Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 66(1) 2012 145-156

On the Influence of Sampling Design on Small Area Estimates

Ralf T. Münnich* **and J. Pablo Burgard**
University of Trier, Trier, Germany
Received 13 May 2011; Revised 20 August 2011; Accepted 20 August 2011

SUMMARY

Recent advances in small area statistics applications raised the question on the influence of sampling designs on model based estimates. On the one hand, weighting was introduced in the modelling (cf. You and Rao 2002). On the other hand, Gelman (2007) argues that sampling designs with highly variable design weights should be avoided in order to support statistical modelling and especially Bayesian modelling.

The present article gives some ideas on the interplay of modelling and survey weights based on a realistic simulation study motivated by the experience from several research projects. Further, recommendations are given on how to control the size of survey weights in optimal sampling designs via a box-constraint optimal allocation, introduced by Gabler *et al.* (2010). A practical study gives some ideas on the impact of design-effects and the variability of design weights on model versus design-based estimation methods.

Keywords : Complex sampling designs, Monte Carlo study, Design effects, Gelman factor, Small area estimation.

1. INTRODUCTION

Gelman (2007) reopened a discussion on the sense of sampling designs by stating *survey weights are a mess*. The main arguments are motivated by the difficulty of drawing correct inferences in Bayesian statistics once the data is gathered from complicated sampling designs with high variations in survey weights. This invited discussion lead to a vivid discussion on statistical modelling using survey data.

Meng *et al.* (2009) emphasized that the spread of survey weights should not exceed 10 and is unacceptable beyond 100. If π_i is the inclusion probability for the i'th unit in the selected sample under the sampling design S in a finite population with N elements, the spread of design weights is defined by

$$\text{GF:} = \frac{\max_{1 \le i, j \le N} \pi_j}{\pi_i} \tag{1}$$

which we denote as the Gelman-factor GF. The ideal case for statistical modelling is the use of simple random sampling where GF = 1. Business statistics application in sample surveys, however, may lead to very large Gelman-factors which may easily exceed 1,000 and, hence, are unacceptable following the argumentation of Meng *et al.* (2009).

The question arises whether a large GF always influences statistical modelling negatively or whether certain alternative conditions may lead to varying impacts of the GF on the reliability of modelling. Consequently, one may be interested, whether problems in statistical modelling also influences negatively survey estimates which, in general, are restricted to means, totals, and proportions, that are based on statistical models.

In this paper, we focus on statistical modelling in the small area estimation context. Special emphasis will

**Corresponding author* : Ralf T. Münnich
E-mail address : muennich@uni-trier.de

be put on classical two-stage random sampling designs which are common in many household surveys. The impact of sampling designs on the efficiency of estimates is thoroughly studied for design-based methods. In the case of stratified random sampling the solution is the optimal allocation (cf. Neyman 1934 and Tschuprov 1923). One can also find in standard textbooks (cf. Lohr 1999 and Särndal *et al.* 2003) that cluster sampling, in general, reduces the efficiency of the estimate. For small area estimation, it seems necessary to guarantee minimal sample sizes in all strata. An equal allocation ensures a good spread of the total sample size over the areas which optimizes the accuracy in each area. Costa *et al.* (2004) indicate that a convex combination of equal and proportional allocation is a compromise between national- and area-level minimization of the MSE. Longford (2006) proposes a sample allocation to the areas according to a relative importance of area-specific variances. Choudhry *et al.* (2011) use non-linear programming techniques to minimize the overall sample size while considering certain minimal efficiency requirements measured by the coefficient of variation on area and aggregate estimates. We propose the use of a box-constraint optimal allocation which minimizes the 2-norm of the vector of the relative root mean squared errors of the estimates while considering certain minimal and maximal sampling fractions (see Gabler *et al.* 2010). This allocation can also be applied to stratified random samples within the areas. Further, it allows to control the Gelman-factors.

In the next Section, different sampling designs are presented that are of interest for the Monte-Carlo study following later. The focus is put on the recent methodology of using box-constraints in stratified random sampling. In the third Section, a selection of classical and modern small area estimators are presented which are widely used in practice. Section 4 gives an overview of the results from a comparative Monte-Carlo study. The focus of the study is put on showing the impact of design on the accuracy of the small area estimates. The study is based on the synthetic data set AMELIA (cf. Alfons *et al.* 2011) which provides a highly skewed study variable.

2. SELECTED SAMPLING DESIGNS IN SMALL AREA APPLICATIONS

Household surveys generally use two- or more-stage designs mainly due to practical reasons. Within the scope of two-stage designs stratified random sampling and cluster sampling can be seen as two *opposite* designs, especially by efficiency reasons. In general, stratified samples may be used to gain efficiency of survey estimates. In case size and standard deviation of the strata are less correlated, the application of the optimal allocation may lead to a further gain in efficiency. The very low design effect, however, may lead to highly varying design weights between the strata which automatically yield a high GF. On the other hand, cluster sampling may lead to large design effects whereas it is unusual that the weights vary a lot. These results are well-known for survey estimates in a design-based framework but little is known about these effects in statistical modelling or small area estimation.

In order to appropriately investigate these effects in small area statistics, the following sampling designs are chosen for the comparative study:

Stratified Random Sampling with Equal Allocation

In this case, all strata have the same sample size. The variation of the design weights depend only on the strata sizes.

Stratified Random Sampling with Proportional Allocation

The proportional allocation ensures a GF = 1. The design effect depends on the variance within the strata with respect to the study variable.

Stratified Random Sampling with Optimal Allocation

Against real world application, we have chosen the variable of interest to determine the within stratum variance. This ensures a maximal gain in efficiency but can be done only in simulations or cases were the universe is known (theoretically). In general, this allocation yields the largest GFs within stratified random sampling. The design effect is minimal.

Cluster Sampling with Constant Cluster Size

In this case, the cluster sizes are chosen equal in size. The GF depends on the number of clusters drawn in an area with regard to the size of the area. In this application the allocation used is a proportional allocation of samples over the areas. As such, the GF is moderate. The design effect, however, may be larger.

Cluster Sampling with Random Cluster Size

In many applications cluster sizes are not fixed but random. The variation of cluster sizes makes the estimation more complicated. Also the total sample size over all areas is not fixed. In this study, however, the expected sample size is set to the same sample size as in the before mentioned design.

These allocations are denoted by *equal*, *prop*, *optall*, *fix*, and *ran* respectively. In all cases planned areas were applied.

