Journal of Official Statistics, Vol. 39, No. 4, 2023, pp. 507–533, http://dx.doi.org/10.2478/JOS-2023-0024

Small Area with Multiply Imputed Survey Data

Marina Runge[1] *and Timo Schmid*[2]

In this article, we propose a framework for small area estimation with multiply imputed survey data. Many statistical surveys suffer from (a) high nonresponse rates due to sensitive questions and response burden and (b) too small sample sizes to allow for reliable estimates on (unplanned) disaggregated levels due to budget constraints. One way to deal with missing values is to replace them by several plausible/imputed values based on a model. Small area estimation, such as the model by Fay and Herriot, is applied to estimate regionally disaggregated indicators when direct estimates are imprecise. The framework presented tackles simultaneously multiply imputed values and imprecise direct estimates. In particular, we extend the general class of transformed Fay-Herriot models to account for the additional uncertainty from multiple imputation. We derive three special cases of the Fay-Herriot model with particular transformations and provide point and mean squared error estimators. Depending on the case, the mean squared error is estimated by analytic solutions or resampling methods. Comprehensive simulations in a controlled environment show that the proposed methodology leads to reliable and precise results in terms of bias and mean squared error. The methodology is illustrated by a real data example using European wealth data.

Key words: Fay-Herriot model; mean squared error; multiple imputation; nonresponse; survey statistics.

1. Motivation

Financial reports based on asset data can provide insights into a wide range of issues of major importance for political decisions and can help in the precise allocation of funds. In addition, wealth data can give an overview of the distribution of assets and liabilities, which can be highly relevant for financial stability and play a central role in assessing inequality. For this reason, survey data on wealth are of particular importance. Since questions about assets and income are sensitive issues, such surveys often suffer from high item nonresponse (Riphahn and Serfling 2005). For example, the Household Finance and Consumption Survey (HFCS) reports for France item nonresponse rates of nearly 30% for value of saving accounts and largest mortgage on household main residence and almost 80% for current value of household main residence (HFCN 2020a).

Listwise deletion, retaining only records with no items missing, leads to a loss of information, and the remaining units in this dataset are not a good representation of the population, which can lead to biased estimates. Missing values are a problem because the

[1] Institute of Statistics and Econometrics, Freie Universität Berlin, Garystraße 21, 14195 Berlin, Germany, Email: marina.runge@fu-berlin.de
[2] Institute of Statistics, Otto-Friedrich-Universität Bamberg, Feldkirchenstraße 21, 96045 Bamberg, Germany, Email: timo.schmid@uni-bamberg.de

Acknowledgments: The authors appreciate gratefully the support of the German Research Foundation within the TESAP project (grant number: 281573942). This article uses data from the Eurosystem Household Finance and Consumption Survey (HFCS). The results published and the related observations and analysis may not correspond to results or analysis of the data producers. Finally, the authors are indebted to the Editor-in-Chief, Associate Editor and the referees for comments that significantly improved the article.

incomplete data do not have the regular (matrix) form needed in almost any statistical method, and therefore handling missing values is necessary. In the literature there are various approaches for dealing with missing data in studies, such as in Rubin (1987) or Longford (2005). Van Buuren (2018) gives an extended overview of approaches to handling and imputing of missing data. Rubin (1976) formulated for the first time the concept of missing data mechanisms by using the indicators of the missing values as random variables and posited a model for them. Methods for missing data are generally based on the assumption that the probability of the missing data does not depend on the missing values after conditioning on the observed values (MAR). To obtain valid statistical inferences, appropriate assumptions about the mechanism of missing values must be made (Van Buuren 2018). Two approaches to handling incomplete data are single imputation, where each missing value is imputed once, and multiple imputation (MI), where the missing values are replaced by a small number of plausible values. The advantage of MI is that it reflects the uncertainty of missing data, which is then taken into account in the estimation. There are several surveys of income and wealth data where MI is used, including the Consumer Expenditure Survey, where the income variable is imputed five times (Fisher 2006), and the HFCS, where also five imputations of the data sets are provided to the user (HFCN 2020a).

