The Impact of Sampling Designs on Small Area Estimates for Business Data

Jan Pablo Burgard[1], Ralf Münnich[1], and Thomas Zimmermann[1]

Evidence-based policy making and economic decision making rely on accurate business information on a national level and increasingly also on smaller regions and business classes. In general, traditional design-based methods suffer from low accuracy in the case of very small sample sizes in certain subgroups, whereas model-based methods, such as small area techniques, heavily rely on strong statistical models.

In small area applications in business statistics, two major issues may occur. First, in many countries business registers do not deliver strong auxiliary information for adequate model building. Second, sampling designs in business surveys are generally nonignorable and contain a large variation of survey weights.

The present study focuses on the performance of small area point and accuracy estimates of business statistics under different sampling designs. Different strategies of including sampling design information in the models are discussed. A design-based Monte Carlo simulation study unveils the impact of the variability of design weights and different levels of aggregation on model- versus design-based estimation methods. This study is based on a close to reality data set generated from Italian business data.

Key words: Nonignorable sampling designs; MSE estimation; confidence interval coverage.

1. Business Surveys and Small Area Estimation

Statistical offices increasingly face the challenge of producing estimates on subgroups in addition to national estimates. In business statistics, these subgroups may consist of regions or NACE classes (Nomenclature statistique des activités économiques dans la Communauté européenne, Eurostat 2008). Generally, in business surveys the sampling designs are optimized to furnish national estimates with a desired level of accuracy which may lead to unsuitably small sample sizes for subgroups of interest. Since the precision of direct estimates, for example measured by the variance of the estimator, is inversely proportional to the sample size, the resulting small sample sizes may lead to unreliable direct estimates for these subgroups. Hence, alternative estimators may have to be considered.

Over the past decades, small area estimation techniques have gained popularity. The main idea behind these methods is to borrow information from other subgroups via statistical models in order to increase the effective sample size of the subgroups of interest (cf. Rao 2003). One major reservation in official statistics against the use of model-based methods is the possible lack of design unbiasedness. In the presence of small sample sizes, however, the design biasedness may play a minor role in assessing the precision of the estimates because of the variability caused by small sample sizes. A widely used measure to assess the precision of estimates is the mean square error (MSE), which considers both the squared bias and the variance of an estimator. Model-based small area methods typically have lower variances but may suffer from design bias. In contrast to model-based methods, design-based methods are design unbiased at the expense of large variances with small sample sizes. Thus there is a trade-off between bias and variance of the different estimators. Therefore, the selection of an estimator of either kind has to be made carefully in any application. While small area estimation is increasingly used in many fields of social statistics, such as the estimation of poverty measures (cf. Molina and Rao 2010, or Lehtonen et al. 2011), it has not yet been widely used in the area of business statistics. Small area estimation techniques use models for the prediction of the quantity of interest. This approach relies heavily on the availability of strong predictive variables for modeling the dependent variable. This auxiliary information usually comes from business registers. The higher the predictive power of the model, the better estimates are produced.

In this article, we want to raise and discuss two issues arising in the application of small area estimation methods for business statistics. First, in many countries business registers do not include strong auxiliary information, leaving the data producer with little choice regarding model building and variable selection. Nevertheless, the data producer might be obliged to publish information on subgroups under these less suitable conditions and without sufficient sample sizes for applying design-based methods. Options available to a data producer are discussed.

Secondly, sampling designs in business statistics are often nonignorable due to a high market concentration of important variables, such as total turnover. The designs, mainly stratified or probability proportional to size, work well within a design-based framework for estimating national figures. However, most small area estimators operate in a model-based framework ignoring the sampling design. In the case of informative sampling designs, this may lead to erroneous statistical inferences (cf. Pfeffermann and Sverchkov 2009). In this case, one option is to correct for the design bias due to the informativeness directly (cf. Pfeffermann and Sverchkov 2007). Another approach incorporates the design weights into the estimation of the statistical model. In a Bayesian context, this issue has been addressed by You and Rao (2003) and Little (2012). A discussion on weighting and prediction in the context of small area estimation from a frequentist's viewpoint is given in Pfeffermann (1993) and Pfeffermann et al. (1998) and for multilevel modeling in general in Asparouhov (2006). In our article we compare different frequentist strategies for including design weights in small area modeling.

In Section 2, we describe the sampling designs used in business surveys and discuss their usefulness for small area estimation. This is followed by the presentation of the small area estimators considered in our study, including their properties with respect to complex survey designs. In Section 3, we describe our data set and outline our design-based

simulation strategy followed by a discussion of the results of our simulation study. Finally, we summarize our findings in Section 4.

2. Small Area Estimation and Modeling

2.1. Sampling Designs for Business Surveys

In business surveys, stratified sampling designs are typically applied. The strata are often determined as cross-classifications of variables such as industry classifications, geographical information or employee size classes (cf. Hidiroglou and Lavallee 2009). Since the present article focuses on enterprise-level business surveys, we omitted multistage designs which, in general, are not applied in business statistics (cf. Thompson and Oliver 2012). Some ideas in the context of small area applications for household surveys can be drawn from Münnich and Burgard (2012).

Frequently, the survey planner who designs the survey and chooses the estimator faces a conflict between obtaining reliable estimates for small domains and for national figures. Furthermore, the planner has to consider the impact of the design on the estimator as well as decide on the level of aggregation at which the estimates are required. This decision-theoretic problem may be addressed by specifying a loss function, which is to be minimized under certain constraints.

Longford (2006) minimizes the weighted sum of domain-specific variances and the variance of the national estimators subject to the sample size restriction, where the weights specify the relative importance of each domain and the priority for the national estimate. Choudhry et al. (2012) consider the problem of minimizing the total sample size subject to the upper bounds of the coefficients of variation for the strata means and the national mean by using nonlinear programming techniques. Another approach introduced by Costa et al. (2004) does not require an explicit loss function but consists of a convex combination of the equal and proportional allocation with L strata ($h = 1, \ldots, L$):

$$n_{h,Costa} = kn\frac{N_h}{N} + (1 - k)\frac{n}{L}, \ 0 \leq k \leq 1, h = 1, \ldots, L, \tag{1}$$

where n_h denote the stratum-specific sample sizes with total sample size n, N_h is the number of units in the h-th stratum summing up to the total number of units N, and k is a weighting constant, which yields the equal allocation for $k = 0$ and the proportional allocation for $k = 1$. The idea behind the Costa allocation is that the equal allocation is favorable for domain level estimates but not very efficient for national estimates, whereas the opposite holds for proportional allocation. In addition to reaching a compromise between efficient estimation at different levels of aggregation, allocation (1) is also particularly easy to apply. The optimal allocation due to Neyman (1934) and Tschuprow (1923) minimizes the variance of the national mean estimator $\hat{\mu}$ of the variable of interest Y for stratified random sampling. If we are interested in small domain estimates, however, this will not be sufficient, since the optimal allocation leads to very small domain-specific sample sizes in cases where there is hardly any variation within a stratum. This may yield stratum-specific sample sizes $n_h < 2$ which do not allow unbiased estimation of the variances. We therefore consider the box-constraint optimal allocation proposed by

Gabler et al. (2012), which minimizes the 2−norm of the relative root mean square error (RRMSE) of a set of direct statistics $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \ldots, \hat{\mu}_D)$ under constraints regarding the lower and upper bounds of the domain-specific sample sizes n_d $(d = 1, \ldots, D)$ of D domains and an upper bound of the total sample size n. The 2−norm (cf. Harville 2008, 60) can be seen as a compensatory functional penalizing larger RRMSEs more than smaller ones. The box-constraint optimal allocation technique allows for control of the sample sizes or sampling fractions and, hence, the variation of the design weights. The domain-specific sample sizes emerge as a solution of the following optimization problem:

$$\min_{n_d} \ \|\boldsymbol{RRMSE}_{<.>}(\hat{\mu})\|_2 = \sqrt[2]{\sum_{d=1}^{D} \mathrm{RRMSE}(\mu_{<d>})}$$

$$\text{s.t. } L_d \leq n_d \leq U_d, \ d = 1, \ldots, D \tag{2}$$

$$\sum_{d=1}^{D} n_d \leq n,$$

where L_d and U_d denote the lower and upper bound for the sample size in the dth domain. The issue of obtaining numerically efficient solutions for the optimization problem (2) for very large numbers of strata is explored in detail by Münnich et al. (2012).