Comparing the above sampling designs, the optimal allocation should deliver by definition the most efficient estimates in classical sampling theory. However, the GF may be very high. In order to control it, Gabler *et al.* (2010) have introduced the box-constraint optimal allocation. Let $\hat{\tau}_1, ..., \hat{\tau}_D$ be a set of design-based Horvitz-Thompson statistics on D areas with according variances. Further, let m_h and M_h be the lower and upper boundary of the sample sizes in each area (stratum). Then, the box-constraint optimal allocation for a fixed total sample size with D areas and H strata is the solution of

$$\min_{n_{1,1}, ..., n_{D,1}, ..., n_{D,H}} \left\| \text{RRMSE}\left(\hat{\tau}_*\right) \right\|_2 \qquad (2)$$

where RRMSE denotes the relative root mean squared error for a given estimator and $n_{d,h}$ is the samples size within area d and stratum h. The solution can be seen as a compensatory 2-norm functional of the root MSEs of D area-specific estimates from either simple or stratified random sampling within the areas. Hence, the stratum-specific sample sizes are determined such that the variances of the domain estimates are minimized simultaneously according to the overall functional. The box-constraint optimal allocation can be used to control the GF by tweaking m_h and M_h. Details of the allocation and the analytic solution can be drawn from Gabler *et al.* (2010) who used a small area combined regression estimator. A numerical improvement and comparison of iterative algorithms is given in Münnich *et al.* (2011). In order to compare the impact of different GF bounds, the parameters of interest are chosen as GF = 25, 10 and denoted *opt025* and *opt010* respectively.

3. SMALL AREA ESTIMATORS AND MODELS

Traditional surveys use design-based methods to produce population figures. These methods usually need a high number of observations per area in order to obtain reliable estimates. In many surveys only small sample fractions are available within certain areas leading to imprecise design-based estimates. The precision of an estimate is usually assessed within one sample by an variance estimator for the point estimate. When using design-based estimates like the Horvitz-Thompson estimator (Horvitz and Thompson 1952), which is design unbiased, the use of confidence intervals as a measure of accuracy usually yields reliable results. Model based small area estimators are usually model unbiased but not necessarily design unbiased (cf. Rao 2003). Biased estimates, however, may have a negative impact on the coverage of confidence intervals which cannot be observed using a single sample but in simulation studies only. In the later study, we will assess the quality of the different area-specific point estimates using the RRMSE of the point estimates on the one hand and the simulated coverage rates on the other hand.

One well known and stable design-based estimator for area means is the GREG estimator:

$$\hat{\mu}_{d,\text{GREG}_*} = \bar{X}_d^T \hat{\beta}_* + \frac{1}{N_d} \sum_{i=1}^{n_d} w_i \cdot \underbrace{(y_{i,d} - x_{i,d}^T \hat{\beta}_*)}_{e_{i,d}} \qquad (3)$$

with $$\hat{\beta}_* = (x_*' W_* x_*)^{-1} x_*' W_* y_* \qquad (4)$$

(cf. Särndal *et al.* 2003), where $\hat{\beta}_*$ is the solution to the linear regression model (4). W_* is a diagonal matrix of the weights $w_{i,*}$, $y_{i,*}$ is the variable of interest, and $x_{i,*}$ and $X_{i,*}$ are the auxiliary information (the first column being 1 for all units) of the i'th household within the sample and register respectively for a given cumulation of areas *. If * = d then the cumulation is the area and if * = g the cumulation is a union of sampling units or areas to sensible groups. With *GREGA* we denote a regression estimator that uses the $\hat{\beta}_d$ whereas the *GREGC* uses a $\hat{\beta}_c$ over the whole sample and, hence, is a combined regression estimator. As variance estimator, we use the classical residual variance estimator considering the estimated regression coefficient and the design in use (cf. Lohr 1999). An extension of the GREG approach to categorical data can be found in Lehton and Veijaen (2009), (1998).

Another approach to the estimation of small area means is proposed by Battese *et al.* (1988). The methodology underlying the proposed estimator can be seen as a unit-level mixed model (cf. Jiang and Lahiri 2006). Usually a two-level mixed model is used. The unit-level variation is often referred to as sampling variance, and the area-level variation as area effect. Both are assumed to be normally distributed with mean zero and variances σ_e and σ_u respectively. The variables x, X and y are defined as before. V is the block diagonal variance covariance matrix as defined in Rao (2003, p. 135).

$$\hat{\mu}_{d,\text{ULEBLUP}} = \bar{X}_d \hat{\beta} + \hat{u}_d$$

$$\hat{u}_d = \hat{\gamma}_d \left(\bar{y} - \bar{x}_d \hat{\beta}\right), \text{ and } \hat{\gamma}_d = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \frac{\hat{\sigma}_\varepsilon^2}{n_d}}$$

$$\hat{\beta} = (x' V^{-1} x)^{-1} x' V^{-1} y \qquad (5)$$

In the standard unit-level model the design weights are not being considered. Reliable results are thus only to be expected under simple random sampling. The use of design weights for estimating the β vector

$$\hat{\beta}_w = \left(x' V^{-\frac{1}{2}} W V^{-\frac{1}{2}} x\right)^{-1} x' V^{-\frac{1}{2}} W V^{-\frac{1}{2}} y \qquad (6)$$

may help to correct for unequal probabilities in sampling. The unit-level model without considering weights will be called *ULEBLUP* the one with the weighted β will be called *ULWEBLUP.*

In order to measure the accuracy of these small area estimates, we use the Prasad and Rao (1990) MSE estimator which is divided into three components in the following way:

$$\text{MSE}\left(\hat{\tau}_{d,\text{EBLUP}}(\hat{\psi})\right) \approx g_{1d}(\hat{\psi}) + g_{zd}(\hat{\psi}) + 2 g_{3d}(\hat{\psi}) \qquad (7)$$

The three components are :

$$g_{1d}(\psi) = \frac{\frac{\sigma_u^2 \sigma_\varepsilon^2}{n_d}}{\sigma_u^2 + \frac{\sigma_\varepsilon^2}{n_d}} = \frac{\sigma_u^2 \sigma_\varepsilon^2}{n_d \sigma_u^2 + \sigma_\varepsilon^2}$$

$$g_{2d}(\psi) = \left(\bar{X}_d - (1 - B_d)\bar{x}_d\right)' (X' V^{-1} X)^{-1} \left(\bar{X}_d - (1 - B_d)\bar{x}_d\right)$$

and

$$g_{3d}(\psi) = \frac{n_d^{-2}}{(\sigma_\varepsilon^2 + n_d \sigma_u^2)^3} \{\sigma_\varepsilon^4 I^{uu} + \sigma_u^4 I^{\varepsilon\varepsilon} - 2\sigma_u^2 \sigma_\varepsilon^2 I^{u\varepsilon}\}$$

with,

$$I^{uu} = 2a^{-1} \sum_{u=1}^{k} \left[(n_u - 1)\sigma_\varepsilon^{-4} + \eta_u^{-2}\right], \; I^{\varepsilon\varepsilon} = 2a^{-1} \sum_{u=1}^{k} n_u^2 \eta_u^{-2}$$

$$I^{u\varepsilon} = -2a^{-1} \sum_{u=1}^{k} n_u \eta_u^{-2}, \; \eta_d = \sigma_\varepsilon^2 + n_d \sigma_u^2$$

$$a = \left[\sum_{u=1}^{k} N_u^2 \eta_u^{-2}\right] \left[\sum_{u=1}^{k} \left\{(n_u - 1)\sigma_\varepsilon^{-4} + \eta_u^{-2}\right\}\right] - \left(\sum_{u=1}^{k} n_u \eta_u^{-2}\right)^2$$

