is fixed and δ_1^{**} is to be estimated. Let also

$$\hat{\delta}_{1s}^{**} = \operatorname{argmin}_{\delta_1^{**}} \sum_s \frac{(y_k - f(x_k \mid \delta_1))^2}{a_k \pi_k}$$

and $\hat{\delta}_{1s} = (\delta_1^*, \hat{\delta}_{1s}^{**})$. The GREG estimator is

$$\hat{t}_{\text{greg}} = \left(\sum_U f(x_k \mid \hat{\delta}_{1s}) - \sum_s \frac{f(x_k \mid \hat{\delta}_{1s})}{\pi_k} \right) + \sum_s \frac{y_k}{\pi_k}. \tag{5.1}$$

An approximation to the design MSE of the GREG estimator is of the form (2.2) with $e_k = y_k - f(x_k \mid \hat{\delta}_{1U})$ where $\hat{\delta}_{1U} = (\delta_1^*, \hat{\delta}_{1U}^{**})$ and

$$\hat{\delta}_{1U}^{**} = \operatorname{argmin}_{\delta_1^{**}} \sum_U \frac{(y_k - f(x_k \mid \delta_1))^2}{a_k}.$$

Example 1. Let us consider the case where $f(x_k \mid \delta_1) = \delta_{1,1} x_{1k}^{\delta_{1,J+1}} + \delta_{1,2} x_{2k}^{\delta_{1,J+2}} + \ldots + \delta_{1,J} x_{Jk}^{\delta_{1,2J}}$. Let $\delta_1^* = (\delta_{1,J+1}, \ldots, \delta_{1,2J})$, $\delta_1^{**} = (\delta_{1,1}, \ldots, \delta_{1,J})'$ and $x_k^\delta = (x_{1k}^{\delta_{1,J+1}}, \ldots, x_{Jk}^{\delta_{1,2J}})$. In this case we obtain

$$\hat{\delta}_{1s}^{**} = \left(\sum_s \frac{x_k^{\delta'} x_k^\delta}{a_k \pi_k} \right)^{-1} \sum_s \frac{x_k^{\delta'} y_k}{a_k \pi_k} \quad \text{and} \quad \hat{\delta}_{1U}^{**} = \left(\sum_U \frac{x_k^{\delta'} x_k^\delta}{a_k} \right)^{-1} \sum_U \frac{x_k^{\delta'} y_k}{a_k}.$$

Letting the exponents $\delta_1^* = (\delta_{1,J+1}, \ldots, \delta_{1,2J}) = (1, \ldots, 1)$, we obtain the classical expression of the GREG estimator found in Särndal et al. (1992).

Example 2. The case with only one auxiliary variable, i.e., $f(x_k \mid \delta_1) = \delta_{10} + \delta_{11} x_k^{\delta_{12}}$ with $a_k = 1$, $\delta_1^* = \delta_{12}$ and $\delta_1^{**} = (\delta_{10}, \delta_{11})'$ is known as the regression estimator. In this case we obtain the well known result that the design MSE can be approximated by expression (2.2) with $e_k = y_k - f(x_k \mid \hat{\delta}_{1U})$ where $f(x_k \mid \hat{\delta}_{1U}) = \hat{\delta}_{10} + \hat{\delta}_{11} x_k^{\delta_{12}}$ and

$$\hat{\delta}_{11} = \frac{N \sum_U x_k^{\delta_{12}} y_k - \sum_U x_k^{\delta_{12}} \sum_U y_k}{N \sum_U x_k^{2\delta_{12}} - \left(\sum_U x_k^{\delta_{12}} \right)^2} \quad \text{and} \quad \hat{\delta}_{10} = \frac{1}{N} \sum_U y_k - \hat{\delta}_{11} \frac{1}{N} \sum_U x_k^{\delta_{12}}.$$

The misspecified model

Let us consider again the situation where the statistician uses the working model (2.4) but the true model is of the form (3.1) with $\beta_1 = (\beta_1^*, \beta_1^{**})$, where β_1^* is the counterpart of the fixed component δ_1^*. The following result states a condition under which Result 1 is valid for the GREG estimator.

