Proceedings of the 14th International Conference on Computational and Mathematical Methods in Science and Engineering, CMMSE 2014 3–7July, 2014.

Multinomial logistic estimation in dual frame surveys

Maria del Mar Rueda[1], Antonio Arcos[1], David Molina[1] and Maria Giovanna Ranalli[2]

[1] *Department of Statistics and Operational Research, University of Granada*

[2] *Department of Political Sciences, University of Perugia, Italy*

emails: `mrueda@ugr.es`, `arcos@ugr.es`, `dmolinam@ugr.es`, `giovanna@stat.unipg.it`

Abstract

We will extend dual frame estimation techniques to the case of estimation of proportions when the variable of interest has multinomial outcomes. We describe the joint distribution of the class indicators by a multinomial logistic model. Logistic generalized regression estimators and model calibration estimators are introduced for class frequencies in a population by using two different approaches: "single frame" and "dual frame". Monte Carlo experiments were carried out to compare the efficiency of the proposed procedures in presence of different sets of auxiliary variables. The experiments indicate that the multinomial logistic formulation yields better results than the classical estimators for estimating proportions when sample data are obtained from more than one frame.

Key words: Finite population, Survey sampling, Auxiliary information, Model assisted inference
MSC 2000: 62D05

1 Introduction

Usually, sampling theory assumes the existence of only one sampling frame containing all population units. Then, a probability sample is drawn according to a sampling design and information collected is used for estimation and inference purposes. To ensure quality of the results obtained, the sampling frame must contain every single unit of population of interest (that is, it must be complete) and it must be updated as well. Otherwise, estimations could be affected by a serious bias due to the non-representativeness of samples selected. Unfortunately, this is not an easy task: populations are constantly changing, with new units entering and exiting the population every few time, so getting a good sampling frame can be difficult.

The dual frame approach tries to solve the aforementioned problems. This approach assumes that two frames are available for sampling and that, overall, they cover the entire target population. A sample is selected from each frame using a, possible different, sampling design for each frame. Much attention has been devoted to the introduction of different ways of combining estimates coming from the different frames. See the seminal papers by [5], [3] [1] [6]. However, these techniques were originally proposed to estimate means and totals of quantitative variables, and although their extension to the estimation of proportions in multinomial response variables is possible, it requires further investigation. Questionnaire items with multinomial outcomes are quite common in public opinion research, marketing research, and official surveys (estimating the proportion of voters in favour of each political party, based on a political opinion survey, is just one concrete example of this procedure). Items where respondents must select one in a series of options can be modeled by a multinomial distribution. [7] present estimators for a proportion which use the logistic regression estimator.

This paper focuses on the estimation of proportions in multinomial response variables when data come from two sampling frames. Different estimators for these proportions are proposed following different approaches and its main properties are studied. A simulation study is also presented.

2 Estimation of class frequencies in dual frame surveys

We will employ the notation considered in [9]. Let $\mathcal{U}$ denote a finite population with N units, $\mathcal{U} = \{1, \ldots, k, \ldots, N\}$ and let A and B be two sampling-frames. Let $\mathcal{A}$ be the set of population units in frame A and $\mathcal{B}$ the set of population units in frame B. The population of interest, $\mathcal{U}$, may be divided into three mutually exclusive domains, $a = \mathcal{A} \cap \mathcal{B}^c, b = \mathcal{A}^c \cap \mathcal{B}$ and $ab = \mathcal{A} \cap \mathcal{B}$. Because the population units in the overlap domain ab can be sampled in either survey or both surveys, it is convenient to create a duplicate domain $ba = \mathcal{B} \cap \mathcal{A}$, which is identical to $ab = \mathcal{A} \cap \mathcal{B}$, to denote the domain in the overlapping area, coming from frame B. Let N, N_A, N_B, N_a, N_b, N_{ab}, N_{ba} be the number of population units in $\mathcal{U}$, A, B, a, b, ab, ba, respectively.

In this work we consider the estimation of class frequencies of a discrete response variable. Assume that we collect data form respondents who provide a single choice from a list of alternatives. We code these alternatives $1, 2, \ldots, m$. Therefore, consider a discrete m-valued survey variable y. The objective is to estimate the frequency distribution of the y in the population U. To estimate this frequency distribution, we define a class of indicators z_i $(i = 1, \ldots, m)$ such that for each unit $k \in U$ $z_{ki} = 1$ if $y_k = i$ and $z_{ki} = 0$ otherwise. Our problem thus, is to estimate the proportions $P_i = \frac{1}{N} \sum_{k \in U} z_{ki}$ $i = 1, 2, \ldots, m$.