Since $\psi = (\sigma_u^2, \sigma_\varepsilon^2)$ is unknown, $g_{2d}(\hat{\psi})$ and $g_{3d}(\hat{\psi})$ are used to estimate g_{2d} and g_{3d}. The estimator $g_{1d}(\hat{\psi})$ for $g_{1d}(\psi)$ is biased. For this reason Prasad and Rao (1990) use a bias corrected estimator $g_{1d}(\hat{\psi}) + g_{3d}(\hat{\psi})$ which leads to the estimator in equation (7). Prasad and Rao (1990) show that the estimators for the components have only a bias of low order.

A general presentation of this MSE estimator may be found in Datta and Lahiri (2000). The estimation of the $\hat{\beta}$ and the $\hat{u}$ for the two level mixed model can be found in Battese *et al.* (1988). For more complex models see Rao (2003) chapters 5-8. An extension to this model for binary data including MSE estimators is proposed by González-Manteiga *et al.* (2007).

You and Rao (2002) show a way how to incorporate design weights directly into the small area modelling. Further, they estimate the β under constraints, such that they achieve a self benchmarking property to the national estimate. This estimator will be called *YOURAO* and is given by:

$$\hat{\mu}_{d,\text{YOURAO}} = \hat{\gamma}_{d\tilde{w}} \bar{y}_{d\tilde{w}} + \left(\bar{X}_d - \hat{\gamma}_{d\tilde{w}} \bar{x}_{d\tilde{w}}\right)^T \hat{\beta}_{\tilde{w}} = \bar{X}_d^T \hat{\beta}_{\tilde{w}} + \hat{u}_{d\tilde{w}} \qquad (8)$$

with $\hat{u}_{d\tilde{w}} = \hat{\gamma}_{d\tilde{w}}(\bar{y}_{d\tilde{w}} - \bar{x}_{d\tilde{w}}^T \hat{\beta}_{\tilde{w}})$, $\hat{\gamma}_{d\tilde{w}} = \dfrac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_e^2 \delta_d^2}$,

$$\delta_d^2 = \sum_{i=1}^{n_d} \tilde{w}_{id}^2, \; \tilde{w}_{id} = \frac{w_{id}}{\sum_{i=1}^{n_d} w_{id}},$$

$$\overline{y}_{d\tilde{w}} = \sum_{i}^{n_d} \tilde{w}_{id}\, y_{id}, \quad \overline{x}_{d\tilde{w}} = \sum_{i}^{n_d} \tilde{w}_{id}\, x_{id}$$

and $\hat{\beta}_{\tilde{w}} = (X^T W (X - \gamma_{d\tilde{w}} \overline{X}_{d\tilde{w}}))^{-1} W (X - \gamma_{d\tilde{w}} \overline{X}_{d\tilde{w}}) y)$. The proof and properties of this estimator can be drawn from You and Rao (2002). Torabi and Rao (2010) propose a second-order approximation to the (MSE) for the *YOURAO* estimator. They achieve by this approximation a nearly unbiased MSE estimate. However, for large datasets this estimator needs some additional algorithmic extension to be included in the simulation study.

In many applications no unit-level information is available. For these cases Fay and Herriot (1979) proposed the so called Fay-Harriot estimator. This estimator is closely related to the ULEBLUP (see Datta and Lahiri 2000). In the case of the Fay-Harriot model all information is only available on aggregate level:

$$\hat{\mu}_{d,\,\mathrm{ALEBLUP}} = \overline{X}_d^T \hat{\beta} + \hat{u}_d \qquad (9)$$

where $\hat{u}_d$ is estimated like in equation (5). This estimator will be referred to as *ALEBLUP*. As well for the *ALEBLUP* as for the *YOURAO* estimator the MSE estimation is done via the Prasad and Rao (1990) MSE estimator (7).

4. SIMULATIVE ASSESSMENT

4.1 Design of the Simulation Study

The aim of the simulation study is to evaluate the influence of the above sampling designs on a standard small area setting. As simulation population, the synthetic data set Amelia is used (cf. Alfons *et al.* 2011). It consists of approximately 3.7 Mio. households in 78 areas with a varying size from 3427 to 243092 households. The sample size is 6000 which yields an overall sampling fraction of approximately 0.16%. The simulation set-up was chosen similarly to Burgard and Münnich (2010).

The dependent variable is the total disposable household income (HY020) which is generated according to the structures found in the EU-SILC data. It is a skewed variable with some extreme outliers. In Fig. 1 a density plot of the HY020 variable is drawn. Each curve represents one area. As can be seen on the left part of this figure the mass of the distributions of variable HY020 are overlapping over all areas. Nevertheless, the means and outliers are very different. On the right hand side one can see that there are some very extreme outliers in both directions.