Besides these stratified sampling designs, πps−designs are often used in business surveys as past values of the auxiliary variables are available from the enterprise register (cf. Holmberg et al. 2002). In πps sampling, the inclusion probability of each unit is proportional to the value of some size variable available at the design stage. πps sampling is a very efficient design for design-based estimation strategies in cases where a high correlation exists between the target variable and the size variable and the intercept is close to zero (cf. Tillé 2006). In fact, if the variable of interest is proportional to the size variable, the variance of a Hajék-type estimator on a national level would be zero for fixed size designs (cf. Särndal et al. 2003, 89). One issue with πps sampling is that it tends to lead to highly variable design weights when there is a large variation in the auxiliary variable X. This can negatively influence the statistical modeling. An approach to reduce this variation is to incorporate box constraints to inclusion probabilities π_i $(i = 1, \ldots, N)$ yielding new inclusion probabilities π_i^* according to

$$\min_{\pi_i^*} \ \sum_{i=1}^{N} \frac{(1/2)\left(\pi_i^* - \pi_i\right)^2}{\pi_i}$$

$$\text{s.t. } \sum_{i=1}^{N} \pi_i^* = n, \tag{3}$$

$$\pi_L \leq \pi_i^* \leq \pi_U, \ i = 1, \ldots, N,$$

where π_L and π_U denote the lower and the upper bound for the new box constraint inclusion probabilities. The solution to problem (3) gives the box-constraint inclusion probabilities π_i^* which satisfy the box constraints. In the same spirit as the box-constraint

optimal allocation, the box-constraint πps design allows for control of the range of the design weights directly. As an additional benefit, the box-constraint approach towards πps sampling avoids very small inclusion probabilities, which are a concern for the sample selection algorithms. Another method has been proposed by Falorsi and Righi (2008) whose strategy may be described as a balanced sampling multiway stratification. They consider a situation in which constraints regarding a multivariate response **y** and several partitions hold whilst at the same time the selected sample is balanced on auxiliary variables. Since current algorithms for drawing balanced samples from large universes are still extremely computer intensive, we omitted the approach of Falorsi and Righi (2008) from our simulation study.

2.2. *Small Area Estimators under Complex Designs*

A common aim in small area estimation is the estimation of the domain mean

$$\mu_d = \frac{1}{N_d} \sum_{j=1}^{N_d} y_{dj}, \; d = 1, \ldots, D, \tag{4}$$

where y_{dj} is the variable of interest for unit j in domain d and N_d denotes the population size in domain d. A traditional estimator often used in survey sampling is the direct estimator given by

$$\hat{\mu}_{d,Direct} = \frac{\displaystyle\sum_{j=1}^{n_d} w_{dj} y_{dj}}{\displaystyle\sum_{j=1}^{n_d} w_{dj}}, \tag{5}$$

with w_{dj} as the design weight of unit j in domain d. Note that with planned domains and stratified random sampling where the strata are nested within the domains, the sum in the denominator of (5) is equal to N_d. Though Estimator (5) is design unbiased, estimates for domains with small sample sizes are expected to be inaccurate. We refer to (5) as the Direct estimator. The group of GREG estimators are given by

$$\hat{\mu}_{d,GREG} = \frac{1}{N_d} \left[\sum_{j=1}^{N_d} \hat{y}_{dj} + \sum_{j=1}^{n_d} w_{dj} (y_{dj} - \hat{y}_{dj}) \right]. \tag{6}$$

where $\hat{y}_{dj}$ is the predicted value of the variable of interest for unit j in domain d under a specified regression model. Thus the domain estimate in (6) results as the mean of the predicted values for all population units in domain d plus the mean of the weighted residuals for the sampled units in domain d. There are various choices for the assisting model, such as using linear or possibly nonlinear models, considering mixed models or focusing on fixed effects, and including or omitting design weights when fitting the model. A thorough investigation of model choice for GREG estimators is given in Lehtonen et al. (2003, 2005). In our study, we will focus on a linear fixed effects model and refer to this estimator as the GREG estimator. A detailed account on design-based and model-assisted domain estimation is given by Lehtonen and Veijanen (2009).

The unit-level mixed model, which is also known as the nested error regression model, is given by

$$y_{dj} = \boldsymbol{x}_{dj}^T \boldsymbol{\beta} + u_d + \varepsilon_{dj}, \ \ d = 1, \ldots, D, j = 1, \ldots, N_d, \tag{7}$$

where $u_d \overset{iid}{\sim} N(0, \sigma_u^2), \varepsilon_{dj} \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$. The domain-specific effects u_d are independent of the sampling error ε_{dj}. $\boldsymbol{x}_{dj}$ is the vector of auxiliary information for unit j in domain d, and $\boldsymbol{\beta}$ the vector of fixed regression parameters. Under Model (7) the small area mean is given by $\mu_d = \overline{\boldsymbol{X}}_d^T \boldsymbol{\beta} + u_d$ for all domains $d = 1, \ldots, D$. $\overline{\boldsymbol{X}}_d$ is the vector of the population mean of the auxiliary information in domain d and $\bar{\boldsymbol{x}}_d$ refers to the sample equivalent.

Assuming that Model (7) holds for the sample as well, the following EBLUP (empirical best linear-unbiased predictor) under negligible sampling fractions for the unknown domain mean μ_d can be derived as (cf. Battese et al. 1988)

$$\hat{\mu}_{d,BHF} = \overline{\boldsymbol{X}}_d^T \hat{\boldsymbol{\beta}} + \hat{u}_d, \quad \hat{u}_d = \hat{\gamma}_d (\bar{y}_d - \bar{\boldsymbol{x}}_d^T \hat{\boldsymbol{\beta}}) \qquad \hat{\gamma}_d = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + (\hat{\sigma}_\varepsilon^2 / n_d)}$$

$$\hat{\boldsymbol{\beta}} = \left(\sum_{d=1}^D \boldsymbol{x}_d^T \hat{\boldsymbol{V}}_d^{-1} \boldsymbol{x}_d \right)^{-1} \left(\sum_{d=1}^D \boldsymbol{x}_d^T \hat{\boldsymbol{V}}_d^{-1} \boldsymbol{y}_d \right)$$

$$\tag{8}$$

In Equation (8), $\hat{u}_d$ is the EBLUP of the random effect u_d, $\hat{\gamma}_d$ is the shrinkage factor depending on the estimated variance components ($\hat{\sigma}_u^2$ and $\hat{\sigma}_\varepsilon^2$), and $\hat{\boldsymbol{\beta}}$ is an estimator for $\boldsymbol{\beta}$, $\bar{y}_d = n_d^{-1} \sum_{j=1}^{n_d} y_{dj}$, $\hat{\boldsymbol{V}}^{-1}$ refers to the inverse of the variance-covariance matrix. $\hat{\boldsymbol{V}}_d = \hat{\boldsymbol{V}}_d(\hat{\sigma}_u^2, \hat{\sigma}_\varepsilon^2)$ in domain d (cf. Rao 2003, Sec. 7.2). While Estimator (8) is model unbiased and efficient for self-weighting sampling designs, this is unlikely to hold for general sampling designs. In the following, we will denote the Estimator (8) as BHF for notational convenience since it dates back to Battese et al. (1988).

In typical applications in official statistics, ignoring the design weights may have severe consequences for the quality of model-based estimators (cf. Münnich and Burgard 2012). Several extensions of mixed models to cope with nonignorable sampling designs have been proposed, for example in Pfeffermann et al. (1998), Asparouhov (2006), Rabe-Hesketh and Skrondal (2006) and Lehtonen et al. (2006). Here, we focus on selected approaches which are suitable and easily applicable in official statistics.

A second way of extending the unweighted EBLUP under the unit-level mixed model is by augmenting the design matrix by the design weights. The following model is fitted to the survey data:

$$y_{dj} = \boldsymbol{x}_{dj}^T \boldsymbol{\beta} + \kappa w_{dj} + u_d + \varepsilon_{dj}, \ \ d = 1, \ldots, D, \ j = 1, \ldots, n_d, \tag{9}$$

where κ is the additional regression coefficient for the impact of the weights on the variable of interest, estimated by $\hat{\kappa}$. Alternatively, the size variable might also be used instead of w_{dj} under unequal probability sampling. The EBLUP under Model (9) is obtained as

$$\hat{\mu}_{d,augBHF} = \overline{\boldsymbol{X}}_d^T \hat{\boldsymbol{\beta}} + \hat{\kappa} \overline{W}_d + \hat{u}_d, \tag{10}$$

with $\overline{W}_d = N_d^{-1} \sum_{j=1}^{N_d} w_{dj}$ as the population mean of the design weights in domain d. This estimator was introduced by Verret et al. (2010) in the context of informative

sampling and will be referred to as the augBHF estimator. Note that we could alternatively estimate $\boldsymbol{\beta}$ and κ using design weights.