 ISBN: 978-84-616-9216-3

We can write
$$P_i = \frac{1}{N}(Z_{ai} + \eta Z_{abi} + (1-\eta)Z_{bai} + Z_{bi}), \tag{1}$$
where $0 \leq \eta \leq 1$ and $Z_{ai} = \sum_{k\in a} z_{ki}$, $Z_{abi} = \sum_{k\in ab} z_{ki}$, $Z_{ba} = \sum_{k\in ba} z_{ki}$ and $Z_b = \sum_{k\in b} z_{ki}$.

Two probability samples s_A and s_B are drawn independently from frame A and frame B of sizes n_A and n_B, respectively. Each design induces first-order inclusion probabilities π_{A_k} and π_{B_k}, respectively, and sampling weights $d_{A_k} = 1/\pi_{A_k}$ and $d_{B_k} = 1/\pi_{B_k}$. The sample s_A can be post-stratified as $s_A = s_a \cup s_{ab}$, where $s_a = s_A \cap a$ and $s_{ab} = s_A \cap (ab)$. Similarly, $s_B = s_b \cup s_{ba}$, where $s_b = s_B \cap b$ and $s_{ba} = s_B \cap (ba)$. Note that s_{ab} and s_{ba} are both from the same domain ab, but s_{ab} is part of the frame A sample and s_{ba} is part of the frame B sample. Then, let $s = s_A \cup s_B$.

The Hartley [5] estimator of P_i $i = 1, 2, \ldots, m$ is given by
$$\hat{P}_{Hi}(\eta) = \frac{1}{N}(\hat{Z}_{ai} + \eta\hat{Z}_{abi} + (1-\eta)\hat{Z}_{bai} + \hat{Z}_{ai}), \tag{2}$$
where $\hat{Z}_{ai} = \sum_{k\in s_a} d_{Ak}z_{ki}$ is the Horvitz-Thompson estimator for the proportion of domain a and similarly for the other domains. If we let
$$d_k^{\circ} = \begin{cases} d_{Ak} & \text{if } k \in s_a \\ \eta d_{Ak} & \text{if } k \in s_{ab} \\ (1-\eta)d_{Bk} & \text{if } k \in s_{ba} \\ d_{Bk} & \text{if } k \in s_b \end{cases}, \tag{3}$$
then $\hat{P}_{Hi}(\eta) = \frac{1}{N}(\sum_{k\in s} d_k^{\circ} z_{ki})$. Since each domain is estimated by its Horvitz-Thompson estimator, $\hat{P}_{Hi}(\eta)$ is an unbiased estimator of P_i for a given η.

The estimator developed by [3] incorporates information regarding the estimation of N_{ab} to improve over P_i, but has the drawback of not being a linear combination of z_i values, unless using simple random sampling. [12] propose a modification of the estimator proposed by [3] for simple random sampling to handle complex designs. They introduce a pseudo maximum likelihood (PML) estimator that does not achieve optimality like the FB estimator, but it can be written as a linear combination of the observations and the same set of weights can be used for all variables of interest. Recently, [9] extended the Pseudo-Empirical-Likelihood approach (PEL) proposed by [13] from one-frame surveys to dual-frame surveys following a stratification approach.

3 Estimation of class frequencies using multinomial logistic regression

Auxiliary information is often available in survey sampling. This information, which may come from past censuses or from other administrative sources, can be used to obtain more

 ISBN: 978-84-616-9216-3

accurate estimators. Then, other than the values of y for $k \in s$, suppose we also know the value of the vector of auxiliary variables $\underline{\mathrm{x}}_k$ for $k \in U$. We consider that the population under study $\mathbf{y} = (y_1, ..., y_N)'$ is the determination of a set of super-population random variables $\mathbf{Y} = (Y_1, ..., Y_N)'$ s.t.

$$\mu_{ki} = P(Y_k = i|\underline{\mathrm{x}}_k) = E(Z_{ki}|\underline{\mathrm{x}}_k) = \frac{\exp(\underline{\mathrm{x}}_k^T \beta_i)}{\sum_{r=1,...,m} \exp(\underline{\mathrm{x}}_{kr}^T \beta_r)} + e_{ki}, \quad i = 1, .., m,$$

that is, we use the multinomial logistic model to relate the variables y and $\underline{\mathrm{x}}$.

We denote by $\boldsymbol{\beta}$ the parameter vector $(\beta_1^T, ..., \beta_m^T)$. Now, we will define new estimators for the population proportions of z_i variables. For that, we consider the estimation of the superpopulation parameter $\boldsymbol{\beta}$ by the units of the sample s.