Of particular interest may be subpopulations of households, either regionally disaggregated or sociodemographic such as households with particular composition (of ages, gender, labor market status, or educational levels). Various political decisions or global events, such as the financial crisis of 2007/2008 or the COVID-19 pandemic in 2020/2021, may affect these subgroups, usually referred to as areas or domains, to varying degrees. Some of these domains may be represented by very few units in the sample and direct estimators (based only on these subjects) result in a large variance. This issue may be solved by small area estimation (SAE) methods. The model-based estimators used in SAE supplement information from other areas and other data sources. Pfeffermann (2013), Rao and Molina (2015) and Jiang and Rao (2020) give compact overviews and Tzavidis et al. (2018) propose a general framework for the production of small area statistics. SAE methods can be distinguished in unit-level (e.g., Battese et al. 1988) and area-level (Fay and Herriot 1979) models. Unit-level models have the greater information content, but can only be used when unit-level covariate data are available. In addition, area-level models are often used because they are better suited to account for complex survey designs for point and variance estimates. Therefore, we focus on the Fay-Herriot model in this article. The Fay-Herriot model can be applied to transformed direct estimators to attain normality of the error terms or to ensure that the resulting estimates are within an appropriate range. Slud and Maiti (2006) and Chandra et al. (2017) study the log-transformed Fay-Herriot model and Sugasawa and Kubokawa (2017) consider a general parametric transformation of the response values. Schmid et al. (2017) use an arcsine transformation to estimate literacy rates of Senegal and Casas-Cordero et al. (2016) to estimate poverty rates of Chile.

In the context of SAE, nonresponse rates in combination with small sample sizes could have significant influence on the estimates especially with sensitive data such as income and wealth data. The investigation of the integration of the imputation uncertainty into small area estimators has received some attention. Among the publications are, for example, Longford (2004), who uses a multiple hot-deck imputation method in the UK

Labour Force Survey to estimate unemployment rates using a small area multivariate shrinkage method. Longford (2005) presents methods for dealing with incomplete data and making inferences using small area estimation methods. An approach to modeling the non-missing at random mechanism in SAE under informative sampling and nonresponse can be found in Sverchkov and Pfeffermann (2018). Kreutzmann et al. (2022) and Bijlsma et al. (2020) use a Fay-Herriot model with pooled direct estimators after multiple imputation and take into account the additional uncertainty due to the missing values in the sampling variance. However, both ignore the additional uncertainty in the regression-synthetic part of the model. We extend this approach to address the latter problem in addition to extending the methodology to ratios.

We present an approach in which we combine MI with the transformed Fay-Herriot model. We take the multiply imputed values of the missing values as given by the data provider. To account for the additional uncertainty from imputation, pooled components of the direct estimator are used, as well as pooled components of the regression-synthetic part of the Fay-Herriot model. In particular, the components (direct and regression-synthetic part) are combined for a given transformation in such a way that the resulting MI adjusted model has the known structure of Fay-Herriot models. This approach exploits the existing knowledge about transformations, back-transformations and mean squared error (MSE) approximations of the transformed Fay-Herriot model. We apply the general approach to three special cases relevant to practice and additionally discuss MSE estimators for these special cases:

1. For the general Fay-Herriot model for a mean value, we adapt the Prasad-Rao MSE estimator (Prasad and Rao 1990) to account for the uncertainty owing to missing values.
2. If the distribution of the target indicator is right-skewed, a log transformation can be used. For this case, we use the adapted Prasad-Rao MSE estimator and apply a back-transformation similar to that presented in Rao and Molina (2015).
3. For the Fay-Herriot model for a ratio with an arcsine transformation, we use insights from Hadam et al. (2023) for the back-transformation of the point estimator, as well as for a parametric bootstrap MSE estimator that can reflect the uncertainty due to the missing values.

The validity of the presented point estimators is demonstrated for the three cases outlined above in a simulation study. It is also shown that the additional uncertainty caused by the missing values is accounted for by the proposed MSE estimators.