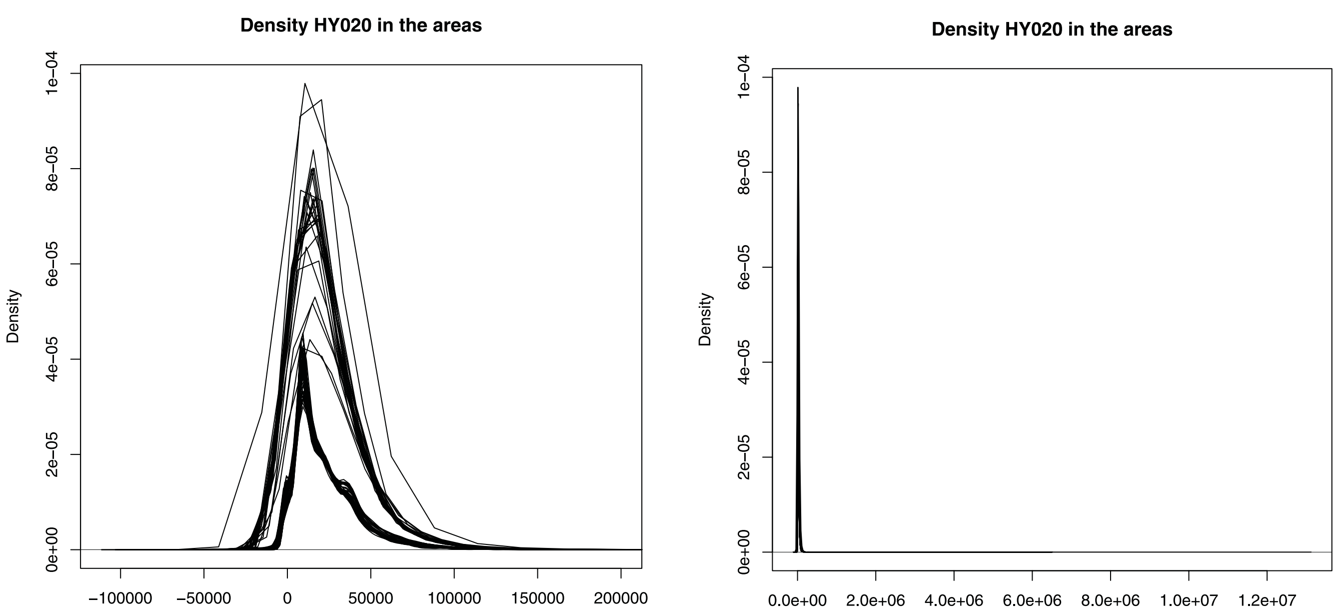

Fig. 1. Density plot of *total disposable household income (HY020)* in each area

In order to outline the assets and drawbacks of the above mentioned designs and allocations in the small area context it is important to have stereotype scenarios. These are created by controlling the scale and the variability of the dependent variable within the strata or clusters. This is achieved by assigning the households to three different groupings (*rand*, *sort*, and *raso*) each describing five equally sized strata.

rand: The households within each area are randomly assigned to five strata.

sort: The five strata are the quintiles of the HY020 by area.

raso: Is a mixture of rand and sort. Stratum 1 contains every 5'th household along the sorted HY020, the strata 2-3 are the first and second quartile from the remaining households respectively, and the strata 4-5 are randomly assigned from the remaining households.

The strata of the groupings are then used as strata for the stratified designs. To the stratified designs the five in Section 2 explained allocations are applied. For the cluster designs with fixed cluster sizes (*fix*), clusters of size ten are created within in each stratum. If the stratum size is not a multiple of ten, then some clusters only have the size of nine. In the cluster designs with random cluster sizes (*ran*) the households are assigned again randomly within the strata to clusters. The cluster sizes were set to 5-15 households. Of every cluster size there are approximately the same amount of clusters within each stratum. The expected sample size is set to be equal to the case of proportional allocation for the respective grouping. For the simulation study 10,000 samples were drawn for every cross combination of groupings and allocations.

In Table 1 the ratio of the design weights over the whole population is stated. The third column shows the GF for all the scenarios. The other columns show the ratios of symmetric quantiles according to their naming. One can see that the GF in the case of the designs *equal*, *prop*, *fix*, *ran* does not change with the grouping since the designs do not depend on the variable of interest. Further, in the case of the grouping *rand* the ratios of designs weights do not vary very much between the different designs. However, only a slightly higher ratio for the optimal allocations *opt010*, *opt025*, and *optall* can be observed. This results from the fact that the stratum-specific variances do not vary much under the grouping *rand*. The GF for the *opt010*, and *opt025* allocations is slightly over 10 and 25 respectively. This is due to the fact that the above described allocation scheme does produce decimals and not integers. Thus,

Table 1. Ratios of the design weights

Allocation	Grouping	GF	*q*95/*q*05	*q*80/*q*20	*q*60/*q*40
Cluster Designs					
fix	all equal	4.25	1.18	1.08	1.02
ran	all equal	4.00	1.18	1.08	1.02
Stratified Designs					
equal	all equal	66.54	8.70	1.61	1.12
prop	all equal	2.04	1.08	1.04	1.01
opt010	rand	3.93	3.14	1.31	1.07
	raso	11.85	10.62	5.03	1.14
	sort	12.82	9.87	5.25	1.21
opt025	rand	7.68	3.20	1.32	1.07
	raso	31.56	19.00	5.42	1.11
	sort	31.68	27.67	9.22	1.25
optall	rand	7.68	3.20	1.32	1.07
	raso	78.55	29.93	6.56	1.12
	sort	128.55	48.51	10.10	1.23

the allocated sample sizes have to be rounded and the GF differs from the theoretical one. Due to the low sampling fractions in the strata the rounding does have a significant effect on the design weights. However, the effect of the box-constrained optimal allocation scheme on the GF is evident.

4.2 Results of the Simulation Study

The following graphs give an overview of the efficiency of the mean disposable income small area estimates under the different designs and groupings. Figs. 2 and 3 compare the impact of the grouping on the estimates. As can be seen in Fig. 2 in the case of *rand*, the allocation of sample sizes to strata does not have a strong effect on the RRMSE of the considered estimators. This is due to the fact that the stratification does not partition the population in more homogeneous sub-populations. As such, the optimal design does not depart much from the proportional allocation case as the variances within strata are all very similar. In this setting, the small area estimators *ULEBLUP, ULWEBLUP, ALEBLUP* do behave slightly better than the classical GREG estimators. Further, it can be seen that the *GREGC* outperforms the *GREGA* in this setting. This is due to the fact, that the *GREGC* uses all information available in the sample to estimate the β. Thus, the estimation is more stable than in the case of *GREGA* where only the sampled observations in one area are used. This pattern holds for all scenarios in this simulation setup as can be seen in Fig. 3. Here, the estimators *GREGA, GREGC, ULWEBLUP* are compared under the groupings *sort* and *raso* and different sampling designs. In the cases of the stratified allocations with *equal* and *prop* allocation, the grouping has almost no effect on the RRMSE of the estimators. In contrast, for the optimal allocations the RRMSE is even higher in the case of the grouping *sort* compared to *raso*. This is due to a more unstable β estimation because of very small sample sizes in strata with low variability of the variable of interest.