3.1 Approach 1: Single frame

When inclusion probabilities in domain ab are known for both frames, and not just for the frame from which the unit was selected, *single-frame* methods ([1], [6]), which combine the observations into a single dataset and adjust the weights in the intersection domain for multiplicity, can be used. To adjust for multiplicity, the weights are defined as follows for all units in frame A and in frame B,

$$\tilde{d}_k = \begin{cases} d_{Ak} & \text{if } k \in s_a \\ (1/d_{Ak} + 1/d_{Bk})^{-1} & \text{if } k \in s_{ab} \cup s_{ba} \\ d_{Bk} & \text{if } k \in s_b \end{cases} .$$

We estimate $\boldsymbol{\beta}$ by maximizing the π-weighted likelihood ([4], [10]) given by

$$L(\boldsymbol{\beta}) = \sum_{i=1,...,m} \sum_{k \in s} \tilde{d}_k \ln \mu_{ki}.$$

This usually requires numerical procedures, and Fisher scoring or Newton-Raphson often work rather well. Most statistical packages include a multinomial logit procedure.

Given the estimate $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, we consider the following auxiliary variable

$$p_{ki} = \widehat{\mu}_{ki} = \frac{\exp(\underline{\mathrm{x}}_k^T \widehat{\beta}_i)}{\sum_{r=1,...,m} \exp(\underline{\mathrm{x}}_{kr}^T \widehat{\beta}_r)}. \tag{4}$$

Since the vector $\underline{\mathrm{x}}_k$ is known for all units of the population U, the values p_{ki} are available $\forall k \in U$ and we propose to use the values p_{ki} to obtain a new estimator for P_i,

$$\widehat{P}_{MLRSi} = \frac{1}{N} \left(\sum_{k \in U} p_{ki} + \sum_{k \in s} \tilde{d}_k (z_{ki} - p_{ki}) \right). \tag{5}$$

 ISBN: 978-84-616-9216-3

We observe that this estimator takes the same model-assisted form as the MLGREG estimator proposed in [7], but here it is adjusted to account for the dual frame sampling setting.

Another important way to incorporate available auxiliary information is given by calibration estimation ([2]), that seeks for new weights that are close (in some sense) to the basic design weights and that, at the same time, match benchmark constraints on auxiliary information. See [8] for the extension of calibration to the dual frame setting. Here, we propose a new calibration estimator

$$\hat{P}_{MLcalSFi} = \frac{1}{N}\sum_{k\in s}\tilde{w}_k z_{ki},$$

where $\tilde{w}_k$ minimizes $\sum_{k\in s} G(\tilde{w}_k, \tilde{d}_k)$ subject to:

$$\sum_{k\in s}\tilde{w}_k \underline{\mathrm{r}}_{ki} = \sum_{k\in U}\underline{\mathrm{r}}_{ki}$$

where the elements of $\underline{\mathrm{r}}_{ki}$ change according to the available auxiliary information. In particular,

- if N_A, N_B N_{ab} are known:

$$\underline{\mathrm{r}}_{ki} = (\delta_k(a), \delta_k(ab) + \delta_k(ba), \delta_k(b), p_{ki})$$

- and if N_A, N_B are known:

$$\underline{\mathrm{r}}_{ki} = (\delta_k(a) + \delta_k(ab) + \delta_k(ba), \delta_k(b) + \delta_k(ba) + \delta_k(ab), p_{ki})$$

with $\delta_k(a), \delta_k(ab), \delta_k(ba)$ and $\delta_k(b)$ the indicator variables for domains a, ab, ba and b, respectively. This is an extension of the Model calibration approach proposed by [14].

3.2 Approach 2: Dual frame

We estimate the probabilities μ_{ki} separately in each frame. For each $k \in \mathcal{A}$, using data of sample s_A one can estimate μ_{ki} by

$$p^A_{ki} = \frac{\exp(\underline{\mathrm{x}}_k^T\widehat{\beta}_i^A)}{\sum_{r=1,\ldots,m}\exp(\underline{\mathrm{x}}_{kr}^T\widehat{\beta}_r^A)} \tag{6}$$

where we estimate $\boldsymbol{\beta}^A$ by maximizing $L(\boldsymbol{\beta}^A) = \sum_{i=1,\ldots,m}\sum_{k\in s_A} d_{Ak}\ln\mu_{ki}$.