Result 2. If ξ_0 is assumed when ξ is the true superpopulation model, $\hat{\delta}_{1s}^{**} \to \hat{\delta}_{1U}^{**}$ as $n \to \infty$ and $\hat{\delta}_{1U}^{**}$ converges to some δ_1^{**} as $N \to \infty$, then

$$\text{MSE}_{\xi p}\left(\hat{t}_{greg}\right) \to \text{MSE}_p\left(\sum_s \frac{f\left(x_k \mid \beta_1\right) - f\left(x_k \mid \delta_1\right)}{\pi_k}\right) + \sigma^2 \sum_U \left(\frac{1}{\pi_k} - 1\right) g\left(x_k \mid \beta_2\right)^2 \qquad (5.2)$$

where $\delta_1 = \left(\delta_1^*, \delta_1^{**}\right)$.

Proof. Note that if $\hat{\delta}_{1s}^{**} \to \delta_{1U}^{**}$, then $\hat{\delta}_{1s} = \left(\delta_1^*, \hat{\delta}_{1s}^{**}\right) \to \left(\delta_1^*, \hat{\delta}_{1U}^{**}\right) = \hat{\delta}_{1U}$. Thus, $f\left(x_k \mid \hat{\delta}_{1s}\right) \to f\left(x_k \mid \hat{\delta}_{1U}\right)$. In turn, if $\hat{\delta}_{1U}^{**} \to \delta_1^{**}$, then $\hat{\delta}_{1U} = \left(\delta_1^*, \hat{\delta}_{1U}^{**}\right) \to \left(\delta_1^*, \delta_1^{**}\right) = \delta_1$. Thus $f\left(x_k \mid \hat{\delta}_{1U}\right) \to f\left(x_k \mid \delta_1\right)$. Therefore,

$$\text{MSE}_{\xi p}\left(\hat{t}_{greg}\right) = \text{MSE}_{\xi p}\left(\left(\sum_U f\left(x_k \mid \hat{\delta}_{1s}\right) - \sum_s \frac{f\left(x_k \mid \hat{\delta}_{1s}\right)}{\pi_k}\right) + \sum_s \frac{y_k}{\pi_k}\right)$$

$$\to \text{MSE}_{\xi p}\left(\left(\sum_U f\left(x_k \mid \delta_1\right) - \sum_s \frac{f\left(x_k \mid \delta_1\right)}{\pi_k}\right) + \sum_s \frac{y_k}{\pi_k}\right),$$

which, by Result 1, is (5.2).

Example 3 (Continuation of Example 1). Let the working model be as in Example 1 and the true model be $f\left(x_k \mid \beta_1\right) = \beta_{1,1} x_{1k}^{\beta_{1,J+1}} + \beta_{1,2} x_{2k}^{\beta_{1,J+2}} + \ldots + \beta_{1,J} x_{Jk}^{\beta_{1,2J}}$. Let also $\beta_1^* = \left(\beta_{1,J+1}, \ldots, \beta_{1,2J}\right)$, $\beta_1^{**} = \left(\beta_{1,1}, \ldots, \beta_{1,J}\right)'$ and $x_k^\beta = \left(x_{1k}^{\beta_{1,J+1}}, \ldots, x_{Jk}^{\beta_{1,2J}}\right)$. In this case, $\hat{\delta}_{1U}^{**} \to A\beta_1^{**}$, where

$$A = \left(\sum_U \frac{x_k^{\delta'} x_k^\delta}{a_k}\right)^{-1} \sum_U \frac{x_k^{\delta'} x_k^\beta}{a_k},$$

and (5.2) becomes

$$\text{MSE}_{\xi p}\left(\hat{t}_{greg}\right) \to \text{MSE}_p\left(\sum_s \frac{\left(x_k^\beta - x_k^\delta A\right)\beta_1^{**}}{\pi_k}\right) + \sigma^2 \left(\sum_U \frac{g\left(x_k \mid \beta_2\right)^2}{\pi_k} - \sum_U g\left(x_k \mid \beta_2\right)^2\right). \qquad (5.3)$$