You and Rao (2002) propose to transform the unit-level model (7) to a survey-weighted domain-level model with normalized weights within the domains. This model is given by

$$\bar{y}_{dw} = \bar{\boldsymbol{x}}_{dw}^T \boldsymbol{\beta}_w + u_d + \bar{\varepsilon}_{dw}, \ d = 1, \ldots, D, \tag{11}$$

with

$$\bar{y}_{dw} = \sum_{j=1}^{n_d} \tilde{w}_{dj} y_{dj}, \bar{\boldsymbol{x}}_{dw} = \sum_{j=1}^{n_d} \tilde{w}_{dj} \boldsymbol{x}_{dw}, \bar{\varepsilon}_{dw} = \sum_{j=1}^{n_d} \tilde{w}_{dj} \varepsilon_{dj} \ \text{and} \ \tilde{w}_{dj} = \frac{w_{dj}}{\sum_{j=1}^{n_d} w_{dj}}.$$

The pseudo-EBLUP under Model (11) follows as (cf. You and Rao 2002):

$$\hat{\mu}_{d,YR} = \hat{\gamma}_{dw} \bar{y}_{dw} + \left(\bar{\boldsymbol{X}}_d - \hat{\gamma}_{dw} \bar{\boldsymbol{x}}_{dw} \right)^T \hat{\boldsymbol{\beta}}_w, \text{with}$$

$$\hat{\gamma}_{dw} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \delta_d^2 \hat{\sigma}_\varepsilon^2}, \tag{12}$$

$$\delta_d^2 = \sum_{j=1}^{n_d} \tilde{w}_{dj}^2, \ \text{and}$$

$$\hat{\boldsymbol{\beta}}_w = \left(\sum_d^D \sum_{j=1}^{n_d} w_{dj} \boldsymbol{x}_{dj} (\boldsymbol{x}_{dj} - \hat{\gamma}_{dw} \bar{\boldsymbol{x}}_{dw})^T \right)^{-1} \left(\sum_d^D \sum_{j=1}^{n_d} w_{dj} y_{dj} (\boldsymbol{x}_{dj} - \hat{\gamma}_{dw} \bar{\boldsymbol{x}}_{dw}) \right)$$

In addition to achieving design consistency, the estimator given by (12) also fulfils the benchmarking property with respect to the national estimate. The Estimator (12) is denoted by YouRao.

In earlier simulation studies, the approach employed by Lehtonen et al. (2011) gave good results for various sampling designs. It is based on incorporating the vector of design weights in the lmer function in the R-package lme4 (cf. Bates et al. 2011; lmer provides a fast mixed-effects model implementation). Despite the fact that it is not meant for including design weights specifically, it has been shown to reduce the bias of the unweighted estimator (8) in many cases. For details regarding the estimation of the model parameters we refer to Bates (2011). This estimator is denoted by wBHF.

In some cases where unit-level data may not be available or the computation of unit-level models may not be feasible, area-level models can be a remedy. An area-level model may be described as follows:

$$\widehat{\bar{y}}_d = \bar{X}_d^T \boldsymbol{\beta} + u_d + \varepsilon_d, \ d = 1, \ldots, D, \tag{13}$$

where $u_d \overset{\text{iid}}{\sim} N(0, \sigma_u^2)$ and $\varepsilon_d \overset{\text{ind}}{\sim} N(0, \sigma_{\varepsilon_d}^2)$, which is independent of u_d (cf. Jiang and Lahiri 2006). Note that in the area-level model (13) the small area means of the direct estimator $\hat{\bar{y}}_d$ are modeled but not the observations themselves. This is due to the fact that auxiliary information is available at the domain level only. In the area-level

literature, σ_u^2 is also referred to as the model variance and $\sigma_{\varepsilon_d}^2$ as the sampling variance of the direct estimator, which depends on the domain-specific sample sizes and is therefore not identically distributed between the domains. The EBLUP under the area-level model (13) is given by

$$\hat{\mu}_{d,FH} = \bar{X}_d^T \hat{\boldsymbol{\beta}}_{FH} + \hat{u}_d \qquad (14)$$

and we will refer to the estimator as FH, since it was introduced by Fay and Herriot (1979). $\hat{\boldsymbol{\beta}}_{FH}$ refers to the estimator of the regression parameters under Model (13) and is given in Rao (2003, 116).

To estimate the prediction mean square error (PMSE) of the aforementioned EBLUP-type estimators BHF, wBHF, augBHF, YouRao and FH we consider two different strategies: one based on Taylor series expansions and the other based on the parametric bootstrap method. A good reference on these methods is Datta (2009). Prasad and Rao (1990) derived the following PMSE decomposition for EBLUP estimators based on results from Kackar and Harville (1984):

$$\text{PMSE}(\hat{\mu}_d(\hat{\theta})) = g_{1d}(\hat{\theta}) + g_{2d}(\hat{\theta}) + 2g_{3d}(\hat{\theta}), \qquad (15)$$

where the terms g_{1d} to g_{3d} depend on the estimated variance components $\hat{\theta}$. Additionally, in the case that the variance components are estimated by Restricted Maximum Likelihood (REML) or Maximum Likelihood (ML), explicit formulae for Estimators (8) and (14) based on decomposition (15) are given in Datta and Lahiri (2000). A second-order correct PMSE estimator for (12) has been derived by Torabi and Rao (2010).

Butar and Lahiri (2003) proposed using parametric bootstrap methods to estimate the PMSE of small area estimators. To account for the finite population, we consider a simplification of the bootstrap proposed by González-Manteiga et al. (2008) to produce PMSE estimates. Their algorithm for computing PMSE estimates for Estimator (8) is as follows:

1. Fit the statistical model to the sample data to obtain the estimates $\hat{\boldsymbol{\beta}}$, $\hat{\sigma}_u^2$ and $\hat{\sigma}_\varepsilon^2$.
2. Construct replicates $y_{dj}^* = \boldsymbol{x}_{dj}^T \hat{\boldsymbol{\beta}} + u_d^* + \varepsilon_{dj}^*$, where $u_d^* \overset{iid}{\sim} N(0, \hat{\sigma}_u^2)$ and $\varepsilon_{dj}^* \overset{iid}{\sim} N(0, \hat{\sigma}_\varepsilon^2)$.
3. Calculate the domain means $\mu_d^* = (1/N_d)\sum_{j=1}^{N_d} y_{dj}^*$.
4. Fit the statistical model to the sampled elements of y_{dj}^* to obtain estimates $\hat{\boldsymbol{\beta}}^*$ and $\hat{u}_d^*$.
5. Compute the estimated domain means $\hat{\mu}_d^* = \bar{X}_d^T \hat{\boldsymbol{\beta}}^* + \hat{u}_d^*$.
6. Repeat the Steps 2 to 5 B times.
7. The estimated PMSE is computed by $\widehat{\text{PMSE}}(\hat{\mu}_{d,PB}) = \frac{1}{B}\sum_{b=1}^{B} \left(\hat{\mu}_d^{*(b)} - \mu_d^{*(b)} \right)^2$.

We also used the parametric bootstrap to obtain PMSE estimates for estimators wBHF, augBHF and FH, using the above mentioned models and formulae for estimating the model parameters and computing the estimated domain means.

3. Simulation Study

3.1. Data Set and Sampling Design

Our design-based simulation study extends the work of Burgard et al. (2012) to cover the issues of PMSE estimation and prediction intervals for small domains. The study is based

on synthetic business data resembling the small and medium enterprises from the Italian business register. This data set is a precursor of the fully synthetic data set TRItalia, which is being produced within the BLUE-ETS project (see Kolb et al. 2013). The parameter of interest is the mean of value added. As auxiliary variables we use turnover and the number of employees. Both variables are available in the Italian business register. We use these auxiliary variables since they are the only noncategorical register variables available at the design stage. From a subject-matter viewpoint, both auxiliary variables may influence the value added. A linear regression model without random effects confirmed that both explanatory variables are highly significant. However, the model only yielded $R^2 = 0.0045$ for our population, indicating that the explanatory power of the model is poor. Even if this case is pessimistic, similar situations may occur in many countries where registers often lack strong covariates. Even if the variable of interest is skewed, the application of transformation methods requires further research on the inclusion of weights, which is beyond the scope of this article.

As a stratification variable we used the first digit of the industry classification within each province (103 Italian provinces), resulting in 927 strata. The stratum-specific population sizes vary from 98 enterprises in the smallest stratum to 114,844 enterprises in the largest stratum. Since our data set is restricted to small and medium enterprises with 1 to 99 employees, our stratification does not contain a census-like stratum where all units within the stratum are sampled with certainty. We account for the problem that statistical agencies have to disseminate information at different levels of aggregation by considering two kinds of domains as scenarios. In the first scenario, the 103 Italian provinces are also the domains of interest, whereas in the second scenario the 927 strata are considered as domains. It is important to note that both scenarios reflect the problem of prediction with planned domains, thus avoiding problems of nonsampled areas.