Similarly we obtain p^B_{ki} for $k \in \mathcal{B}$, and define for each $i = 1, \ldots, m$ the following regression estimator:

 ISBN: 978-84-616-9216-3

$$\widehat{P}_{MLRDi} = \frac{1}{N}\left(\sum_{a} p_{ki}^{A} + \eta \sum_{ab} p_{ki}^{A} + \sum_{b} p_{ki}^{B} + (1-\eta)\sum_{ba} p_{ki}^{B} + \right.$$
$$\left. + \sum_{s_A}(z_{ki} - p_{ki}^{A})d_{Ak} + \sum_{s_B}(z_{ki} - p_{ki}^{B})d_{Bk}\right).$$

Several calibration estimators can be defined using dual frame approach. Although all of them are in the form

$$\hat{P}_{MLcalDFi} = \frac{1}{N}\sum_{k\in s} w_k^{\star} z_{ki}, \tag{7}$$

different sets of weights can be obtained considering different distance functions and different calibration constraints. In particular, let say $\hat{P}_{MLCalDF1i}$ use weights $w_k^{1\star}$ such that (we show only the case N_{ab} unknown for space reason)

$$\min \sum_{k\in s} G(w_k^{1\star}, d_k) \qquad s.t.$$

$$\sum_{k\in s_a} w_k^{1\star}\delta_k(a) + \sum_{k\in s_{ab}} w_k^{1\star}\delta_k(ab) + \sum_{k\in s_{ba}} w_k^{1\star}\delta_k(ba) = N_A,$$
$$\sum_{k\in s_b} w_k^{1\star}\delta_k(b) + \sum_{k\in s_{ba}} w_k^{1\star}\delta_k(ba) + \sum_{k\in s_{ab}} w_k^{1\star}\delta_k(ab) = N_B,$$

and

$$\sum_{k\in s_A} w_k^{1\star} p_{ki}^{A} + \sum_{k\in s_B} w_k^{1\star} p_{ki}^{B} = \sum_{k\in U_a} p_{ki}^{A} + \eta \sum_{k\in U_{ab}} p_{ki}^{A} + (1-\eta)\sum_{k\in U_{ba}} p_{ki}^{B} + \sum_{k\in U_b} p_{ki}^{B}$$

where p_{ki}^{A} are the estimated probabilities defined in (6) and p_{ki}^{B} is its analogous in frame B.

As an alternative, the last single constraint can be replaced by other two, each of them referring to a frame, as follows

$$\sum_{k\in s_A} w_k^{2\star} p_{ki}^{A} = \sum_{k\in U_a} p_{ki}^{A} + \eta \sum_{k\in U_{ab}} p_{ki}^{A}$$
$$\sum_{k\in s_B} w_k^{2\star} p_{ki}^{B} = (1-\eta)\sum_{k\in U_{ba}} p_{ki}^{B} + \sum_{k\in U_b} p_{ki}^{B}$$

From the resulting weights we can calculate a new estimator, $\hat{P}_{MLCalDF2i}$.

Alternatively, another estimator, say $\hat{P}_{MLCalDF3i}$, can be obtained following a methodology quite similar to the one described in section 3.1. In this sense, estimator is calculated from a set of weights $w_k^{3\star}$ verifying that

 ISBN: 978-84-616-9216-3

$$\min \sum_{k \in s} G(w_k^{3\star}, d_k^{\circ}) \qquad \text{s.t.} \tag{8}$$

$$\sum_{k \in s_a} w_k^{3\star} \delta_k(a) + \sum_{k \in s_{ab}} w_k^{3\star} \delta_k(ab) + \sum_{k \in s_{ba}} w_k^{3\star} \delta_k(ba) = N_A,$$

$$\sum_{k \in s_b} w_k^{3\star} \delta_k(b) + \sum_{k \in s_{ba}} w_k^{3\star} \delta_k(ba) + \sum_{k \in s_{ab}} w_k^{3\star} \delta_k(ab) = N_B,$$

and

$$\sum_{k \in s} w_k^{3\star} p_{ki}^{\star} = \sum_{k \in U} p_{ki}^{\star},$$

where probabilities $p_{ki}^{\star}$, $k \in U$, $i = 1, \ldots, m$ are estimated from the whole sample s using an estimate $\hat{\boldsymbol{\beta}}^{\star}$ of $\boldsymbol{\beta}$ obtained by maximizing $L(\boldsymbol{\beta}) = \sum_{i=1,\ldots,m} \sum_{k \in s} d_k^{\circ} \ln \mu_{ki}$, where d_k° are defined in (3).

4 Monte Carlo simulation experiments

For our simulation study we use the `hsbdemo` data set (`http://www.ats.ucla.edu/stat/data/hsbdemo.dta`). The data set contains variables on 200 students. The outcome variable is `prog`, program type, a three-level categorical variable whose categories are `academic`, `general`, `vocation`. The predictor variables are social economic status, `ses`, a three-level categorical variable and a mathematical score, `math`, a continuous variable. We estimate a multinomial logistic regression model. We create a new data set with 50 copies of predictor variables `ses` and `math` and with the predicted values for the variable `prog`. The simulated populations, namely POP1, have, therefore, dimension $N = 10000$.