The article is structured as follows. Sections 2, 3, and 4 describe the statistical methodology. In Section 2, the transformed Fay-Herriot model is presented, which serves as the basis for the combination with MI. Section 3 describes how the direct and regression-synthetic components of the transformed Fay-Herriot model are combined after MI, which leads to a MI adjusted Fay-Herriot model. In Section 4, we consider three common special cases of the model from Section 3 and present associated uncertainty measures. The proposed methodology is evaluated in simulation experiments in Section 5 and then applied to HFCS data in Section 6. Section 7 summarizes the main findings, discusses limitations of the approach and outlines further research potential.

2. Transformed Fay-Herriot Model

In the following the transformed Fay-Herriot model is introduced, where the transformation is described by a known function h. Let N be the size of a finite population which is partitioned into $d = 1, \ldots, D$ domains and n the sample size with $i = 1, \ldots, n_d$ units per domain so that $n = \sum_{d=1}^{D} n_d$. The Fay-Herriot model involves in the first stage a sampling model in which it is supposed that the direct estimator consists of the true domain-specific population indicator θ_d and a sampling error e_d:

$$\hat{\theta}_d^{Dir} = \theta_d + e_d, \qquad e_d \overset{ind}{\sim} N\left(0, \sigma_{e_d}^2\right).$$

It is assumed that the sampling errors e_d are independently normally distributed with known variance $\sigma_{e_d}^2$. Although the sampling variances $\sigma_{e_d}^2$ are assumed to be known, in practice they are estimated by unit-level data (Rivest and Vandal 2002; Wang and Fuller 2003: You and Chapman 2006). Another unit-level approach to address the problem of unknown sampling variances is proposed by Maiti et al. (2014) and Sugasawa et al. (2017) by shrinking and simultaneous modeling of small area means and variances. When the indicator of interest is a mean value, a domain specific direct estimator is the weighted average of the sampled values:

$$\hat{\theta}_d^{Dir} = \frac{\sum_{i=1}^{n_d} w_{id} y_{id}}{\sum_{i=1}^{n_d} w_{id}}.$$

The incorporation of sampling weights w_{id} makes the point estimator design unbiased. Note that the population and the outcomes y_{id} are assumed to be fixed, and the sampling mechanism is the only source of uncertainty. The sampling weights reflect a complex design in the estimation of the associated variance. The second stage of the Fay-Herriot model is a linking model, which links covariate information to the population indicator. x_d is a $p \times 1$ vector with area-level population covariates and β is the corresponding $p \times 1$ vector with regression coefficients. υ_d are normally distributed domain specific random effects:

$$\theta_d = x_d^T \beta + \upsilon_d, \qquad \upsilon_d \overset{iid}{\sim} N\left(0, \sigma_v^2\right). \tag{1}$$

Combining the sampling and the linking model results in:

$$\hat{\theta}_d^{Dir} = x_d^T \beta + \upsilon_d + e_d, \;\; \upsilon_d \overset{iid}{\sim} N\left(0, \sigma_v^2\right), \;\; e_d \overset{ind}{\sim} N\left(0, \sigma_{ed}^2\right). \tag{2}$$

If a smooth and monotone transformation function h is applied to the direct estimator, $\hat{\theta}_d^{Dir}$ is replaced by $\hat{\theta}_d^{Dir*} := h\left(\hat{\theta}_d^{Dir}\right)$ in Equation (2) and we want to predict $h^{-1}(\theta_d)$. The transformed Fay-Herriot model is then defined, for example, as in Sugasawa and Kubokawa (2017):

$$h\left(\hat{\theta}_d^{Dir}\right) = x_d^T \beta + \upsilon_d + e_d, \qquad \upsilon_d \overset{iid}{\sim} N\left(0, \sigma_v^2\right), \qquad e_d \overset{ind}{\sim} N\left(0, \sigma_{ed}^{2*}\right). \tag{3}$$