The expected total sample size is set to $n = 60,000$. For the (box-constraint) optimal allocations the auxiliary variable turnover was used to compute the stratum-specific sample sizes. Besides these stratified sampling designs we also consider unequal probability within the strata, where the expected sample size within each stratum is set to the sample size allocated by proportional allocation. As in the case of optimal allocation, we use turnover to compute the inclusion probabilities. We use turnover as a size measure because it is the variable in our data set which has the highest correlation with our dependent variable. Since turnover does not have zero values in our data set, its use as a size measure is straightforward. A major difference between optimal allocation and unequal probability designs is that the former leads to design weights which vary between the strata, whereas for the latter the weights also vary within the strata. Even though the computation of inclusion probabilities is straightforward when using a strictly positive auxiliary variable, the sample selection is very computer intensive. For the case of unconstrained inclusion probabilities, Midzuno's method as described in Tillé (2006, Algorithm 6.13) programmed in C failed to produce the desired samples in due time. This problem was resolved by means of the box-constraint inclusion probabilities given in (3). In accordance with Münnich and Burgard (2012) the Gelman factor (GF) is defined as the ratio of the largest to the smallest design weight. This definition of a Gelman factor should not be confused with the Gelman-Rubin factor which is related to

Table 1.   Sampling designs

Abbreviation	Design	Gelman factor
COSTA$_{50}$	Costa-type allocation with $k = 0.5$	47.66
EQUAL	equal allocation	1,153.85
BCOpt$_{25}$	box-constraint optimal allocation with $GF = 25$	30.88
BCOpt$_{50}$	box-constraint optimal allocation with $GF = 50$	60.83
OPT	optimal allocation	554.92
PROP	proportional allocation	1.78
UPS	unequal probability sampling	44,085, 380.58
UPS$_{10}$	unequal probability sampling under constraint $\max \frac{\pi_L^{-1}}{\pi_U^{-1}} \leq 10$	10
UPS$_{100}$	unequal probability sampling under constraint $\max \frac{\pi_L^{-1}}{\pi_U^{-1}} \leq 100$	100

MCMC convergence diagnostics. The GF is given by

$$\text{GF} = \frac{\max\limits_{i=1,\ldots,N} \frac{1}{\pi_i}}{\min\limits_{i=1,\ldots,N} \frac{1}{\pi_i}}. \tag{16}$$

The GF for equal allocation thus varies with the stratum sizes. If all the strata are of roughly the same size, then the equal allocation is almost equivalent to the proportional
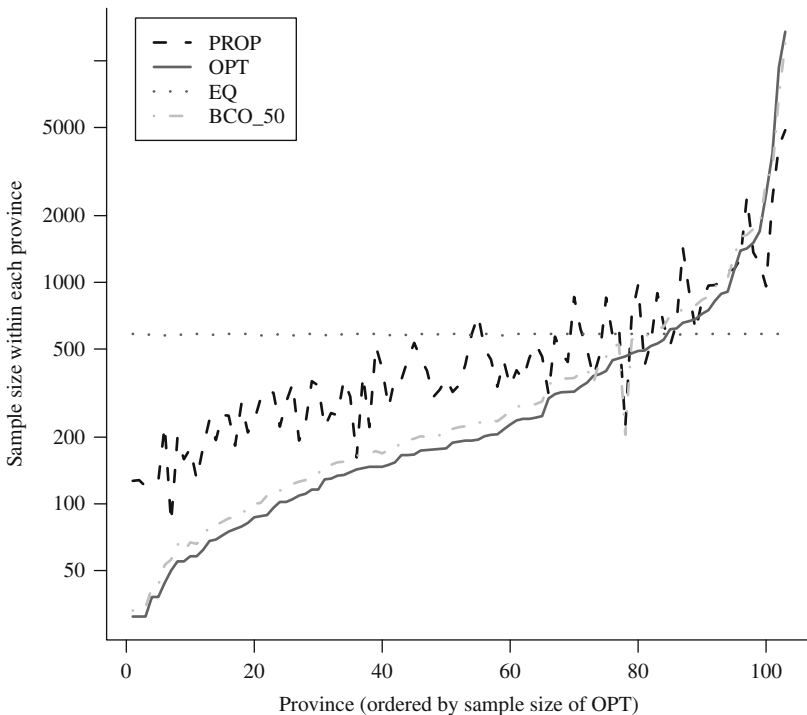

Fig. 1.   Domain-specific sample sizes – 1st Scenario (103 domains)

Table 2. Domain-specific sample sizes – 1st Scenario (103 domains)

	PROP	EQ	COSTA$_{50}$	BCOpt$_{25}$	BCOpt$_{50}$	OPT
Min	82	576	330	35	33	31
Max	4,862	585	2,727	11,765	11,978	13,522

allocation and thus the GF will approach 1. If the N_h are highly variable, the equal allocation leads to highly dispersed design weights. Additionally, in the case of πps designs we get GF $\doteq \max z_i / \min z_i$ with z being the size variable used for the calculation of the πps inclusion probabilities. Typically, the variation of the auxiliary variable in business surveys is very large and thus the GF is very large as well. Table 1 lists our sampling designs with the abbreviations used in the following and the GF. Turning our attention to the box-constraint optimal allocations, we recognize that these allocations do not satisfy the box constraints exactly. This is due to the fact that our approach produces non-integer-valued numbers, which have to be rounded, rather than integer-valued constraints.

The variation of the domain-specific sample sizes under the first scenario is illustrated in Figure 1. We see that the optimal allocation on the one hand and the equal allocation on the other hand are the two extreme cases. The minimum and the maximum for the domain-specific sample sizes under Scenarios 1 and 2 are given in Tables 2 and 3. These tables further illustrate that under Scenario 2 the minimum domain-specific sample sizes are very small except for the equal allocation.

To evaluate the results of our simulation study, we consider several different quality measures related to the accuracy of point estimates and the reliability of confidence intervals. A common measure to estimate the bias of a point estimator is the relative bias. It is given by

$$\text{RB}(\hat{\mu}_d) = \frac{\left(\frac{1}{R}\right)\sum_{l=1}^{R}\hat{\mu}_{l,d} - \mu_d}{\mu_d}, \; d = 1, \ldots, D, \tag{17}$$

where R denotes the number of Monte Carlo replicates. The relative bias takes values from $-\infty$ to ∞, whilst a relative bias close to 0 is desirable, indicating that the point estimates are on average identical to the true values. Another quality measure is the relative root mean square error (RRMSE), which is computed as

$$\text{RRMSE}(\hat{\mu}_d) = \frac{\sqrt{\frac{1}{R}\sum_{l=1}^{R}(\hat{\mu}_{l,d} - \mu_d)^2}}{\mu_d}, \; d = 1, \ldots, D. \tag{18}$$

Table 3. Domain-specific sample sizes – 2nd Scenario (927 domains)

	PROP	EQ	COSTA$_{50}$	BCOpt$_{25}$	BCOpt$_{50}$	OPT
Min	2	64	34	2	2	2
Max	1,605	65	836	3,998	3,935	4,766

Table 4.   Computing times in seconds

Estimator	BHF	wBHF	augBHF	YouRao	GREG	Direct
Seconds	2732.17	2870.49	2789.30	3.83	0.40	0.21

The values of the RRMSE are in the range between 0 and ∞, where a value close to 0 indicates good results. Moreover, we consider summary statistics of the quality measures over all domains. With respect to the relative bias, we compute the mean absolute relative bias (MARB)

$$\text{MARB}(\hat{\mu}_d) = \frac{1}{D}\sum_{d=1}^{D}|\text{RB}(\hat{\mu}_d)| \tag{19}$$

and for the RRMSE we consider the average relative root mean square error (ARRMSE)

$$\text{ARRMSE}(\hat{\mu}_d) = \frac{1}{D}\sum_{d=1}^{D}\text{RRMSE}(\hat{\mu}_d). \tag{20}$$

We construct confidence intervals based on MSE or PMSE estimators as described in Subsection 2.2. The traditional approach is to compute the confidence interval (CI) as follows (cf. Chatterjee et al. 2008):

$$\text{CI}(\hat{\mu}_d)_{1-\alpha} = \left[\hat{\mu}_d - \sqrt{\widehat{\text{PMSE}}(\hat{\mu}_d)}\cdot z_{1-\alpha/2}; \hat{\mu}_d + \sqrt{\widehat{\text{PMSE}}(\hat{\mu}_d)}\cdot z_{1-\alpha/2}\right] \tag{21}$$

with $z_{1-\alpha/2}$ as the $(1-\alpha/2)$-quantile of the standard normal distribution. Since $\sqrt{\widehat{\text{PMSE}}(\hat{\mu}_d)}$ is estimated, confidence intervals based on quantiles of the t-distribution with $(n_d - 1)$ degrees of freedom could be considered. Note that differences between these two approaches to computing confidence intervals vanish as the domain-specific sample size n_d increases. Additionally, we also considered using bootstrap confidence intervals as proposed by Chatterjee et al. (2008). The reliability of confidence intervals is measured by the coverage rate, computed as the percentage of confidence intervals covering the true value μ_d.