Example 4 (Continuation of Example 2). Let the working model be as in Example 2 and the true model be $f\left(x_k \mid \beta_1\right) = \beta_{10} + \beta_{11} x_k^{\beta_{12}}$ with $\beta_1^* = \beta_{12}$ and $\beta_1^{**} = \left(\beta_{10}, \beta_{11}\right)'$. It can be shown that (5.2) becomes

$$\text{MSE}_{\xi p}\left(\hat{t}_{greg}\right) \to \beta_{11}^2 \text{MSE}_p\left(\sum_s \frac{v_k}{\pi_k}\right) + \sigma^2 \left(\sum_U \frac{g\left(x_k \mid \beta_2\right)^2}{\pi_k} - \sum_U g\left(x_k \mid \beta_2\right)^2\right) \qquad (5.4)$$

with

$$v_k = \left(x_k^{\beta_{12}} - \overline{x}^{\beta_{12}}\right) - \left(x_k^{\delta_{12}} - \overline{x}^{\delta_{12}}\right) \frac{S_{\beta,\delta}}{S_{\delta,\delta}}, \qquad (5.5)$$

and

$$\overline{x}^{\beta_{12}} = \frac{1}{N} \sum_U x_k^{\beta_{12}} \qquad S_{\beta,\delta} = \frac{1}{N-1} \sum_U \left(x_k^{\beta_{12}} - \overline{x}^{\beta_{12}}\right)\left(x_k^{\delta_{12}} - \overline{x}^{\delta_{12}}\right)$$

$$\overline{x}^{\delta_{12}} = \frac{1}{N} \sum_U x_k^{\delta_{12}} \qquad S_{\delta,\delta} = \frac{1}{N-1} \sum_U \left(x_k^{\delta_{12}} - \overline{x}^{\delta_{12}}\right)^2.$$

Note that (5.4) does not depend on β_{10}.

For the particular case developed in Examples 2 and 4, where $f(x_k | \beta) = \beta_{10} + \beta_{11} x_k^{\beta_{12}}$ and $f(x_k | \delta) = \delta_{10} + \delta_{11} x_k^{\delta_{12}}$, an alternative approximation of σ^2 is (Proof in the Appendix)

$$\sigma^2 \approx \beta_{11}^2 F_0 \quad \text{with} \quad F_0 = \frac{1}{\bar{x}^{2\beta_2}} \frac{S_{1,\beta}^2}{S_{1,1}} \left(\frac{1}{R_{x,y}^2} - \frac{1}{R_{1,\beta}^2} \right) \tag{5.6}$$

where

$$\bar{x}^{2\beta_2} = \frac{1}{N} \sum_U x_k^{2\beta_2} \quad S_{1,\beta} = \frac{1}{N} \sum_U (x_k - \bar{x})(x_k^{\beta_{12}} - \bar{x}^{\beta_{12}}) \quad S_{1,1} = \frac{1}{N} \sum_U (x_k - \bar{x})^2$$

with $|R_{x,y}| \leq |R_{1,\beta}|$ and $R_{1,\beta}$ and $R_{x,y}$ are, respectively, the correlation coefficients between x and $x^{\beta_{12}}$ and between x and y. The latter is unknown but often some decent guess about it is available.

The approximation of σ^2 in (5.6) is more convenient than the one in (4.2) as now we have that (5.4) is approximated by

$$\text{MSE}_{\xi p}(\hat{t}_{\text{greg}}) \approx \beta_{11}^2 \left[\text{MSE}_p \left(\sum_s \frac{v_k}{\pi_k} \right) + F_0 \left(\sum_U \frac{g(x_k | \beta)^2}{\pi_k} - \sum_U g(x_k | \beta)^2 \right) \right] \tag{5.7}$$

with v_k given by (5.5). This expression depends neither on the intercept β_{01} nor the parameter σ, and the slope β_{11} becomes a proportionality constant that can be ignored.