Units are randomly assigned to the two frames, A and B, according to three different scenarios depending on the overlap domain size N_{ab}. We first generate copies the sequence "a", "b", or "ab" to have the required domain sizes in the population and generate N normal random numbers, $\varepsilon_k, k = 1, \ldots, N$. Then, we sort the data by ε. The first scenario has a *small* overlap domain size N_{ab}=1000 and the resulting sizes of the two frames are N_A=6000 and N_B=5000.The second and the third scenarios have respectively *large* and *medium* overlap domain size. The resulting frame sizes in the second scenario are given by N_A=6000 and N_B=7000 and the overlap domain size is N_{ab}=3000, while for the third scenario we have N_A=8000, N_B=7000 and N_{ab}=5000.

Similarly, POP2 is built first by assigning units to the frames and second by fitting a multinomial logistic regression model separately in each frame (with the same predictor variables).

Samples from frame A are selected by means of Midzuno sampling, with inclusion probabilities proportional to variable `cid`. Samples from frame B are selected by means

 ISBN: 978-84-616-9216-3

of simple random sampling. For each scenario, we draw a combination of sample sizes for frame A and frame B, as follows: ($n_A = 180$, $n_B = 232$).

This makes a 3×2 design for the simulation study. For each of the 6 settings, we compute the multinomial logistic regression estimator under single frame (PMLRS) and dual frame (PMLRD) approach, the multinomial logistic calibration estimators under single frame (PMLCalSF) and dual frame (PMLCalDF1, PMLCalDF2, PMLCalDF3) approach.

We compute also the Hartley estimator [5], the Pseudo Maximum Likelihood estimator (PML) when N_{ab} is unknown [11], the single frame estimator (BKA) [1] and [6], the Fuller-Bunmeister estimator [3] and the Raking Ratio estimator (SFRR) [11] for the purpose of comparison. The Pseudo Empirical Likelihhod estimator (PEL) [9] and the dual frame and the single frame calibration estimator (CalDF and CalSF) [8] are also computed using the auxiliary information on `ses` and `math`. When needed (and for comparative purposes) the value of η has been estimated using $\eta = v(\hat{N}_{ba})/(v(\hat{N}_{ab}) + v(\hat{N}_{ba}))$ for all compared estimators, where $v(\hat{N}_{ab})$ is an estimate of the variance of the Horvitz-Thompson estimator $\hat{N}_{ab}$ for the size of overlap domain, and similarly for $v(\hat{N}_{ba})$.

For each estimator, we compute the percent relative bias $RB\% = E_{MC}(\hat{Y} - Y)/Y * 100$, the percent relative mean squared error $RMSE\% = E_{MC}[(\hat{Y} - Y)^2]/Y^2 * 100$ for each category of the main variable `prog` and the minimum, maximum and mean percent over categories, based on 1000 simulation runs.

Tables 1 to 2 report results. From these tables we can see that relative biases are negligible in all cases, as a consequence, efficiency comparisons can be based on variances. The performance in terms of efficiency of the estimators is essentially driven by the set of auxiliary variables employed. When no auxiliary information about `ses` and `math` is used, the efficiency is small (SFRR, Hartley, FB, PML). When `ses` and `math` are employed in calibration process (CalSF, PEL, CalDF), the efficiency increases and where `ses` and `math` are also used trough a model, is the most effective as expected (PMLRS, PMLCalSF, PMLRD, PMLCalDF1, PMLCalDF2, PMLCalDF3). There is not a relevant difference in efficiency between single frame and dual frame approach, irrespective to the use of a multinomial logistic estimator or a multinomial calibration estimator. With regard to the relative efficiency, comparisons do not allow any proposed estimators to emerge among others and do suggest that all the estimators considered tend to perform well, and better than using a simple linear regression model (compare with CalSF and CalDF). Furthermore the proposed estimators have the additional advantage that the estimates of proportions for each category add to 1.

Acknowledgements

This study was partially supported by Ministerio de Educación y Ciencia (grant MTM2012-35650, Spain), by Consejería de Economía, Innovación, Ciencia y Empleo (grant SEJ2954,

 ISBN: 978-84-616-9216-3

Junta de Andalucía, Spain), and under the support of the project PRIN-SURWEY (grant 2012F42NS8, Italy).

References

[1] M. D. Bankier, *Estimators based on several stratified samples with applications to multiple frame surveys*, Journal of the American Statistical Association (1986) 1074–1079.

[2] J. C. Deville and C. E. Särndal, . *Calibration estimators in survey sampling*, Journal of the American Statistical Association (1992) 376–382.