In the following section we will report results based on 1,000 Monte Carlo replications. For the parametric bootstrap methods we use 499 bootstrap replications. Due to the small number of bootstrap replications, the bootstrap confidence intervals are outperformed

Table 5.   Types of estimators

Abbreviation	Estimator
Direct	Hajek-type estimator (5)
GREG	linear fixed effects generalized regression estimator (6)
YouRao	pseudo-EBLUP (12)
wBHF	weighted EBLUP using weights option in lmer
augBHF	augmented EBLUP (10)
BHF	EBLUP (8) under unit-level mixed model
FH	EBLUP (14) under area-level mixed model

Table 6.   MARB – 1st Scenario (103 domains)

	PROP	EQ	UPS_{10}	UPS_{100}	$COSTA_{50}$	$BCOpt_{25}$	$BCOpt_{50}$	OPT
Direct	0.004	0.004	0.005	0.006	0.003	0.005	0.005	0.007
GREG	0.004	0.004	0.027	0.045	0.004	0.005	0.006	0.007
YouRao	0.023	0.026	0.034	0.069	0.020	0.059	0.066	0.089
wBHF	0.022	0.014	0.019	0.014	0.021	0.012	0.010	0.009
augBHF	0.022	0.023	0.226	0.532	0.021	0.019	0.018	0.019
BHF	0.022	0.023	0.226	0.533	0.020	0.040	0.041	0.040
FH	0.051	0.059	0.069	0.090	0.046	0.100	0.106	0.115

by the other methods. Therefore, the bootstrap confidence intervals are not presented in Subsection 3.3. The average CPU time (AMD Opteron 6164 HE with 1.7 GHz and 4GB RAM for each kernel) for the estimators is given in Table 4, where in case of the BHF, wBHF, and augBHF 499 bootstrap resamples are performed.

3.2.   Results of Point Estimates

In this section we summarize the most important aspects regarding the simulation results on the point estimates. For convenience, our estimators are listed in Table 5. We also considered a GREG based on mixed models, but we did not observe major differences between a GREG with or without random effects. To keep the presentation of the results as short as possible, the focus subsequently lies on the GREG without random effects.

For our first scenario (103 domains) the mean absolute relative bias over all domains is given in Table 6. The analysis of MARB in Table 6 indicates that the Direct estimator is indeed unbiased under all sampling designs. The model-assisted GREG has some problems under unequal probability designs. This can be traced back to the fact that we did not include design weights when estimating β. With respect to the estimators based on unit-level models, we see that there are only minor differences under proportional allocation, equal allocation and convex combinations thereof. As soon as we consider (box-constraint) optimal allocations, the bias of the unweighted BHF estimator is more pronounced compared to the wBHF and augBHF estimator. With respect to unequal probability designs, we observe severe biases for the augBHF and BHF estimators, whereas the wBHF estimator is still accurate. The YouRao estimator performs similarly to the wBHF under proportional and equal allocation, but its bias increases for other designs and is higher than the bias of the unweighted BHF estimator for optimal

Table 7.   ARRMSE – 1st Scenario (103 domains)

	PROP	EQ	UPS_{10}	UPS_{100}	$COSTA_{50}$	$BCOpt_{25}$	$BCOpt_{50}$	OPT
Direct	0.142	0.160	0.162	0.193	0.135	0.220	0.234	0.256
GREG	0.140	0.158	0.167	0.205	0.134	0.219	0.233	0.255
YouRao	0.139	0.156	0.160	0.191	0.133	0.211	0.224	0.245
wBHF	0.029	0.071	0.042	0.074	0.033	0.103	0.123	0.155
augBHF	0.030	0.033	0.229	0.539	0.029	0.031	0.032	0.034
BHF	0.029	0.032	0.229	0.539	0.030	0.044	0.046	0.046
FH	0.122	0.136	0.136	0.153	0.119	0.166	0.172	0.182

Table 8.   MARB – 2nd Scenario (927 domains)

	PROP	EQ	UPS_{10}	UPS_{100}	$COSTA_{50}$	$BCOpt_{25}$	$BCOpt_{50}$	OPT
Direct	0.015	0.008	0.017	0.024	0.010	0.022	0.023	0.026
GREG	0.015	0.009	0.016	0.020	0.010	0.022	0.023	0.026
YouRao	0.247	0.107	0.294	0.439	0.131	0.391	0.409	0.457
wBHF	0.081	0.068	0.073	0.060	0.078	0.047	0.042	0.037
augBHF	0.080	0.075	0.251	0.553	0.074	0.068	0.071	0.074
BHF	0.081	0.076	0.251	0.553	0.078	0.094	0.096	0.096
FH	0.221	0.127	0.257	0.304	0.152	0.314	0.318	0.322

allocations. The FH estimator exhibits bias under all designs considered due to full shrinkage towards the synthetic component.

As soon as we consider the ARRMSE given in Table 7, the picture is completely different. We observe that there is no single best estimator under all designs and hence sampling design plays an important role. The results under proportional and Costa-type allocations are almost the same. Under these designs, all model-based unit-level estimators work well. Even though the area-level FH estimator is biased, it has the lower ARRMSE compared to design-based estimators and the YouRao estimator. Under equal allocation, the results are similar to the proportional and Costa-type allocations except for the weighted wBHF estimator, which has a considerably higher ARRMSE than the other unit-level estimators. For designs based on (box-constraint) optimal allocations, the augmented augBHF estimator performs best, with the unweighted BHF estimator as the only other estimator with an ARRMSE under ten percent. The wBHF estimator suffers from a much higher ARRMSE despite a lower bias, a result that can be attributed to the increase of the variability of the model parameter estimates. Moreover, under unequal probability designs, the wBHF estimator is the only reasonable estimator in terms of an ARRMSE under ten percent. The performance of the augBHF and BHF is identical up to the third decimal number, indicating that augmenting the design matrix does not increase the precision under unequal probability sampling in this setting. A closer look at Tables 6 and 7 reveals that due to the shrinkage to the synthetic component the FH estimator is biased under all designs but does not perform badly with respect to RRMSE. The comparison between the Direct estimator and the GREG shows that the working model does not have much predictive power. Concentrating on the Direct estimator, we note that designs optimized for national-level estimation are not the best choice for domain estimation.

Table 9.   ARRMSE – 2nd Scenario (927 domains)

	PROP	EQ	UPS_{10}	UPS_{100}	$COSTA_{50}$	$BCOpt_{25}$	$BCOpt_{50}$	OPT
Direct	0.575	0.333	0.650	0.771	0.374	0.889	0.940	1.006
GREG	0.569	0.329	0.649	0.769	0.370	0.885	0.935	1.002
YouRao	0.451	0.312	0.498	0.573	0.344	0.570	0.581	0.608
wBHF	0.085	0.114	0.103	0.196	0.088	0.344	0.441	0.556
augBHF	0.091	0.082	0.256	0.561	0.080	0.080	0.085	0.089
BHF	0.085	0.082	0.256	0.561	0.083	0.099	0.101	0.102
FH	0.350	0.281	0.385	0.412	0.302	0.420	0.425	0.433

The MARB in the presence of smaller domains (Scenario 2, 927 domains) are given in Table 8. Compared to Scenario 1 (Table 6), the biases increase (almost) uniformly which is due to the smaller sample sizes. We see that for all designs the GREG and Direct estimators have the lowest relative bias. With respect to the unit-level estimators the impact of designs is similar to the first scenario, but with generally higher absolute relative biases. The FH suffers most from severe bias, especially under designs with largely varying sample sizes such as (box-constraint) optimal allocations or proportional allocation and the unequal probability designs. The most striking aspect about these results is the large bias of the YouRao estimator.