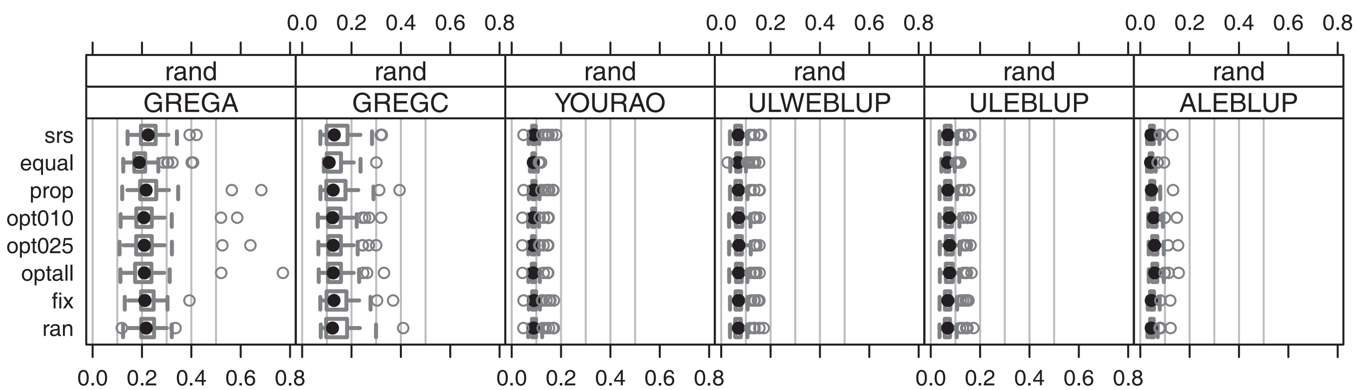

Fig. 2. RRMSEs of the mean income estimates under the grouping *rand*

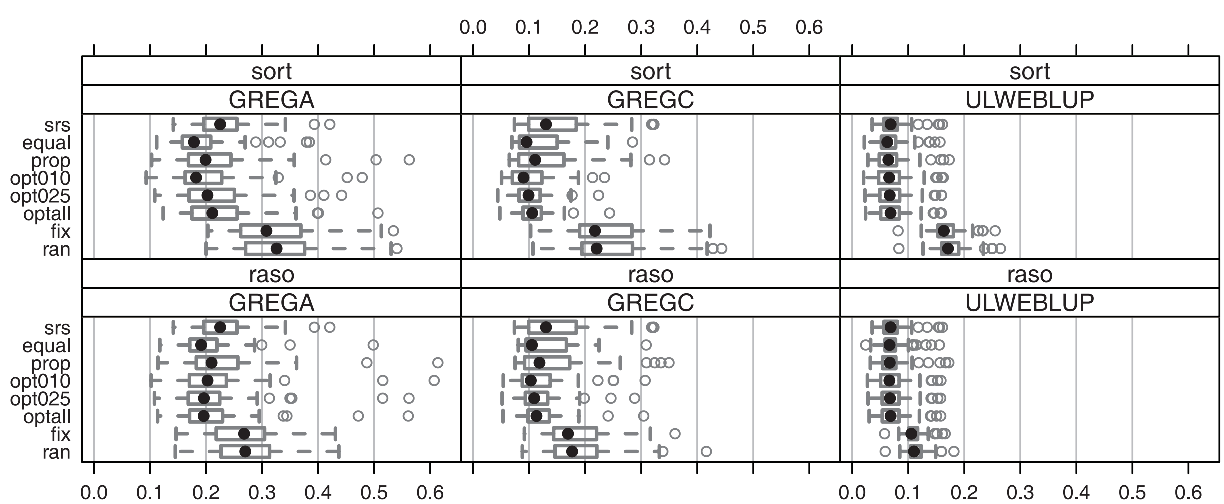

Fig. 3. RRMSEs of the mean income of *GREGA*, *GREGC* and *ULWEBLUP* under grouping *sort* and *raso*.

As expected, one can see that the optimal allocation compared to the proportional allocation improves the estimation in terms of RRMSE. Further, the *GREGC* yields the best results for the optimal allocation under *sort*. Surprisingly, the *GREGA* with optimal allocation seems to perform better under *raso* than under *sort*. In case of small sample sizes or small supports for the β estimates, the interaction between the estimates of the auxiliary variable and the β's cannot be neglected which results in higher variances of the *GREGA*. Similar observation have been made in the German Census research project. The combined regression estimators outperformed separate regression estimators even if the areas are not very small (cf. Münnich *et al.* 2012). However, these differences tend to be very small and the results cannot be generalised.

In case of the cluster design, one can observe strong effects of the grouping on the performance of the estimators. For all estimators the RRMSE under the grouping *sort* is higher than under *raso* and *rand*. Hence, the well-known cluster effect in design-based theory seems also to play a role in modelling, most likely by the indirectly reduced sample size due to clustering. Also important to note is that the impact of clustering is more evident than a high GF. The effect of varying cluster sizes results in slightly higher RRMSEs. This little impact was expected due to the little variation of cluster sizes, which may be much more evident in practice. Hence, it seems also worth controlling the variability of cluster sizes for statistical modelling.

In Fig. 4, different small area estimators are compared. The impact of the sampling designs on the unweighted small area estimator *ULEBLUP* is very strong. In this case, we can easily recognize the influence of the GF which is even more problematic under the grouping *sort*. Once an optimal allocation is applied, the RRMSE of the *ULEBLUP* rises considerably. Especially in the situation where the optimal designs from a design-based perspective is most efficient (combination *sort* and *optall*) the *ULEBLUP* is near to useless. For the unweighted estimator, the GF really shows to be an appropriate indicator for the possible design impact. The lower the GF is, the lower is the RRMSE.

In the situation of *raso*, the optimal allocation schemes do not have very different effects. Hence, the restriction of the GF is not necessary. In contrast, when the β is estimated using the design weights like the *ULWEBLUP* then the unit-level small area model can gain efficiency by the design. Again, in the case of cluster designs, both models *ULEBLUP* and *ULWEBLUP* do have an higher RRMSE. The inclusion of design weights in the estimation of the β does not play a major role in this comparison as the design weights do not vary too much.