In the following, * always refers to the transformed scale of the direct estimator, its variance and the Fay-Herriot estimator presented at the end of this section. The model parameters, the model variance σ_v^2 and the regression coefficients β are not known and must be estimated. There are various methods to obtain estimates of σ_v^2, for example, restricted maximum likelihood (REML), maximum likelihood (ML) and the FH method-of-moments. More details on the estimation methods of the model variance can be found in Chapter 6 in Rao and Molina (2015). A drawback of ML is that it does not account for the loss in degrees of freedom arising from the estimation of the regression coefficients β (Rao and Molina 2015). Therefore, we use in this article the REML method. The regression coefficients β and the random effects v_d are estimated by:

$$\hat{\beta} = \hat{\beta}(\hat{\sigma}_v^2) = \left(\sum_{d=1}^{D} \frac{x_d x_d^T}{\sigma_{e_d}^{2^*} + \hat{\sigma}_v^2}\right)^{-1} \left(\sum_{d=1}^{D} \frac{x_d \hat{\theta}_d^{Dir^*}}{\sigma_{e_d}^{2^*} + \hat{\sigma}_v^2}\right), \tag{4}$$

$$\hat{v}_d = \frac{\hat{\sigma}_v^2}{\sigma_{e_d}^{2^*} + \hat{\sigma}_v^2}\left(\hat{\theta}_d^{Dir^*} - x_d^T \hat{\beta}\right). \tag{5}$$

Plugging those predictors into Equation (1) leads to the empirical best linear unbiased predictor (EBLUP), that is, the transformed Fay-Herriot estimator:

$$\hat{\theta}_d^{FH*} = x_d^T \hat{\beta} + \hat{v}_d. \tag{6}$$

This estimator can be expressed as a convex combination of the direct estimator and the regressionsynthetic component, resulting in an optimal combination of the two components. If the variance of the direct estimator is large, more weight is given to the synthetic component, and vice versa:

$$\hat{\theta}_d^{FH^*} = \hat{\gamma}_d \hat{\theta}_d^{Dir^*} + (1 - \hat{\gamma}_d) x_d^T \hat{\beta} \quad \text{with} \quad \hat{\gamma}_d = \frac{\hat{\sigma}_v^2}{\hat{\sigma}_{e_d}^{2^*} + \hat{\sigma}_v^2}. \tag{7}$$

At this point $\hat{\theta}_d^{FH^*}$ is still on the transformed scale and has to be transformed to the original scale to obtain $\hat{\theta}_d^{FH^*}$.

3. Combining Transformed Fay-Herriot Models after Multiple Imputation

An often applied technique to handle missing values is MI, where the missing values are replaced by several plausible values. To obtain these values, an imputation model is required. It is not sufficient to generate only one imputation, since the imputation is treated as if it were true, and the uncertainties arising from the nonresponse are ignored. On the contrary, a large number of imputations is usually not necessary, and M between 5 and 20 is sufficient, but it may be advantageous to choose a higher value (20–100) if the non-response is high and there is a large uncertainty about the estimand (Van Buuren 2018). The procedure for MI involves two steps: the imputation step and the analysis step. In the former, the imputer, usually the data provider, generates the M replicate completions of the survey data using a suitable imputation model and provides them to the analyst. In the second step, the analyst applies a statistical model suitable for the complete data separately to each imputed data set. The focus of this article is on the latter. If θ is the indicator of interest and $\hat{\theta}$ its estimator, the analysis model is calculated with each imputed data set, so we obtain $\hat{\theta}_m$ and $\widehat{\text{Var}}\ (\hat{\theta}_m)$ for $m = 1, ..., M$. The results are then combined with the application of pooling rules developed by Rubin (1987) for point estimates and their variances, which include the additional variability and uncertainty induced by the missing data. Rubin's rules (RR) are defined as follows. The pooled estimator of θ is the mean value of the M estimators:

$$\hat{\theta}^{RR} = \frac{1}{M}\sum_{m=1}^{M}\hat{\theta}_m. \tag{8}$$

The variance of the pooled estimator $\hat{\sigma}^{2RR}$ is composed by the mean value of the individual variances of each estimator (within-variance) and the variance between the M estimates (between-variance) with an correction due to the finite sample size:

$$\hat{\sigma}^{2^{RR}} = \widehat{\mathrm{Var}}\left(\hat{\theta}^{RR}\right) = \frac{1}{M}\sum_{m=1}^{M}\widehat{\mathrm{Var}}\left(\hat{\theta}_m\right) + \frac{M+1}{M}\frac{1}{M-1}\sum_{m=1}^{M}\left(\hat{\theta}_m - \hat{\theta}^{RR}\right)^2. \tag{9}$$

In the next sections, we describe how the combining rules are applied to the components of the transformed Fay-Herriot model from Section 2.

3.1. Component Pooling

With the M multiply imputed sampling values $y_{id,m}$ of each unit $i = 1, ..., n_d$ and domains $d = 1, ..., D$, the transformed direct estimators $\hat{\theta}_{d,m}^{Dir^*} = h\left(\hat{\theta}_{d,m}^{Dir}\right)$ of the target indicator and their corresponding sample variances $\sigma_{e_{d,m}}^{2^*}$ are calculated for each domain $d = 1, ..., D$ and $m = 1, ..., M$. Rubin's rules are based on asymptotic theory, and the resulting combined estimate is more accurate if the distribution of the indicator of interest is better approximated by the normal distribution (Rubin 1987). Van Buuren (2018) states that to promote approximate normality, target indicators can be transformed, then pooled and back-transformed. Therefore, the M direct estimators $\hat{\theta}_{d,m}^{Dir^*}$ and their variances $\sigma_{e_{d,m}}^{2^*}$ are pooled on the transformed scale and substituted in Equations (8) and (9). Kreutzmann et al. (2022) present a Fay-Herriot estimator which uses pooled direct components on the original scale, which are substituted in the (log transformed) Fay-Herriot model. We extend this approach and transform the direct components of each imputed data set to estimate the regression-synthetic components. This allows the uncertainty of the missing values to be included not only in the direct components, but also in those of the linking model. The model components of the linking model are estimated for each imputed data set. The estimated variances of the random effects $\hat{v}_{d,m}$ are combined according to Rubin's rule:

$$W = \frac{1}{M}\sum_{m=1}^{M}\hat{\sigma}_{v_m}^2 \quad \text{and} \quad B_d = \frac{M+1}{M}\frac{1}{M-1}\sum_{m=1}^{M}\left(\hat{v}_{d,m} - \frac{1}{M}\sum_{m=1}^{M}\hat{v}_{d,m}\right)^2. \tag{10}$$

The mean squared distance of the random effects of the domains of the M imputed data sets and the pooled random effects per domain is different between the areas. In order to guarantee that the random effects have a common variance, further pooling has to be applied. Therefore, the mean value of the between variance is taken. Together with Equation (10) this leads to the pooled model variance:

$$\hat{\sigma}_v^{2^{RR}} = W + \frac{1}{D}\sum_{d=1}^{D}B_d. \tag{11}$$

The pooled model variance $\hat{\sigma}_v^{2^{RR}}$ and the pooled direct components are now used to obtain MI adjusted estimates of the regression coefficients and random effects $\hat{\sigma}_v^{2^{RR}}$, $\hat{\sigma}_{e_d}^{2^{RR^*}}$ and $\hat{\theta}_d^{Dir.RR^*}$ are inserted into Equation (4) to obtain the MI adjusted regression coefficients $\hat{\beta}$ and then together into Equation (5) to obtain the MI adjusted random effects $\hat{v}_d$.