With regards to the ARRMSE, which is shown in Table 9, we see that the ARRMSE increases drastically compared to Scenario 1. Unlike Scenario 1, we now observe significant differences between Costa-type allocation and proportional allocation, which is due to the fact that the proportional allocation leads to very small domain-specific sample sizes in Scenario 2 (see Table 3). This causes a severe loss in estimation quality in comparison to the Costa-type allocation for design-based estimators. This does not apply for model-based estimators, which manage to borrow strength from other domains to
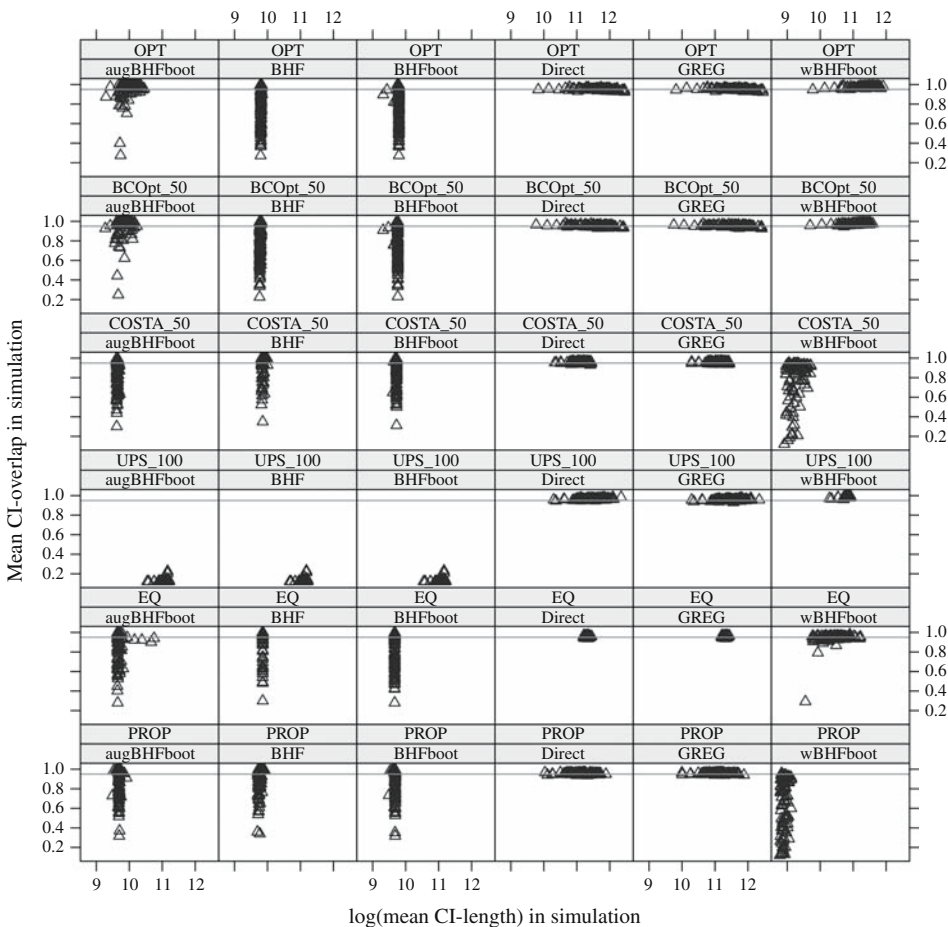

Fig. 2.   Coverage Rates – 1st Scenario (103 domains)

Table 10.   Mean of coverage rates – 1st Scenario (103 domains)

	PROP	EQ	UPS_{100}	$COSTA_{50}$	$BCOpt_{50}$	OPT
augBHFboot	0.867	0.845	0.127	0.850	0.937	0.944
BHF	0.898	0.927	0.128	0.933	0.748	0.765
BHFboot	0.861	0.831	0.127	0.868	0.738	0.762
Direct	0.955	0.954	0.965	0.956	0.951	0.949
GREG	0.955	0.955	0.956	0.956	0.951	0.949
wBHFboot	0.670	0.939	0.987	0.781	0.973	0.973

compensate for small domain-specific sample sizes and perform best under both designs. With respect to the equal allocation the ranking of the estimators is similar, except that the weighted wBHF estimator performs worse than the other two unit-level estimators. Under (box-constraint) optimal allocations, the augmented augBHF estimator performs slightly better than the BHF estimator due to the lower bias. Other estimators cannot be recommended in this case, since their ARRMSE is 30 percent and more. Under unequal probability designs, the weighted wBHF estimator seems the only reasonable choice, even though its ARRMSE is close to 20 percent when the Gelman factors are constrained to 100.

3.3.   Results of Precision Estimates

In this section, we report results of precision estimates for the most interesting designs and estimators only. The coverage rates and mean confidence interval lengths for the first scenario are illustrated in Figure 2 and means of the coverage rates are given in Table 10. Figure 2 depicts both the confidence interval coverage rates and mean confidence interval length for each domain. Ideally, these points would lie on the horizontal line, indicating a 95 % coverage rate, and at the left side in each panel, demonstrating high accuracy of the point estimates by a shorter average length of confidence intervals. It is obvious that some small area estimators yield lower coverages, which is mainly caused by a worse fit of the statistical model in several areas. This reflects the situation of many registers only containing a limited set of potentially predictive covariates.

Focusing on the length of the confidence intervals, we see that equal allocation is the best choice if one wishes to use design-based estimators. Whereas the coverage rates of the Direct and GREG estimators are reasonable under all the sampling designs considered, this does not apply for the other estimators. Under the UPS_{100} design, most model-based estimation methods suffer from severe undercoverage. With respect to the two approaches to PMSE estimation, either by Taylor approximation or by parametric bootstrap, we hardly observe any differences in the case of the unweighted BHF estimator. Interestingly, the augBHF does not perform badly in the case of (box-constraint) optimal allocations, even though it does not achieve the nominal coverage rate on average in any design. The mean coverage rates indicate overcoverage for the wBHF using parametric bootstrap under (box-constraint) optimal allocations, which clearly shows that the confidence intervals are not efficient. Altogether, it is indisputable that the coverage rates of the model-based estimators are not satisfactory.

The coverage rates for the second scenario are depicted in Figure 3. In addition to the dark "+" signs related to the computation of the confidence intervals based on quantiles of
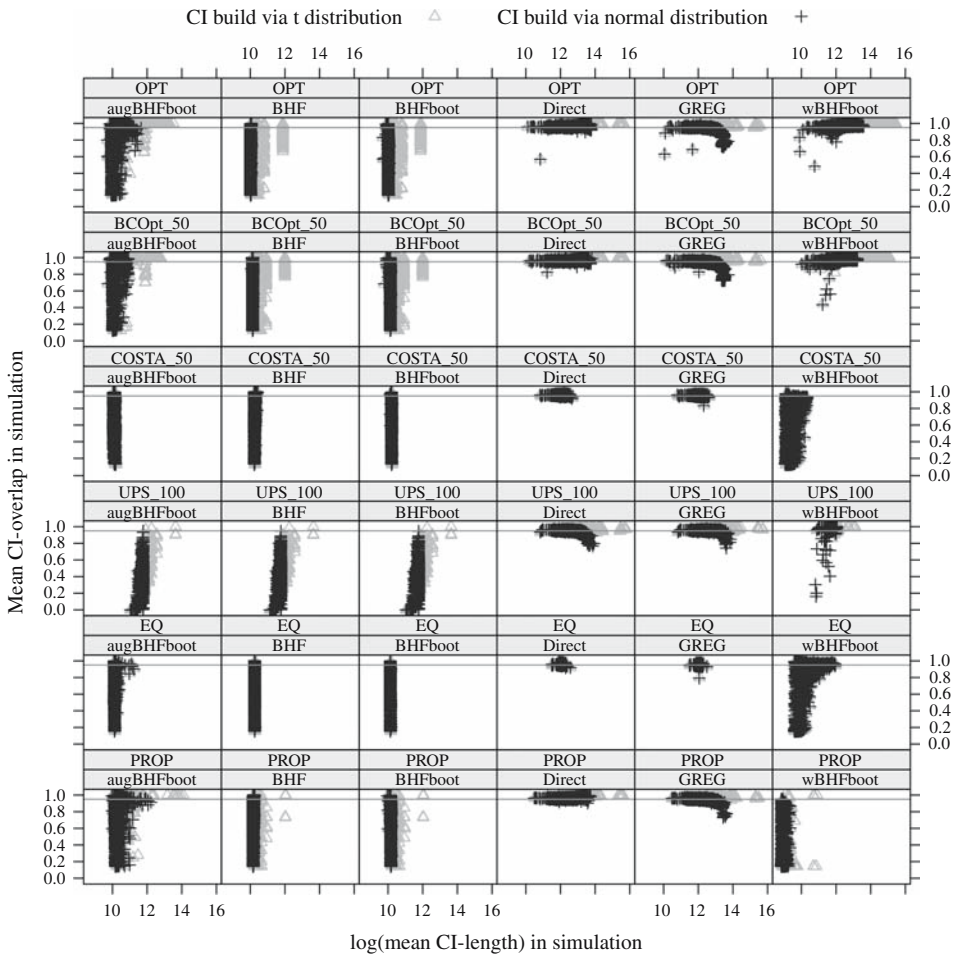
Fig. 3.   *Coverage Rates – 2nd Scenario (927 domains)*

the normal distribution, lighter triangles indicate confidence intervals based on quantiles of t-distributions with $(n_d - 1)$ degrees of freedom as explained at the end of Subsection 3.1. These two methods differ only in the presence of very small domain-specific sample sizes n_d, which was not a concern in Scenario 1. Looking at the x-scale, we observe that the CI-length increases dramatically compared to Scenario 1. Furthermore, in the case of the Direct estimator, we observe some problems under UPS_{100} for the CIs built via the normal distribution. These problems vanish as soon as we use the t-distribution, which seems to be the better choice for very small domains. For the GREG estimator the use of normal-quantiles is critical except under equal and Costa-type allocations. With respect to the model-based estimators the poor performance of all strategies is striking, as can also be seen from Tables 11 and 12.