[3] W. A. Fuller an L. F. Burmeister, *Estimation for samples selected from two overlapping frames*, Proceedings of the American Statistical Association, Social Statistics Section (1972) 245–249.

[4] V. P. Godambe and M. E. Thompson, *Parameters of superpopulation and survey population: their relationships and estimation*, International Statistical Review (1986) 127–138.

[5] H. O. Hartley, *Multiple frame surveys*, Proceedings of the American Statistical Association, Social Statistics Section (1962) 203–206.

[6] G. Kalton and D. W. Anderson, *Sampling rare populations*, Journal of the Royal Statistical Society. Series A (1986) 65–82.

[7] R. Lehtonen and A. Veijanen, *On multinomial logistic generalized regression estimators*, Preprint from the Department of Statistics, University of Jyväskylä **22** (1998).

[8] M. G. Ranalli, A. Arcos, M. Rueda, M. and A. Teodoro, *Calibration estimators in dual frames surveys* arXiv:1312.0761 [stat.ME] (2013)

[9] J. N. K. Rao and C. Wu, *Pseudo empirical likelihood inference for multiple frame surveys*, Journal of the American Statistical Association **105** (2010) 1494–1503.

[10] C. E. Särndal, B. Swensson and J. Wretman, *Model-assisted survey sampling*, Springer-Verlag, New York, 1992.

[11] C. J. Skinner, *On the efficiency of raking ratio estimation for multiple frame surveys*, Journal of the American Statistical Association (1991) 779–784.

[12] C. J. Skinner and J. N. K. Rao, *Estimation in dual frame surveys with complex designs*, Journal of the American Statistical Association **91(443)** (1996) 349–356.

[13] C. Wu and J. N. K. Rao, *Pseudo empirical likelihood ratio confidence intervals for complex surveys*, The Canadian Journal of Statistics **34** (2006) 359–375.

[14] C. Wu and R. R. Sitter, *A model-calibration approach to using complete auxiliary information from survey data*, Journal of the American Statistical Association **96** (2001) 185–193.

 ISBN: 978-84-616-9216-3

Table 1: Relative efficiency (respecto to the BKA estimator) of compared estimator.