3.2. MI Adjusted Fay-Herriot Model

The pooled direct components together with the pooled and MI adjusted regression-synthetic parts of the model lead to the MI adjusted Fay-Herriot model, which preserves the structure of the transformed Fay-Herriot model. The area-level population auxiliary information x_d, obtained from external sources, such as the census, is fixed and complete as in Equation (1). The model can be written analogously to Equation (3) with pooled direct components and the pooled model variance. Using the estimators of unknown model parameters as elaborated in Subsection 3.1 leads to the proposed FH.MI estimator $\hat{\theta}_d^{FH.MI^*}$, which can be written analogously to Equation (7) with $\hat{\theta}_{e_d}^{Dir.RR^*}$, $\hat{\sigma}_{e_d}^{2^{RR^*}}$ and $\hat{\sigma}_v^{2^{RR}}$ plugged in:

$$\hat{\theta}_d^{FH.MI^*} = \hat{\gamma}_d \hat{\theta}_d^{Dir.RR^*} + (1 - \hat{\gamma}_d) x_d^T \hat{\beta} \quad \text{with} \quad \hat{\gamma}_d = \frac{\hat{\sigma}_v^{2^{RR}}}{\hat{\sigma}_{e_d}^{2^{RR^*}} + \hat{\sigma}_v^{2^{RR}}}. \tag{12}$$

The presented $\hat{\theta}_d^{FH.MI^*}$ estimator preserves the representations of the Fay-Herriot estimator. As $\hat{\theta}_d^{FH.MI^*}$ is on the transformed scale, a suitable back transformation depending on h has to be applied to obtain $\hat{\theta}_d^{FH.MI}$.

Small area estimators with multiply imputed data can be derived in two ways: 1. Fit the Fay-Herriot model to each of the M imputed data sets and combine the Fay-Herriot estimators with Rubin's rule. 2. Estimate the direct and the regression synthetic components M times and combine them using Rubin's rules as described in Subsection 3.1 and then estimate the shrinkage estimator in Equation (12). The advantage of the first approach is that it is simple. However, it loses the structure of the Fay-Herriot model and the representation of the estimator as a weighted combination of the direct and regression synthetic components. In addition, it is unclear how the uncertainty of the M Fay-Herriot estimators is combined, since Rubin's rule is commonly used for variances and it is unclear how this rule can be applied to the MSE. The advantage of the second (the proposed) approach and the resulting FH.MI estimator is that the model structure of the Fay-Herriot model is preserved, the interpretability of the components is maintained, and the existing knowledge about MSE estimators is directly transferable and extensible. The estimator of the first approach is used as a benchmark in the model-based simulation study in Section 5 and is denoted by FH.RR.

4. MI Adjusted Fay-Herriot Estimators with Uncertainty Measures

In the following sections, we focus on three special cases of the transformed MI adjusted Fay-Herriot estimator (12). For each case we specify the FH.MI point estimator and an associated MSE estimator.

4.1. Estimator for a Mean

The (population) mean of a quantity of interest for domain d is estimated by the weighted sample average per imputed data set m:

$$\hat{\theta}_{d,m}^{Dir} = \frac{\sum_{i=1}^{n_d} w_{id} y_{id,m}}{\sum_{i=1}^{n_d} w_{id}} \quad \text{for} \quad d = 1, .., D \quad \text{and} \quad m = 1, ..., M. \tag{13}$$

If no transformation is required for the direct estimator, $\hat{\theta}_d^{FH.MI^*}$ is on the original scale such that $\hat{\theta}_d^{FH.MI} = \hat{\theta}_d^{FH.MI^*}$. With the pooled and MI adjusted estimators presented in Section 3, the FH.MI estimator $\hat{\theta}_d^{FH.MI}$ can be calculated according to Equation (12). As a measure of uncertainty which captures the additional uncertainty due to multiple imputation, we adapt the MSE estimator of Prasad and Rao (1990) in the following. The second-order approximation of the MSE of $\hat{\theta}_d^{FH}$ is given by:

$$\text{MSE}\left(\hat{\theta}_d^{FH}\right) \approx g_{1d}\left(\sigma_v^2\right) + g_{2d}\left(\sigma_v^2\right) + g_{3d}\left(\sigma_v^2\right).$$