4. Summary

This article explores two major issues official statistics face when implementing small area estimation techniques in business surveys. First, business registers of many countries

Table 11. Mean of coverage rates – 2nd Scenario (927 domains) – Normal Quantiles

	PROP	EQ	UPS_{100}	$COSTA_{50}$	$BCOpt_{50}$	OPT
augBHFboot	0.717	0.619	0.386	0.642	0.745	0.726
BHF	0.686	0.687	0.389	0.699	0.638	0.588
BHFboot	0.672	0.612	0.386	0.663	0.634	0.585
Direct	0.962	0.960	0.941	0.959	0.966	0.967
GREG	0.943	0.954	0.946	0.952	0.904	0.895
wBHFboot	0.349	0.735	0.984	0.476	0.968	0.965

do not yield many variables with strong predictive power. Second, the sampling designs applied, in general, are nonignorable and may have a major impact on model-based estimates. In this context, several strategies for incorporating design weights into statistical models are discussed. The application focuses on registers where only a few variables with limited predictive power are available. This reflects the situation in many countries and several branches of official statistics and shows the usefulness of the estimators under less favorable circumstances.

Our results suggest that model-based estimators should be considered in addition to purely design-based estimators due to lower RRMSEs in many settings. Furthermore, estimators ignoring the sampling design cannot be recommended since they may yield considerably biased estimates. Besides the influence of the range of design weights, our results stress the relevance of the source of design weight variation – between or within areas and strata. Altogether, our study illustrates the efficiency gains made possible by using model-based small area estimators even under less favorable circumstances.

A comparison of the augBHF and the wBHF estimator illustrates that the origin of the variation of the design weights is an essential basis for selecting the appropriate estimator. Under purely stratified designs with large Gelman factors the augBHF estimator gives reasonable results and should be the estimator of choice with respect to minimal ARRMSE, whilst the wBHF estimator suffers from the variability of $\boldsymbol{\beta}$ estimates. In contrast, under unequal probability designs the wBHF estimator is clearly the best estimator in both scenarios if one wishes to minimize the ARRMSE of the estimates. The poor performance of the augBHF estimators in this case is partly explained by the huge discrepancy between $\overline{W}_d$ and the expected mean of the sampling weights in domain d under unequal probability sampling. This causes a bias due to informative sampling where the model which holds for the population does not hold for the sample as well (cf. Pfeffermann and Sverchkov 2009). This problem of the augBHF estimator under unequal

Table 12. Mean of coverage rates – 2nd Scenario (927 domains) – t Quantiles

	PROP	EQ	UPS_{100}	$COSTA_{50}$	$BCOpt_{50}$	OPT
augBHFboot	0.744	0.623	0.443	0.648	0.834	0.825
BHF	0.714	0.692	0.446	0.706	0.766	0.738
BHFboot	0.699	0.616	0.442	0.669	0.763	0.735
Direct	0.973	0.963	0.958	0.964	0.982	0.983
GREG	0.959	0.958	0.963	0.958	0.960	0.960
wBHFboot	0.366	0.741	0.993	0.483	0.981	0.980

probability sampling could be resolved by estimating the model parameters using design weights. In our simulations, the YouRao estimator especially suffers from a poor model. In other simulations, the YouRao performed much better when auxiliary information with better predictive power was available. Similar results hold for the area-level FH estimator.

In addition to the Gelman factors and their sources of variation, the domain-specific sample size plays a crucial role for domain estimation. This can be seen from the comparatively good results of most estimators under equal and Costa-type allocation achieved at the expense of less efficient estimation at the national level. Furthermore, we note that under Scenario 2 with many small sample sizes the precision of domain estimates generally decreases compared to the first scenario with larger domains. This decrease is most pronounced for design-based estimators which cannot compensate for the small sample sizes by borrowing strength from other domains.

Focusing on the precision estimates, we observe that the confidence interval coverage rates of the design-based estimators are as expected. The shortest CI lengths result under equal allocation designs. Minor problems of the design-based estimators with very small domain-specific sample sizes are corrected by plugging-in quantiles from a t_{n_d-1} distribution. The coverage rates for the BHF were not satisfactory under either Taylor linearization of the PMSE or PMSE estimation by parametric bootstrap due to high biases. We have seen that very small domains may be problematic for precision estimates, as the severe cases of under coverage in Scenario 2 point out. Moreover, our results indicate that under (box constraint) optimal allocations in Scenario 1, the reliability of the confidence intervals of the augBHF estimator is better than the reliability of the unweighted BHF estimator. With respect to the parametric bootstrap method for the wBHF estimator, mainly in Scenario 1, we have seen overcoverage for the (box constraint) optimal allocations and unequal probability sampling, implying that the PMSE estimates are too conservative.

The present application used small and medium enterprises. When dealing with large enterprises one could expect extremely skewed distributions with outliers. Under these settings, either transformation methods (Berg and Chandra 2012 or Shlomo and Priam 2013) or robust models should be considered (Sinha and Rao 2009 or Chambers and Tzavidis 2006). A comparison of robust small area methods including computational issues can be drawn from Schmid (2012). When using nonignorable sampling designs in business surveys, the robustification of design weights should be investigated in addition to the robust modeling.

6. References

Asparouhov, T. 2006. "General Multi-Level Modeling with Sampling Weights." *Communications in Statistics – Theory and Methods* 35: 439–460. DOI: http://dx.doi.org/10.1080/03610920500476598.

Bates, D. 2011. *Computational Methods for Mixed Models*. Available at: http://cran.r-project.org/web/packages/lme4/vignettes/Theory.pdf (accessed October 16, 2014).

Bates, D., M. Maechler, and B. Bolker. 2011. *Linear Mixed-Effects Models Using S4 Classes. R package Version: 0.999375-42*. Available at: http://www.r-project.org/conferences/useR-2012/TutorialBates.pdf (accessed October 14, 2014).

Battese, G.E., R.M. Harter, and W.A. Fuller. 1988. "An Error Component Model for Prediction of County Crop Areas Using Survey and Satellite Data." *Journal of the*

American Statistical Association 83: 28–36. DOI: http://dx.doi.org/10.1080/01621459. 1988.10478561.

Berg, E. and H. Chandra. 2012. "Small Area Prediction for a Unit Level Lognormal Model." Federal Committee on Statistical Methodology Research Conference. DOI: http://dx.doi.org/10.1016/j.csda.2014.03.007.

Burgard, J.P., R. Münnich, and T. Zimmermann. 2012. "Small Area Modelling Under Complex Survey Designs for Business Data." In Proceedings of the Fourth International Conference of Establishment Surveys, June 11–14, 2012. Montreal. Available at: http://www.amstat.org/meetings/ices/2012/papers/301906.pdf (accessed Oct 16, 2014).

Butar, F.B. and P. Lahiri. 2003. "On Measures of Uncertainty of Empirical Bayes Small-Area Estimators." *Journal of Statistical Planning and Inference* 112: 63–76. DOI: http://dx.doi.org/10.1016/S0378-3758(02)00323-3.

Chambers, R., and N. Tzavidis. 2006. "M-Quantile Models for Small Area Estimation." *Biometrika* 93: 255–268. DOI: http://dx.doi.org/10.1093/biomet/93.2.255.

Chatterjee, S., P. Lahiri, and H. Li. 2008. "Parametric Bootstrap Approximation to the Distribution of EBLUP and Related Prediction Intervals in Linear Mixed Models." *The Annals of Statistics* 36: 1221–1245.

Choudhry, G.H., J.N.K. Rao, and M.A. Hidiroglou. 2012. "On Sample Allocation for Efficient Domain Estimation." *Survey Methodology* 38: 23–29.

Costa, A., A. Satorra, and E. Ventura. 2004. "Improving Both Domain and Total Area Estimation by Composition." *Statistics and Operations Research Transactions* 28: 69–86.

Datta, G.S. 2009. "Model-Based Approach to Small Area Estimation." In *Handbook of Statistics*, Vol. 29B, 251–288. New York: Elsevier.

Datta, G.S., and P. Lahiri. 2000. "A Unified Measure of Uncertainty of Estimated Best Linear Unbiased Predictors in Small Area Estimation Problems." *Statistica Sinica* 10: 613–627.

Eurostat. 2008. "NACE Rev. 2: Statistical Classification of Economic Activities in the European Community." Eurostat methodologies and working papers, European Communities, cat. No. KS-RA-07-015-EN-N.

Falorsi, P.D. and P. Righi. 2008. "A Balanced Sampling Approach for Multi-Way Stratification Designs for Small Area Estimation." *Survey Methodology* 34: 223–234.