In many applications only aggregate data are available for small area estimation. In this case an area model has to be used. As can be seen in Fig. 2, in the nice case where the allocation has almost no effect on the inclusion probabilities, the *ALEBLUP* performs very

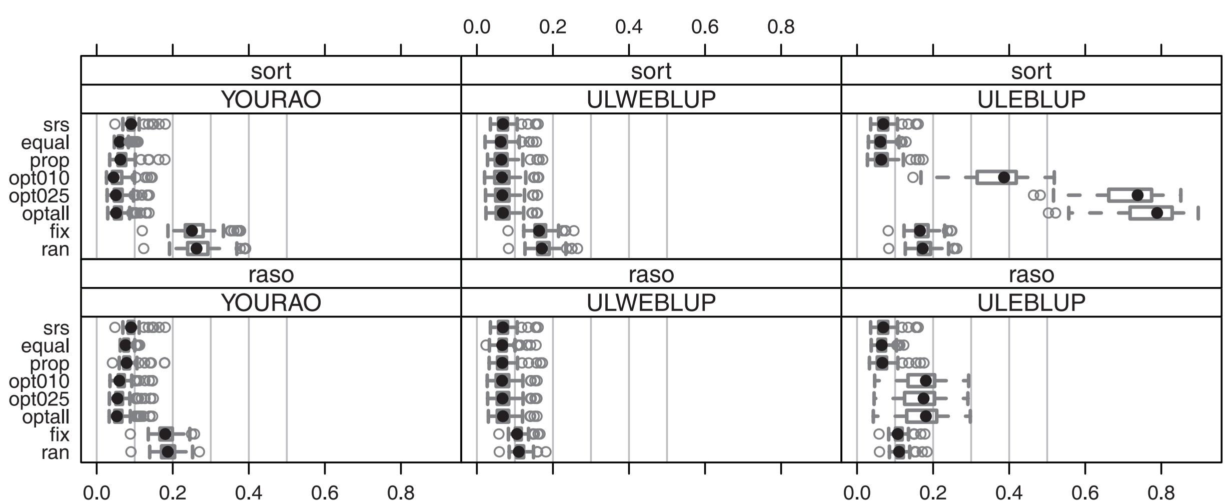

Fig. 4. RRMSEs of the mean income estimates of *YOURAO*, *ULEBLUP*, and *ULWEBLUP*

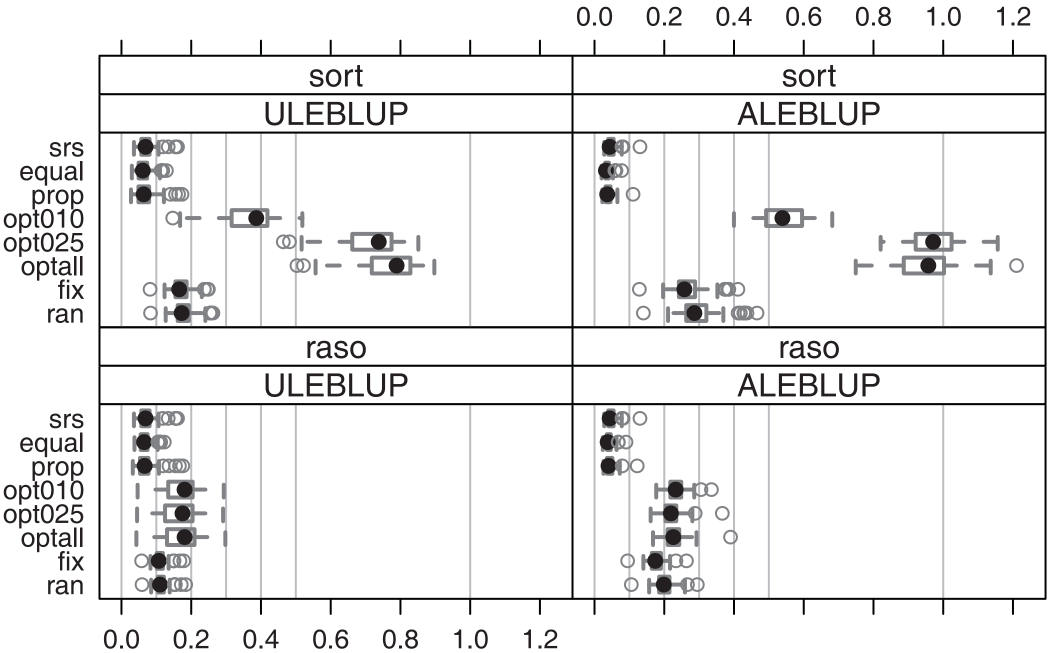

Fig. 5. RRMSEs of the mean income estimates of the unweighted models

well. In Fig. 5, the *ALEBLUPB* and the *ULEBLUP* are compared. As can be seen, both estimators have similar problems with the sampling design. However, the *ALEBLUP* is still more negatively influenced by high GFs which results in much higher RRMSEs. Surprisingly, little differences can be observed between *equal* and *prop* allocations.

In Fig. 6 the 95% confidence interval coverage rates are plotted against the mean confidence interval length divided by the true mean. Ideally the points for the areas would lie on the left of the 95% line, the more to the left the better. As could be seen in the Fig. 3 the *GREGC* outperforms the *GREGA* in terms of RRMSE in almost all cases. In Fig. 6 one can further observe that also the variance estimates for the *GREGC* performs better than for the *GREGA*. The residual variance estimator seems to be negatively influenced by the higher variability of the β estimates. Enhanced residual variance estimation methods may reduce this effect (e.g. Särndal *et al.* 2003, p.401).

Further, the cluster designs show a negative effect on the coverage rates which follows from relatively poor variance estimates. Amazingly, also the *GREGA* and *GREGC* have serious problems which partly is dependant on the grouping. The effect seems to be induced by the skewed distribution of the variable of interest.

Another interesting finding is that in case of the grouping *rand* the MSE estimate for the *ALEBLUB* measures the variability of the point estimate quite accurately in all designs. However, if the grouping departs from randomness the MSE estimates perform rather poorly, in case of optimal designs they are useless. A little surprise was the impact of sorting on the *YOURAO* under cluster sampling. It is not clear whether this can be reduced by using the Torabi and Rao (2010) MSE estimator instead of Prasad and Rao (1990).

Finally, Fig. 7 indicates that the unweighted methods show much larger design effects than all other estimators. In these cases, sampling designs with a high GF seem to be worst.

In general, one would recommend using weighted small area estimators. These work fine as there is not unfortunate clustering where the clusters are effected by some pre-sorting. Without random groupings, the impact of the settings is much less than in the other cases. An alternative approach to incorporate sample information can be achieved by augmenting the model with design variables (cf. Verret *et al.* 2010).