	POP1						POP2					
	acad.	gen.	voc.	min	max	mean	acad.	gen.	voc.	min	max	mean
						Medium						
PMLRS	348.12	181.25	252.44	181.25	348.12	260.60	285.76	154.12	205.13	154.12	285.76	215.00
PMLCalSF	358.10	180.97	258.85	180.97	358.10	265.97	316.20	154.22	225.59	154.22	316.20	232.00
PMLRD	350.18	187.65	257.22	187.65	350.18	265.01	290.04	206.85	208.62	206.85	290.04	235.17
PMLCalDF1	358.28	185.99	262.45	185.99	358.28	268.90	320.63	208.45	228.48	208.45	320.63	252.52
PMLCalDF2	358.93	186.31	263.52	186.31	358.93	269.58	322.70	206.94	228.85	206.94	322.70	252.83
PMLCalDF3	356.87	181.05	258.60	181.05	356.87	265.50	315.27	154.64	224.88	154.64	315.27	231.59
SFRR	99.63	100.26	99.59	99.59	100.26	99.82	100.57	99.71	101.41	99.71	101.41	100.56
CalSF	149.94	142.21	132.30	132.30	149.94	141.48	179.37	134.31	147.57	134.31	179.37	153.75
Hartley	99.24	97.79	98.61	97.79	99.24	98.54	99.60	97.35	99.81	97.35	99.81	98.92
FB	97.37	97.77	97.89	97.37	97.89	97.67	98.52	97.03	99.70	97.03	99.70	98.41
PML	99.50	99.97	99.60	99.50	99.97	99.69	100.57	99.39	101.49	99.39	101.49	100.48
PEL	217.89	135.87	177.26	135.87	217.89	177.00	233.98	137.14	197.58	137.14	233.98	189.56
CalDF	213.91	134.83	175.14	134.83	213.91	174.62	240.90	138.51	205.54	138.51	240.90	194.98
						Small						
PMLRS	331.75	163.33	248.08	163.33	331.75	247.72	252.11	157.96	228.84	157.96	252.11	212.97
PMLCalSF	353.77	163.17	265.85	163.17	353.77	260.93	268.97	158.53	248.22	158.53	268.97	225.24
PMLRD	343.94	164.70	257.75	164.70	343.94	255.46	309.54	229.33	225.33	225.33	309.54	254.73
PMLCalDF1	365.90	165.05	274.66	165.05	365.90	268.53	339.40	229.52	244.28	229.52	339.40	271.06
PMLCalDF2	365.15	163.94	275.28	163.94	365.15	268.12	345.97	230.52	249.17	230.52	345.97	275.22
PMLCalDF3	353.76	163.06	265.66	163.06	353.76	260.82	268.72	158.59	248.06	158.59	268.72	225.12
SFRR	99.96	99.90	99.84	99.84	99.96	99.90	100.52	100.01	100.20	100.01	100.52	100.24
CalSF	155.30	137.56	140.60	137.56	155.30	144.48	161.56	148.10	141.66	141.66	161.56	150.44
Hartley	99.76	97.59	98.98	97.59	99.76	98.77	98.48	98.30	99.15	98.30	99.15	98.64
FB	98.10	97.59	98.51	97.59	98.51	98.06	98.62	97.82	98.97	97.82	98.97	98.47
PML	99.81	99.89	99.60	99.60	99.89	99.76	100.50	99.86	100.23	99.86	100.50	100.19
PEL	232.55	147.36	198.25	147.36	232.55	192.72	224.18	165.83	177.18	165.83	224.18	189.06
CalDF	210.50	134.54	179.08	134.54	210.50	174.70	222.13	164.14	174.61	164.14	222.13	186.96
						Large						
PMLRS	356.73	161.87	257.40	161.87	356.73	258.66	345.31	130.70	263.30	130.70	345.31	246.43
PMLCalSF	375.21	161.38	267.54	161.38	375.21	268.04	384.82	133.27	282.32	133.27	384.82	266.80
PMLRD	362.07	168.39	265.88	168.39	362.07	265.44	318.17	146.83	257.61	146.83	318.17	240.87
PMLCalDF1	381.24	174.49	276.55	174.49	381.24	277.42	307.90	114.90	275.49	114.90	307.90	232.76
PMLCalDF2	376.11	167.22	274.78	167.22	376.11	272.70	353.73	145.56	280.22	145.56	353.73	259.83
PMLCalDF3	371.74	161.23	266.64	161.23	371.74	266.53	379.61	132.47	282.95	132.47	379.61	265.01
SFRR	100.20	99.50	100.31	99.50	100.31	100.00	103.12	103.35	100.42	100.42	103.35	102.29
CalSF	147.60	130.53	138.13	130.53	147.60	138.75	160.52	121.03	129.70	121.03	160.52	137.08
Hartley	98.16	96.01	97.42	96.01	98.16	97.19	94.01	96.85	99.28	94.01	99.28	96.71
FB	99.29	96.17	99.18	96.17	99.29	98.21	101.01	98.18	98.99	98.18	101.01	99.39
PML	99.95	99.11	100.19	99.11	100.19	99.75	102.47	103.34	99.75	99.75	103.34	101.85
PEL	193.48	124.99	173.21	124.99	193.48	163.89	194.09	147.68	169.70	147.68	194.09	170.49
CalDF	192.10	125.72	170.56	125.72	192.10	162.79	189.76	163.55	181.28	163.55	189.76	178.19

 ISBN: 978-84-616-9216-3

Table 2: Relative bias of compared estimator.