The risk measure

As in Section 4, the asymptotic model expected MSE of the GREG estimator given by Result 2 can be seen as the loss incurred by assuming that δ is the true parameter when it is, in fact, β. Assuming a prior distribution on β, the risk (4.1) can be calculated.

6. Numerical examples

In Sections 2 and 3 we have established that the strategy $\pi \text{ps}(\delta_2) - \text{diff}(\delta_1)$ is optimal under a superpopulation model, but it is not robust to misspecifications of this model. In Subsection 6.1 we present a small Monte Carlo simulation study carried out to illustrate these results by comparing the optimal strategy and three alternatives.

In Sections 4 and 5 we introduced a measure that allows for quantifying the risk of implementing a sampling design, so allowing to guide the choice of design. In Subsection 6.2 we illustrate the use of the risk measure with real survey data.

6.1 Simulation study under a misspecified model

We compare the efficiency and robustness of four strategies through a simulation study. The strategies to be compared are πps together with the difference estimator (which is optimal when the model is correct), πps together with the GREG estimator (optimal design), stratified simple random sampling

(STSI) together with the difference estimator (optimal estimator) and STSI together with the GREG estimator.

Our implementation of πps makes use of Pareto πps (Rosén, 1997). There is a host of other schemes for drawing πps samples. Nevertheless, Pareto πps is a convenient method with good properties, see for example Rosén (2000).

Our implementation of STSI makes use of model-based stratification (Wright, 1983). We consider $H = 5$ strata with boundaries defined using Dalenius and Hodges (1959) $\mathrm{cum}\sqrt{f}$-rule on $g(x_k | \delta_2)$ which is well described in (Särndal et al., 1992, page 463) and the sample is allocated using Neyman allocation, $n_h \propto N_h S_{gh}$. Using the $\mathrm{cum}\sqrt{f}$-rule may be suboptimal (see Särndal et al., 1992, page 464) but the efficiency of stratification by a continuous size variable is fairly insensitive to the exact choice of boundaries.

We consider only misspecification of the spread. The trend term is of the form $f(x_k | \beta_1) = \beta_{10} + \beta_{11} x_k^{\beta_{12}}$ with $\beta_{10} = 1{,}000$, $\beta_{11} = 1$ and $\beta_{12} = 0.75$, 1 and 1.25. The true spread is $g(x_k | \beta_2) = x_k^{\beta_2}$ with $\beta_2 = 0.5, 0.75$ and 1. The working spread is $g(x_k | \delta_2) = x_k^{\delta_2}$ with $\delta_2 = 0.5, 0.75$ and 1.

We will use the difference estimator (2.1) calibrated on $f(x_k | \beta_1)$. Regarding the GREG estimator, we will fix β_{12}, whereas the coefficients β_{10} and β_{11} will be estimated.

The simulation is set out as follows. The population size is $N = 5{,}000$. The x-values are independent realizations from a gamma distribution with shape $\alpha = 4/100$ and scale $\lambda = 1{,}200$ plus one unit, whereas y_k is a realization from a gamma distribution with shape and scale

$$
\alpha_k = \frac{\left(\beta_{10} + \beta_{11} x_k^{\beta_{12}}\right)^2}{\sigma_0^2 x_k^{2\beta_2}} \quad \text{and} \quad \lambda_k = \frac{\sigma_0^2 x_k^{2\beta_2}}{\beta_{10} + \beta_{11} x_k^{\beta_{12}}},
$$

where σ^2 was set in such a way that the correlation between x and y is $\rho = 0.95$. The design MSE of a sample of size $n = 500$ is then computed for each strategy. Holding the x-values fixed, the process is iterated $B = 5{,}000$ times.