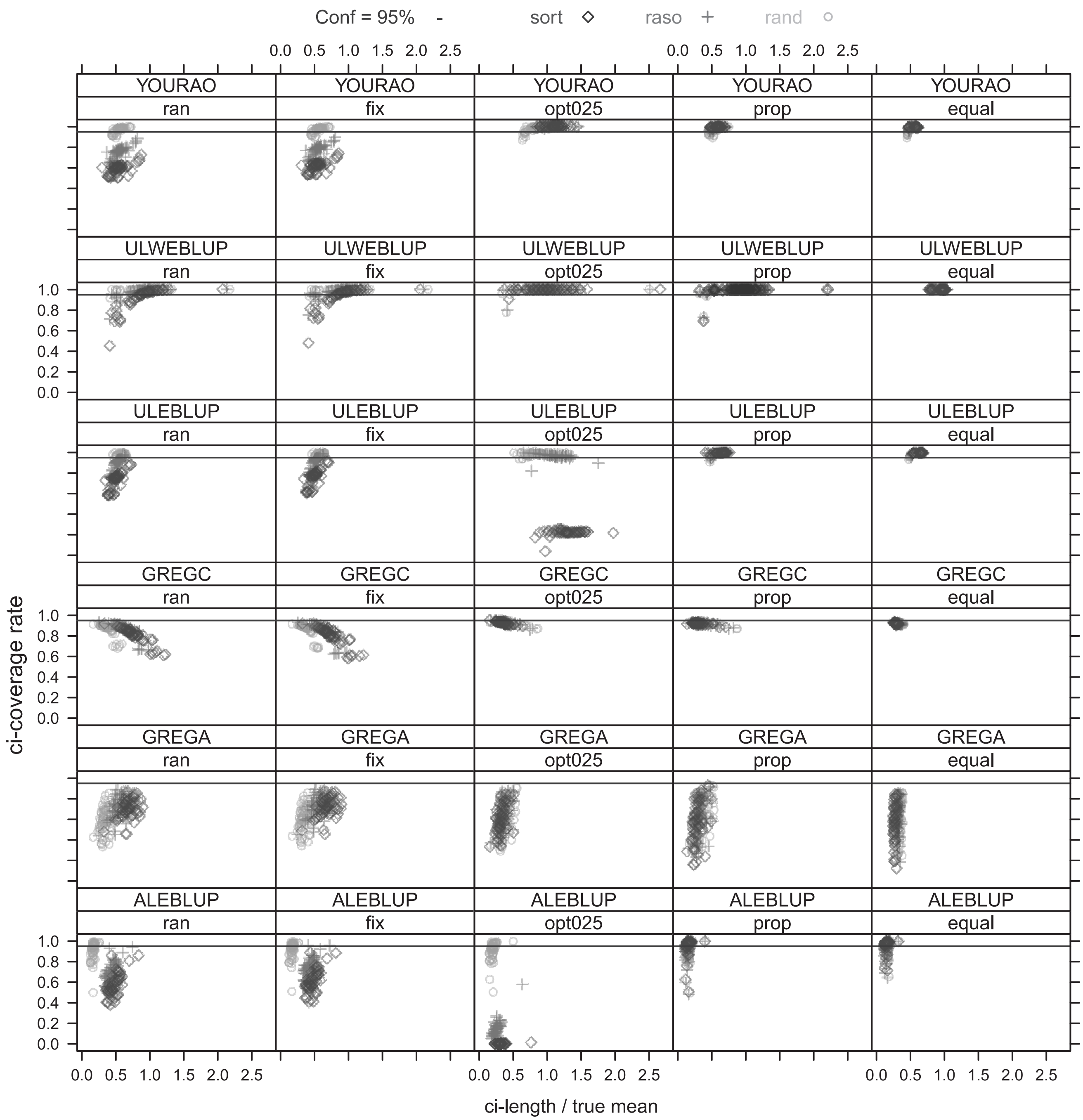

Fig. 6. Confidence interval coverage rates versus lengths for mean income estimates

5. CONCLUSION AND OUTLOOK

The aim of the paper was to investigate the influence of design weights on different small area estimators. According to Gelman's critique on survey weights, we were interested whether the major problems arise in surveys with a high GF or whether further issues have to be acknowledged.

In the Monte-Carlo study, a set of different estimators and designs were compared under 3 differently grouped population. One could observe that a high GF within stratified designs really gives an impact on modelling. However, this negative effect can be reduced considerably by using design weighted regression methods, and especially the *YOURAO*

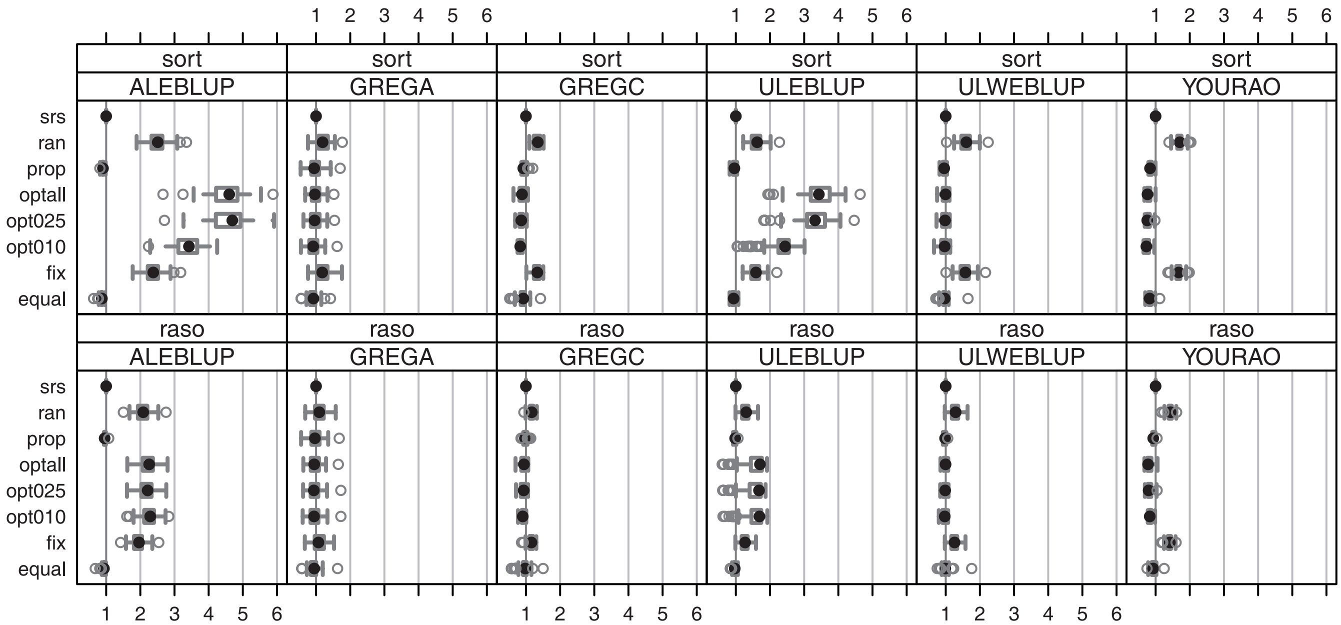

Fig. 7. Design effects for the grouping *sort*

estimator. Further, including box-constraints in the allocation seemed to be a helpful tool in order to control the GF and ensure minimal efficiencies in the area-specific estimates. Major negative effects occurred in cluster sampling once some pre-sorting could be observed within the data. In case of random allocation of the clusters the estimators were much more stable. This effect is also known from design-based theory.