The first component g_{1d} is based on the prediction of the random effects and g_{2d} reflects the variability arising from the estimation of the regression coefficients. g_{1d} and g_{2d} are independent of the estimation method of the model variance σ_v^2, whereas, g_{3d} reflects the uncertainty caused by the estimation of σ_v^2 and depends on the estimation method through its asymptotic variance $\bar{V}\left(\hat{\sigma}_v^2\right)$ (as $D \to \infty$) (see e.g., Rao and Molina 2015). According to Prasad and Rao (1990) a second-order unbiased estimator of MSE ($\hat{\theta}_d^{FH}$) is:

$$\widehat{\text{MSE}}\left(\hat{\theta}_d^{FH}\right) = g_{1d}\left(\hat{\sigma}_v^2\right) + g_{2d}\left(\hat{\sigma}_v^2\right) + 2g_{3d}\left(\hat{\sigma}_v^2\right).$$

The components of the Prasad-Rao estimator using REML are defined as follows:

$$g_{1d}\left(\hat{\sigma}_v^2\right) = \hat{\gamma}_d^2 \sigma_{e_d}^2, \tag{14}$$

$$g_{2d}\left(\hat{\sigma}_v^2\right) = (1 - \hat{\gamma}_d)^2 x_d^T \left\{ \sum_{d=1}^{D} \frac{x_d x_d^T}{\sigma_{e_d}^2 + \hat{\sigma}_v^2} \right\}^{-1} x_d, \tag{15}$$

$$g_{3d}\left(\hat{\sigma}_v^2\right) = (\sigma_{e_d}^2)^2 (\sigma_{e_d}^2 + \hat{\sigma}_v^2)^{-3} \bar{V}(\hat{\sigma}_v^2), \tag{16}$$

$$\bar{V}(\hat{\sigma}_v^2) = 2 \left\{ \sum_{d=1}^{D} \frac{1}{\left(\sigma_{e_d}^2 + \hat{\sigma}_v^2\right)^2} \right\}^{-1}.$$

In the same way as in Subsection 3.1, where we obtain M estimates of the model variance, that is, $\hat{\sigma}_{v_m}^2$ for $m = 1, ..., M$, we obtain M corresponding asymptotic ($D \to \infty$) variances $\bar{V}_m(\hat{\sigma}_{v_m}^2)$ for $m = 1, ..., M$. To adjust the MSE estimator for this additional uncertainty, the asymptotic variances are pooled with Rubin's rule for variances (9):

$$\bar{V}^{RR}\left(\hat{\sigma}_{v}^{2^{RR}}\right) = \frac{1}{M}\sum_{m=1}^{M}\bar{V}_{m}\left(\hat{\sigma}_{v_m}^{2}\right) + \frac{M+1}{M}\frac{1}{M-1}\sum_{m=1}^{M}\left(\hat{\sigma}_{v_m}^{2} - \hat{\sigma}_{v}^{2^{RR}}\right)^{2}$$
$$\text{with} \quad \bar{V}\left(\hat{\sigma}_{v_m}^{2}\right) = 2\left\{\sum_{d=1}^{D}\frac{1}{\left(\sigma_{e_{d,m}}^{2} + \hat{\sigma}_{v_m}^{2}\right)^{2}}\right\}^{-1} \quad \text{for} \quad m = 1, ..., M. \tag{17}$$

Using $\hat{\sigma}_{v}^{2^{RR}}$ and $\sigma_{e_d}^{2^{RR}}$ in Equations (14), (15), and (16) together with the pooled asymptotic variance (17) takes into account the uncertainty about the missing values. Note that instead of plugging the pooled variance terms into the asymptotic variance formula, the pooled asymptotic variance $\bar{V}^{RR}\left(\hat{\sigma}_{v}^{2^{RR}}\right)$ is used, introducing an additional term into the estimator due to the between-variation. This leads to the proposed MSE estimator for $\hat{\theta}_{d}^{FH.MI}$, which captures the uncertainty due to missing values:

$$\hat{\mathrm{V}}\left(\hat{\theta}_{d}^{FH.MI}\right) = g_{1d}\left(\hat{\sigma}_{v}^{2^{RR}}\right) + g_{2d}\left(\hat{\sigma}_{v}^{2^{RR}}\right) + 2\left(\sigma_{e_d}^{2^{RR}}\right)^{2}\left(\sigma_{e_d}^{2^{RR}} + \hat{\sigma}_{v}^{2^{RR}}\right)^{-3}\bar{V}^{RR}\left(\hat{\sigma}_{v}^{2^{RR}}\right). \tag{18}$$