Table 6.1 shows the results of the simulation study. The first three columns indicate the model parameters. The fourth column shows the (simulated) model expected MSE of the strategy πps $-$ dif, whereas the last three columns show the (simulated) efficiency of the strategies πps $-$ GREG, STSI $-$ dif and STSI $-$ GREG compared to πps $-$ dif (as a percentage), with efficiency defined as eff $=$ $\mathrm{MSE}_{\xi, \pi\mathrm{ps}}(\hat{t}_y)/\mathrm{MSE}_{\xi, p}(\hat{t}_y)$ where the model expected MSEs are approximated by their simulated counterparts,

$$
\mathrm{MSE}_{\xi, \mathrm{p}}(\hat{t}_y) = \mathrm{E}_\xi \mathrm{MSE}_\mathrm{p}(\hat{t}_y) \approx \frac{1}{B} \sum_{r=1}^{B} \mathrm{MSE}_\mathrm{p}^{(r)}(\hat{t}_y),
$$

in such a way that a value of 100 indicates that the strategy is as efficient as πps $-$ dif and values smaller (larger) than 100 indicate that the strategy is less (more) efficient than πps $-$ dif.

The upper part of Table 6.1 shows the case when the working model coincides with the true model. As expected, the strategy that couples πps with the difference estimator (πps – dif) was always more efficient than the remaining strategies. Nevertheless, the loss in efficiency due to estimating some parameters through the GREG estimator is negligible. On the other hand, there is a remarkable loss in efficiency due to the use of STSI instead of πps. Finally, it is noted from (2.6) that as the anticipated MSE for all strategies does not depend on the trend f but only on the spread g, the efficiency remains constant under the same value of δ_2, independently of the value of β_{12}.

Table 6.1
Efficiency of three strategies as a percentage of the model expected MSE of πps – dif

			Correct model			
β_{12}	β_2	δ_2	πps – dif	πps – GREG	STSI – dif	STSI – GREG
0.75	0.50	0.50	$2.78 \cdot 10^5$	99.9	57.3	57.3
0.75	0.75	0.75	$4.82 \cdot 10^4$	99.6	77.9	77.9
0.75	1.00	1.00	$1.90 \cdot 10^4$	99.0	83.2	83.2
1.00	0.50	0.50	$7.64 \cdot 10^6$	99.9	57.3	57.3
1.00	0.75	0.75	$7.20 \cdot 10^5$	99.7	77.9	77.9
1.00	1.00	1.00	$2.14 \cdot 10^5$	99.1	83.1	83.1
1.25	0.50	0.50	$1.46 \cdot 10^8$	99.9	57.3	57.3
1.25	0.75	0.75	$7.85 \cdot 10^6$	99.7	77.9	78.0
1.25	1.00	1.00	$1.81 \cdot 10^6$	99.2	83.1	83.1
			Misspecified model			
β_{12}	β_2	δ_2	πps – dif	πps – GREG	STSI – dif	STSI – GREG
0.75	0.50	0.75	$3.98 \cdot 10^5$	99.9	98.9	98.9
0.75	0.75	1.00	$6.45 \cdot 10^4$	99.5	114.5	114.4
0.75	1.00	0.50	$4.73 \cdot 10^4$	100.1	133.9	134.0
1.00	0.50	1.00	$2.14 \cdot 10^7$	99.9	185.6	185.6
1.00	0.75	0.50	$1.03 \cdot 10^6$	100.1	93.1	93.2
1.00	1.00	0.75	$2.77 \cdot 10^5$	99.8	88.9	89.0
1.25	0.50	0.75	$2.09 \cdot 10^8$	99.9	98.9	98.9
1.25	0.75	1.00	$1.05 \cdot 10^7$	99.6	114.5	114.5
1.25	1.00	0.50	$4.50 \cdot 10^6$	100.3	134.0	134.2

The lower part of Table 6.1 shows some comparisons under a misspecified model, in particular, a misspecified spread. It can be noted that even under this mild misspecification of the model, πps – dif is not necessarily the best strategy anymore as the strategies using STSI were more efficient in several cases. However, it is not evident when will STSI be more efficient than πps or vice versa. The risk measure introduced in Section 4 can be used to guide the choice between designs. The results shown in this section agree with those shown by for example Holmberg and Swensson (2001).