As can be seen in Fig. 1 the data contains some influential outliers. These outliers may inflict heavily on the estimation of models leading to bad estimates. One way to overcome these peculiarities is the application of robust methods. Two such robust small area methods are the outlier robust EBLUP estimator for linear mixed models from Sinha and Rao (2009) and the robust M-quantile approach from Chambers and Tzavidis (2006). If spatial dependencies are present in the data, then spatial robust small area models may be applied, as the spatial robust EBLUP (SREBLUP) proposed by Schmid and Münnich (2012) or the M-quantile geographically weighted regression model proposed by Salvati *et al.* (2011).

Further, one should note that the given designs omitted the appearance of unsampled areas. In case of unsampled areas or domains caution has to be taken, that the selection of the area with allocated samples is independent from the variable of interest. Otherwise adapted small area estimators under informative sampling (Pfeffermann *et al.* 2006) have to be applied.

ACKNOWLEDGEMENTS

The research was started within the AMELI project in which the simulation environment and the dataset were generated. The main part of the research is included in the BLUE-ETS project. Both projects are financially supported by the European Commission within the 7th Framework Programme (cf. http://ameli.surveystatistics.net and http://www.blue-ets.eu).

REFERENCES

Alfons, A., Templ, M., Filzmoser, P., Kraft, S., Hulliger, B., Kolb, J.P. and Münnich, R. (2011). Synthetic Data Generation of SILC Data. Technical Report, AMELI deliverable D6.2.

Battese, G.E., Harter, R.M. and Fuller, W.A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *J. Amer. Statist. Assoc.*, **83(401)**, 28-36.

Burgard, J.P. and Münnich, R. (2010). Modelling over and under counts for design-based Monte Carlo studies in small area estimation : An application to the German register-assisted census. *Compu. Statist. Data Anal. DOI* : 10.1016/;csda.2010.11.002.

Chambers, R. and Tzavidis, N. (2006). M-Quantile models for small area estimation. *Biometrika*, **93(2)**, 255-268.

Choudhry, G.H., Rao, J.N.K. and Hidiroglou, M.A. (2011). On sample size allocation for planned small domains. Submitted.

Costa, À., Satorra, A. and Ventura, E. (2004). Improving both domain and total area estimation by composition. *SORT: Statist. Oper. Res. Trans.*

Datta, G.S. and Lahiri, P.A. (2000). Unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statist. Sinica*, **10**, 613-627.

Fay, R.E. and Herriot, R.A. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *J. Amer. Statist. Assoc.*, **74(366)**, 269-277.

Gabler, S., Ganninger, M. and Münnich, R. (2010). Optimal allocation of the sample size to strata under box constraints. *Metrika*. DOI : 10.1007/s00184-010-0319-3.

Gelman, A. (2007). Struggles with survey weighting and regression modeling. *Statist. Sci.*, **22(2)**, 153-164.

González-Manteiga, W., Lombarda, M.J., Molina, I., Morales, D. and Santamara, L. (2007). Estimation of the mean squared error of predictors of small area linear parameters under a logistic mixed model. *Comput. Statist. Data Anal.*, **51**, 2720-2733.

Horvitz, D.G. and Thompson, D.J.A. (1952). Generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.*, **47(260)**, 663-685.

Jiang, J. and Lahiri, P. (2006). Mixed model prediction and small area estimation. *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research*, **15(1)**, 1-96.

Lehtonen, R. and Veijanen, A. (2009). Design-based methods of estimation for domains and small areas. In: C.R. Rao, editors, *Handbook of Statistics - Sample Surveys: Inference and Analysis,* 219-249. Elsevier.

Lehtonen, R. and Veijanen, A. (1998). Logistic generalized regression estimators. *Survey Methodology*, **24(1)**, 51-55.

Lohr, S.L. (1999). *Sampling: Design and Analysis*. Duxbury Press, 1 edition.

Longford, N.T. (2006). Sample size calculation for small-area estimation. *Survey Methodology*, **32(1)**, 87-96.

Meng, X.L., Duan, N., Chen, C. and Alegria, M. (2009). Power-shrinkage: An alternative method for dealing with excessive weights. *Joint Statistical Meetings*.

Münnich, R., Gabler, S., Ganninger, M., Burgard, J.P. and Kolb, J.P. (2012). *Design und Schätzung im Zensus 2011* of Statistik und Wissenschaft. Statistisches Bundesamt. In submission.

Münnich, R., Sachs, E. and Wagner, M. (2011). A comparison of box constraint optimization algorithms for optimal allocation in stratified random sampling. *Adv. Statist. Anal. DOI :* 10.1007/s10182-011-0176-z.

Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *J. Roy. Statist. Soc.*, **97**, 558-606.

Pfeffermann, D., Da Silva Moura, F.A. and Do Nascimento Silva, P.L. (2006). Multi-level modelling under informative sampling. *Biometrika*, **93(4),** 943-959.

Prasad, N.G.N. and Rao, J.N.K. (1990). The estimation of the mean squared error of small-area estimators. *J. Amer. Statist. Assoc.*, **85(409)**, 163-171.

Rao, J.N.K. (2003.). *Small Area Estimation* of Wiley series in survey methodology. John Wiley and Sons, New York.

Salvati, N., Tzavidis, N., Pratesi, M. and Chambers, R. (2011). Small area estimation via M-quantile geographically weighted regression. Forthcoming in *TEST*.

Särndal, C.E., Swensson, B. and Wretman, J. (2003). *Model Assisted Survey Sampling*. Springer, New York.

Schmid, T. and Münnich, R. (2012). Spatial robust small area estimation. In submission.

Sinha, S.K. and Rao, J.N.K. (2009). Robust small area estimation. *Canad. J. Statist.*, **37(3)**, 381-399.

Torabi, M. and Rao, J.N.K. (2010). Mean squared error estimators of small area means using survey weights. *Canad. J. Statist.*, **38(4)**, 598-608.

Tschuprov, A. (1923). On the mathematical expectation of the moments of frequency distributions in the case of correlated observations. *Metron*, **2**, 461-493.

Verret, F., Hidiroglou, M.A. and Rao, J.N.K. (2010). Small Area Estimation under Informative Sampling. *Proceedings of the Survey Methods Section of the SSC Annual Meeting*.

You, Y. and Rao, J.N.K. (2002.). A pseudo-empirical best linear unbiased prediction approach to smallarea estimation using survey weights. *Canad. J. Statist. / La Revue Canadienne de Statistique*, **30(3)**, 431-439.