Fay, R.E. and R.A. Herriot. 1979. "Estimation of Income for Small Places: An Application of James-Stein Procedures to Census Data." *Journal of the American Statistical Association* 74: 269–277. DOI: http://dx.doi.org/10.1080/01621459.1979.10482505.

Gabler, S., M. Ganninger, and R. Münnich. 2012. "Optimal Allocation of the Sample Size to Strata Under Box Constraints." *Metrika* 75: 15–161. DOI: http://dx.doi.org/10.1007/s00184-010-0319-3.

Gonzalez-Manteiga, W., M.J. Lombardía, I. Molina, D. Morales, and L. Santamaría. 2008. "Bootstrap Mean Squared Error of a Small-Area EBLUP." *Journal of Statistical Computation and Simulation* 78: 443–462.

Harville, D.A. 2008. *Matrix Algebra from a Statistician's Perspective*. New York: Springer.

Hidiroglou, M.A. and P. Lavallee. 2009. "Sampling and Estimation in Business Surveys." In *Handbook of Statistics*, Vol. 29A, 441–470. New York: Elsevier.

Holmberg, A., P. Flisberg, and M. Rönnqvist. 2002. "On the Choice of Sampling Design in Business Surveys with Several Important Study Variables." R&D Report 2002:3, Statistics Sweden.

Jiang, J. and P. Lahiri. 2006. "Mixed Model Prediction and Small Area Estimation." *Test* 15: 1–96. DOI: http://dx.doi.org/10.1007/BF02595419.

Kackar, R.N., and D.A. Harville. 1984. "Approximations for Standard Errors of Estimators of Fixed and Random Effect in Mixed Linear Models." *Journal of the American Statistical Association* 79: 853–862. DOI: http://dx.doi.org/10.1080/01621459.1984.10477102.

Kolb, J.-P., R. Münnich, F. Volk, and T. Zimmermann. 2013. "TRItalia dataset." In *BLUE-ETS Deliverable D6.2: Best practice recommendations on variance estimation and small area estimation in business surveys*, edited by R. Bernardini Papalia, C. Bruch, T. Enderle, S. Falorsi, A. Fasulo, E. Fernandez-Vazquez, M. Ferrante, J.P. Kolb, R. Münnich, S. Pacei, R. Priam, P. Righi, T. Schmid, N. Shlomo, F. Volk, and T. Zimmermann, 168–188. Available at: http://www.blue-ets.istat.it/fileadmin/deliverables/Deliverable6.2.pdf (accessed October 16, 2014).

Lehtonen, R., C.-E. Särndal, and A. Veijanen. 2003. "The Effect of Model Choice in Estimation for Domains, Including Small Domains." *Survey Methodology* 29: 33–44.

Lehtonen, R., C.-E. Särndal, and A. Veijanen. 2005. "Does the Model Matter? Comparing Model-Assisted and Model-Dependent Estimators of Class Frequencies for Domains." *Statistics in Transition* 7: 649–673.

Lehtonen, R., C.-E. Särndal, A. Veijanen, and M. Myrskylä. 2006. "The Role of Models in Model-Assisted and Model-Dependent Estimation for Domains and Small Areas." Workshop on survey sampling and methodology, Ventspils. Available at: http://home.lu.lv/~pm90015/workshop2006/papers/Workshop2006_04_Lehtonen.pdf (accessed October 16, 2014).

Lehtonen, R. and A. Veijanen. 2009. "Design-Based Methods of Estimation for Domains and Small Areas." In *Handbook of Statistics*, Vol. 29B, 219–249. New York: Elsevier.

Lehtonen, R., A. Veijanen, M. Myrskylä, and M. Valaste. 2011. "*Small Area Estimation of Indicators on Poverty and Social Exclusion*." Technical report, AMELI deliverable D2.2. Available at: http://www.uni-trier.de/fileadmin/fb4/projekte/SurveyStatisticsNet/Ameli_Deliverables/AMELI-WP2-D2.2.20110402.pdf (accessed October 16, 2014).

Little, R.J. 2012. "Calibrated Bayes, an Alternative Inferential Paradigm for Official Statistics." *Journal of Official Statistics* 28: 309–334.

Longford, N.T. 2006. "Sample Size Calculation for Small Area Estimation." *Survey Methodology* 32: 87–96.

Molina, I., and J.N.K. Rao. 2010. "Small Area Estimation of Poverty Indicators." *Canadian Journal of Statistics* 38: 369–385. DOI: http://dx.doi.org/10.1002/cjs.10051.

Münnich, R. and J.P. Burgard. 2012. "On the Influence of Sampling Design on Small Area Estimates." *Journal of the Indian Society of Agricultural Statistics* 66: 145–156.

Münnich, R., E. Sachs, and M. Wagner. 2012. "Numerical Solution of Optimal Allocation Problems in Stratified Sampling Under Box Constraints." *Advances in Statistical Analysis* 96: 435–450. DOI: http://dx.doi.org/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." *Journal of the Royal Statistical Society* 97: 558–625. DOI: http://dx.doi.org/10.2307/2342192.

Pfeffermann, D. 1993. "The Role of Sampling Weights When Modeling Survey Data." *International Statistical Review* 61: 317–337. DOI: http://dx.doi.org/10.2307/1403631.

Pfeffermann, D., C.J. Skinner, D.J. Holmes, H. Goldstein, and J. Rasbash. 1998. "Weighting for Unequal Selection Probabilities in Multilevel Models." *Journal of the Royal Statistical Society Series B* 60: 23–40. DOI: http://dx.doi.org/10.1111/1467-9868.00106.

Pfeffermann, D. and M. Sverchkov. 2007. "Small-Area Estimation Under Informative Probability Sampling of Areas and Within the Selected Areas." *Journal of the American Statistical Association* 102: 1427–1439. DOI: http://dx.doi.org/10.1198/016214507000001094.

Pfeffermann, D. and M. Sverchkov. 2009. "Inference Under Informative Sampling." In *Handbook of Statistics*, Vol. 29B, 455–487. New York: Elsevier.

Prasad, N.G.N., and J.N.K. Rao. 1990. "The Estimation of the Mean Squared Error of Small Area Estimators." *Journal of the American Statistical Association* 85: 163–171. DOI: http://dx.doi.org/10.1080/01621459.1990.10475320.

Rabe-Hesketh, S. and A. Skrondal. 2006. "Multilevel Modelling of Complex Survey Data." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 169: 805–827. DOI: http://dx.doi.org/10.1111/j.1467-985X.2006.00426.x.

Rao, J.N.K. 2003. *Small Area Estimation*. New York: John Wiley and Sons.

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

Schmid, T. 2012. "*Spatial Robust Small Area Estimation applied on Business Data*." Ph.D. thesis, University of Trier.

Shlomo, N. and R. Priam. 2013. "Improving Estimation in Business Surveys." In *BLUE-ETS Deliverable D6.2: Best practice recommendations on variance estimation and small area estimation in business surveys*, edited by R. Bernardini Papalia, C. Bruch, T. Enderle, S. Falorsi, A. Fasulo, E. Fernandez-Vazquez, M. Ferrante, J.P. Kolb, R. Münnich, S. Pacei, R. Priam, P. Righi, T. Schmid, N. Shlomo, F. Volk, and T. Zimmermann, 52–70. Available at: http://www.blue-ets.istat.it/fileadmin/deliverables/Deliverable6.2.pdf (accessed October 16, 2014).

Sinha, S.K. and J.N.K. Rao. 2009. "Robust Small Area Estimation." *Canadian Journal of Statistics* 37(3): 381–399. ISSN 1708-945X.

Thompson, K.J. and B.E. Oliver. 2012. "Response Rates in Business Surveys: Going Beyond the Usual Performance Measure." *Journal of Official Statistics* 28: 221–237.

Tillé, Y. 2006. *Sampling Algorithms. Springer Series in Statistics*. New York: Springer.

Torabi, M. and J.N.K. Rao. 2010. "Mean Squared Error Estimators of Small Area Means Using Survey Weights." *Canadian Journal of Statistics* 38: 598–608.

Tschuprow, A. 1923. "On the Mathematical Expectation of the Moments of Frequency Distributions in the Case of Correlated Observations." *Metron* 2: 461–493.

Verret, F., M.A. Hidiroglou, and J.N.K. Rao. 2010. "Small Area Estimation Under Informative Sampling." In Proceedings of the Survey Methods Section SSC Annual Meeting, May 2010.

You, Y. and J.N.K. Rao. 2002. "A Pseudo-Empirical Best Linear Unbiased Prediction Approach to Small Area Estimation Using Survey Weights." *The Canadian Journal of Statistics* 30: 431–439.

You, Y. and J.N.K. Rao. 2003. "Pseudo Hierarchical Bayes Small Area Estimation Combining Unit Level Models and Survey Weights." *Journal of Statistical Planning and Inference* 111: 197–208. DOI: http://dx.doi.org/10.1016/S0378-3758(02)00301-4.