6.2 Using the risk measure for choosing the design in a real survey

In this subsection we illustrate the implementation of the risk measure using data from a real survey. We want to estimate $t_y = \sum_U y_k$ where U is the set of residential properties in Bogotá, Colombia (of size $N = 681{,}276$) and y_k is the value of the kth property in 2017 in COP. x_k, the built-up area of the

k^{th} property in square meters, is known for every $k \in U$. The auxiliary variable x has mean 184, standard deviation 110 and skewness 2.57. The desired sample size is $n = 1,000$.

We assume that a model of the type ξ_0 with $f(x_k | \delta_1) = \delta_{10} + \delta_{11} x_k^{\delta_{12}}$ and $g(x_k | \delta_2) = x_k^{\delta_2}$ adequately describes the association between x and y. We plan to use the GREG estimator for estimating δ_{10} and δ_{11}, i.e., $\delta_1^{***} = (\delta_{10}, \delta_{11})$. As this model has the form shown in Example 4, the model expected MSE can be approximated by expression (5.7).

We will use the risk (4.1) in order to assist the decision between πps or STSI using $H = 6$ strata. We take $h(\beta_{12}, \beta_2)$ as a bivariate normal distribution with no correlation between β_{12} and β_2. The integral is approximated using package `cubature` (Narasimhan, Johnson, Hahn, Bouvier and Kiêu, 2019) developed for the statistical software environment `R` (R Core Team, 2020).

We consider two cases with different degrees of confidence regarding the working model.

Case 1. In this case no information about δ_{12}, δ_2 or $R_{x,y}$ is available. Naive values of $\delta_{12} = 1$, $\delta_2 = 1$ and $R_{x,y} = 0.75$ are considered. In order to reflect the uncertainty, $h(\beta)$ should have a large variance, therefore we set

$$\begin{bmatrix} \beta_{12} \\ \beta_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 1.0 \\ 1.0 \end{bmatrix}, \begin{bmatrix} 0.3295^2 & 0 \\ 0 & 0.3295^2 \end{bmatrix} \right).$$

The variance was chosen in such a way that 99% of the mass lies in the circle of radius 1. Evaluation of (4.1) yields $R(\pi \text{ps}) = 6.89 \cdot 10^{15} \beta_{11}^2$ and $R(\text{STSI}) = 1.59 \cdot 10^{15} \beta_{11}^2$, suggesting that a stratified design should be used.

The design MSE of both strategies is computed and we get, $\text{MSE}_{\pi \text{ps}}(\hat{t}_{\text{greg}}) = 2.29 \cdot 10^{25}$ and $\text{MSE}_{\text{STSI}}(\hat{t}_{\text{greg}}) = 1.36 \cdot 10^{25}$. The strategy suggested by (4.1) was indeed the best choice.

Case 2. Using a sample from 2010, prior values of $\delta_{12} = 1.9$, $\delta_2 = 2$ and $R_{x,y} = 0.7$ are proposed. As the uncertainty here is smaller than that in Case 1, we set a smaller variance,

$$\begin{bmatrix} \beta_{12} \\ \beta_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 1.9 \\ 2.0 \end{bmatrix}, \begin{bmatrix} 0.2471^2 & 0 \\ 0 & 0.2471^2 \end{bmatrix} \right).$$

The variance was chosen in such a way that 99% of the mass lies in the circle of radius 0.75. Evaluation of (4.1) yields $R(\pi \text{ps}) = 7.08 \cdot 10^{22} \beta_{11}^2$ and $R(\text{STSI}) = 4.06 \cdot 10^{18} \beta_{11}^2$, suggesting that a stratified design should be used.