	POP1						POP2					
	acad.	gen.	voc.	min	max	mean	acad.	gen.	voc.	min	max	mean
						Medium						
BKA	0.06	-0.13	0.31	0.06	0.31	0.16	-0.84	0.06	0.12	0.06	0.84	0.34
PMLRS	-0.03	0.48	-0.07	0.03	0.48	0.19	0.04	-0.39	0.03	0.03	0.39	0.15
PMLCalSF	-0.07	0.46	0.05	0.05	0.46	0.19	0.02	-0.53	0.13	0.02	0.53	0.23
PMLRD	-0.04	1.05	-0.23	0.04	1.05	0.44	-0.01	0.02	0.02	0.01	0.02	0.02
PMLCalDF1	-0.07	1.09	-0.17	0.07	1.09	0.44	0.01	-0.14	0.03	0.01	0.14	0.06
PMLCalDF2	-0.12	1.06	-0.02	0.02	1.06	0.40	-0.04	-0.19	0.16	0.04	0.19	0.13
PMLCalDF3	-0.07	0.44	0.06	0.06	0.44	0.19	0.02	-0.53	0.13	0.02	0.53	0.23
SFRR	0.04	-0.13	0.32	0.04	0.32	0.16	-0.83	0.04	0.17	0.04	0.83	0.35
CalSF	0.02	-0.02	0.05	0.02	0.05	0.03	-0.79	0.21	-0.26	0.21	0.79	0.42
Hartley	-1.54	-0.14	0.01	0.01	1.54	0.56	-2.37	0.03	-0.20	0.03	2.37	0.87
FB	-1.47	0.03	0.14	0.03	1.47	0.54	-2.25	0.18	-0.01	0.01	2.25	0.82
PML	0.15	-0.13	0.36	0.13	0.36	0.21	-0.80	0.07	0.15	0.07	0.80	0.34
PEL	-0.01	-0.01	0.03	0.01	0.03	0.02	-0.84	0.03	0.18	0.03	0.84	0.35
CalDF	0.10	0.06	-0.18	0.06	0.18	0.11	-0.69	0.11	-0.05	0.05	0.69	0.28
						Small						
BKA	-0.33	-0.06	0.26	0.06	0.33	0.21	-0.02	-0.01	0.04	0.01	0.04	0.02
PMLRS	-0.06	0.83	-0.11	0.06	0.83	0.33	0.09	0.27	-0.38	0.09	0.38	0.24
PMLCalSF	-0.10	0.84	-0.01	0.01	0.84	0.32	0.07	0.20	-0.29	0.07	0.29	0.18
PMLRD	-0.07	1.35	-0.24	0.07	1.35	0.55	0.00	1.24	-0.50	0.00	1.24	0.58
PMLCalDF1	-0.12	1.45	-0.15	0.12	1.45	0.57	-0.08	1.57	-0.38	0.08	1.57	0.68
PMLCalDF2	-0.15	1.41	-0.05	0.05	1.41	0.54	-0.05	1.37	-0.39	0.05	1.37	0.61
PMLCalDF3	-0.10	0.86	-0.01	0.01	0.86	0.32	0.06	0.21	-0.29	0.06	0.29	0.19
SFRR	-0.34	-0.06	0.26	0.06	0.34	0.22	-0.04	-0.01	0.04	0.01	0.04	0.03
CalSF	-0.05	0.04	-0.08	0.04	0.08	0.06	-0.30	0.13	-0.28	0.13	0.30	0.23
Hartley	-1.84	-0.12	-0.07	0.07	1.84	0.68	-1.15	-0.05	-0.38	0.05	1.15	0.53
FB	-1.68	0.12	0.10	0.10	1.68	0.63	-0.95	0.17	-0.19	0.17	0.95	0.44
PML	-0.24	-0.06	0.30	0.06	0.30	0.20	-0.02	0.01	0.06	0.01	0.06	0.03
PEL	0.05	-0.03	0.05	0.03	0.05	0.04	-0.25	0.07	-0.15	0.07	0.25	0.16
CalDF	0.22	0.04	-0.17	0.04	0.22	0.14	-0.01	0.16	-0.48	0.01	0.48	0.22
						Large						
BKA	-0.29	0.13	-0.24	0.13	0.29	0.22	0.04	-0.10	0.26	0.04	0.26	0.13
PMLRS	0.04	0.34	-0.22	0.04	0.34	0.20	-0.12	0.85	0.10	0.10	0.85	0.36
PMLCalSF	-0.02	0.29	-0.03	0.02	0.29	0.11	-0.17	0.68	0.27	0.17	0.68	0.37
PMLRD	0.02	0.66	-0.27	0.02	0.66	0.32	-0.11	1.60	-0.16	0.11	1.60	0.62
PMLCalDF1	-0.02	0.46	-0.09	0.02	0.46	0.19	-0.65	7.06	-0.20	0.20	7.06	2.64
PMLCalDF2	-0.08	0.66	0.01	0.01	0.66	0.25	-0.18	1.39	0.09	0.09	1.39	0.56
PMLCalDF3	-0.03	0.36	-0.04	0.03	0.36	0.14	-0.16	0.72	0.24	0.16	0.72	0.38
SFRR	-0.31	0.12	-0.22	0.12	0.31	0.22	0.02	-0.09	0.24	0.02	0.24	0.12
CalSF	-0.59	0.19	-0.31	0.19	0.59	0.36	0.32	-0.02	-0.04	0.02	0.32	0.13
Hartley	-1.95	0.17	-0.54	0.17	1.95	0.89	-2.08	-0.05	-0.12	0.05	2.08	0.75
FB	-1.99	0.27	-0.53	0.27	1.99	0.93	-2.01	0.08	-0.14	0.08	2.01	0.74
PML	-0.18	0.14	-0.22	0.14	0.22	0.18	0.13	-0.11	0.35	0.11	0.35	0.20
PEL	-0.66	0.11	-0.08	0.08	0.66	0.28	0.49	-0.13	0.22	0.13	0.49	0.28
CalDF	-0.37	0.17	-0.32	0.17	0.37	0.29	1.19	0.01	-0.36	0.01	1.19	0.52

ISBN: 978-84-616-9216-3