The design MSE of both strategies is computed and we get $\text{MSE}_{\pi \text{ps}}(\hat{t}_{\text{greg}}) = 1.85 \cdot 10^{28}$ and $\text{MSE}_{\text{STSI}}(\hat{t}_{\text{greg}}) = 1.91 \cdot 10^{25}$. Note that the use of (4.1) prevented us from using πps, whose MSE is almost one thousand times bigger than the one under stratified sampling!

7. Conclusions

The strategy that couples πps with the difference estimator is optimal when the parameters of the superpopulation model are known. Taking into account that these assumptions are seldom satisfied, it was shown in Section 3 and illustrated in Subsection 6.1 that this optimality breaks down even under small misspecifications of the model.

In Section 4 we propose a method for choosing the sampling design, which is extended to its use with the GREG estimator in Section 5. The method allows for taking the uncertainty about the model parameters into account by introducing a prior distribution on them. Although it could be argued that a source of subjectivity is added by introducing a prior distribution on the parameters, our view is that it is more subjective to choose the design without any type of assessment of the assumptions. Furthermore, inference is still design-based, as the prior is used only for choosing the design.

The method was illustrated with a real dataset, yielding satisfactory results. It should be noted that although the illustrations used stratified simple random sampling, the method in this article is valid for any sampling design.

Appendix

Proof of (4.2)

Proof. The following expectations are required in the proof,

$$\mathrm{E}_{\xi} Y_k \;=\; \mathrm{E}_{\xi}\left[f\left(x_k \,|\, \beta_1\right) + \varepsilon_k \right] = f\left(x_k \,|\, \beta_1\right) \tag{A.1}$$

$$\mathrm{E}_{\xi} Y_k^2 \;=\; \mathrm{E}_{\xi}\left[\left(f\left(x_k \,|\, \beta_1\right) + \varepsilon_k \right)^2 \right] = f\left(x_k \,|\, \beta_1\right)^2 + \sigma^2 g\left(x_k \,|\, \beta_2\right)^2 \tag{A.2}$$

$\mathrm{E}_{\xi} \bar{Y}$, $\mathrm{E}_{\xi} \bar{Y}^2$ and $\mathrm{E}_{\xi} \overline{fY}$ are obtained using (A.1) and (A.2),

$$\mathrm{E}_{\xi} \bar{Y} \;=\; \mathrm{E}_{\xi}\left[\frac{1}{N} \sum_U Y_k \right] = \frac{1}{N} \sum_U \mathrm{E}_{\xi} Y_k = \frac{1}{N} \sum_U f\left(x_k \,|\, \beta_1\right) \equiv \bar{f} \tag{A.3}$$

$$\mathrm{E}_{\xi} \bar{Y}^2 \;=\; \mathrm{E}_{\xi}\left[\frac{1}{N} \sum_U Y_k^2 \right] = \frac{1}{N} \sum_U \left(f\left(x_k \,|\, \beta_1\right)^2 + \sigma^2 g\left(x_k \,|\, \beta_2\right)^2 \right) \equiv \overline{f^2} + \sigma^2 \overline{g^2} \tag{A.4}$$

$$\mathrm{E}_{\xi} \overline{fY} \;=\; \mathrm{E}_{\xi}\left[\frac{1}{N} \sum_U f\left(x_k \,|\, \beta\right) Y_k \right] = \frac{1}{N} \sum_U f\left(x_k \,|\, \beta\right) \mathrm{E}_{\xi} Y_k = \frac{1}{N} \sum_U f\left(x_k \,|\, \beta\right)^2 = \overline{f^2}. \tag{A.5}$$

Now, using (A.3), (A.4) and (A.5) we get

$$\mathrm{E}_{\xi}\left[\overline{fY} - \bar{f}\bar{Y} \right] \;=\; \overline{f^2} - \bar{f}^2 = S_{f,f} \tag{A.6}$$

$$\mathrm{E}_{\xi}\left[\bar{Y}^2 - \bar{Y}^2 \right] = \overline{f^2} + \sigma^2 \overline{g^2} - \bar{f}^2 = S_{f,f} + \sigma^2 \overline{g^2}. \tag{A.7}$$

Using (A.6) and (A.7), we obtain an approximation to the correlation coefficient, $R_{f,y}$,

$$R_{f,y}^2 \;=\; \frac{\left(\overline{fy} - \overline{f}\,\overline{y}\right)^2}{\left(\overline{f^2} - \overline{f}^2\right)\left(\overline{y^2} - \overline{y}^2\right)} \;\approx\; \frac{\mathrm{E}_\xi^2\left[\overline{fY} - \overline{f}\,\overline{Y}\right]}{\mathrm{E}_\xi\left[\left(\overline{f^2} - \overline{f}^2\right)\left(\overline{Y^2} - \overline{Y}^2\right)\right]} \;=\; \frac{S_{f,f}}{S_{f,f} + \sigma^2 \overline{g}^2}. \qquad (A.8)$$

Solving (A.8) for σ^2 we get (4.2), as desired. The proof of (5.6) is analogous.

References

Beaumont, J.-F., Haziza, D. and Ruiz-Gazen, A. (2013). A unified approach to robust estimation in finite population sampling. *Biometrika*, 100, 3, 555-569.

Bramati, M. (2012). Robust Lavallée-Hidiroglou stratified sampling strategy. *Survey Research Methods*, 6, 3, 137-143.

Cassel, C.M., Särndal, C.-E. and Wretman, J. (1976). Some results on generalized difference estimation and generalized regression estimation for finite populations. *Biometrika*, 63, 3, 615-620.

Cassel, C.M., Särndal, C.-E. and Wretman, J. (1977). *Foundations of Inference in Survey Sampling*. New York: John Wiley & Sons, Inc.

Dalenius, T., and Hodges, J.L. (1959). Minimum variance stratification. *Journal of the American Statistical Association*, 54, 88-101.

Godambe, V.P. (1955). A unified theory of sampling from finite populations. *Journal of the Royal Statistical Society, Series B*, 17, 269-278.

Hájek, J. (1959). Optimal strategy and other problems in probability sampling. *Casopis Pro Pestování Matematiky*, 84, 4, 387-423.

Holmberg, A., and Swensson, B. (2001). On pareto πps sampling: Reflections on unequal probability sampling strategies. *Theory of Stochastic Processes*, 7(23), 142-155.

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 superpopulation model. *Journal of the American Statistical Association*, 77, 89-96.

Lanke, J. (1973). On UMV-estimators in survey sampling. *Metrika*, 20, 196-202.

Narasimhan, B., Johnson, S., Hahn, T., Bouvier, A. and Kiêu, K. (2019). *Cubature: Adaptive Multivariate Integration Over Hypercubes.* R package version 2.0.4. https://CRAN.R-project.org/package=cubature.

Nedyalkova, D., and Tillé, Y. (2008). Optimal sampling and estimation strategies under the linear model. *Biometrika*, 95, 3, 521-537.

R Core Team (2020). R: A language and environment for statistical computing. *The R Foundation for Statistical Computing*, Vienna, Austria. URL https://www.R-project.org/.

Rosén, B. (1997). On sampling with probability proportional to size. *Journal of Statistical Planning and Inference*, 62, 159-191.

Rosén, B. (2000). *Generalized Regression Estimation and Pareto πps.* R&D Report 2000:5. Statistics Sweden.

Royall, R.M., and Herson, J. (1973). Robust estimation in finite populations I. *Journal of the American Statistical Association*, 68, 344, 880-889.

Särndal, C.-E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer.

Tillé, Y., and Wilhelm, M. (2017). Probability sampling designs: Principles for choice of design and balancing. *Statistical Science*, 32(2), 176-189.

Wright, R.L. (1983). Finite population sampling with multivariate auxiliary information. *Journal of the American Statistical Association*, 78, 879-884.

Zhai, Z., and Wiens, D. (2015). Robust model-based stratification sampling designs. *The Canadian Journal of Statistics*, 43, 4, 554-577.