4.2. Estimator for a log Mean

Domain specific mean values of income and wealth data are often skewed to the right, or the relationship with the auxiliary information may be non-linear. In such a case, the linear Fay-Herriot model (Subsection 4.1) may be more appropriate for the log-transformed direct estimator. Using the direct estimator from Equation (13) and $h: z \rightarrow log(z)$ the direct components of the model for the M imputed data sets are:

$$\hat{\theta}_{d,m}^{Dir^{*}} = \log\left(\hat{\theta}_{d,m}^{Dir}\right) \quad \text{with variances} \quad \sigma_{e_{d,m}}^{2^{*}} \approx \left(\hat{\theta}_{d,m}^{Dir}\right)^{-2}\sigma_{e_{d,m}}^{2}$$
$$\text{for} \quad d = 1, .., D, \quad m = 1, ..., M.$$

Using a Taylor expansion for moments, the sample variance, that is, the variance of the direct estimator, can be moved to the logarithmic scale. Although this is an approximation for large samples, it is used in SAE as in Neves et al. (2013). Citro and Kalton (2000) use the same approximation with a minor modification based on the properties of the log-normal distribution, while noting that the results do not differ considerably. Calculating the direct and the regression-synthetic components as described in Subsection 3.1 with $h: z \rightarrow log(z)$ and together with Equation (12) leads to the Fay-Herriot-MI estimator $\hat{\theta}_{d}^{FH.MI^{*}}$, which is still on the log-scale. The estimates can be transformed back to the original scale by several methods. Slud and Maiti (2006) present a bias-correction under a log-transformed Fay-Herriot model and propose a corresponding estimator for the MSE. Chandra et al. (2017) extend this estimator by an additional bias correction that accounts for the sampling variation of the estimator. These methods can be applied only to observed/sampled areas. We apply a method that is suitable even for domains/areas with no observations. To obtain the point estimator on the original scale, properties of the log-normal distribution are used and the back-transformation for the MSE estimator is based on a Taylor expansion similar to that presented in Rao and Molina (2015). A short derivation can be found in the Appendix (Section 8). The back-transformation is defined as follows:

$$\hat{\theta}_d^{FH.MI} = exp\left\{\hat{\theta}_d^{FH.MI^*} + 0.5\widehat{\text{MSE}}\left(\hat{\theta}_d^{FH.MI^*}\right)\right\},$$

$$\widehat{\text{MSE}}\left(\hat{\theta}_d^{FH.MI}\right) = exp\left\{\hat{\theta}_d^{FH.MI^*} + 0.5\widehat{\text{MSE}}\left(\hat{\theta}_d^{FH.MI^*}\right)\right\}^2 \widehat{\text{MSE}}\left(\hat{\theta}_d^{FH.MI^*}\right).$$

$\widehat{\text{MSE}}(\hat{\theta}_d^{FH.MI^*})$ denotes at this point the adapted Prasad-Rao MSE estimator defined in Equation (18).

4.3. *Estimator for an arcsine Ratio*

The Fay-Herriot model is widely used for estimating poverty or literacy rates with high regional resolution. In order to guarantee that the estimated rates are between 0 and 1 suitable transformations are frequently used. The arcsine transformation h: $z \rightarrow sin^{-1}(\sqrt{z})$, of which the inverse maps its values to $[0, 1]$, is commonly used. Schmid et al. (2017) compared in a design-based simulation the arcsine transformation with an estimator based on a normal-logistic distribution. Both estimators provided very similar results regarding bias and root mean squared error (RMSE). We concentrate on the arcsine transformation because, unlike the logit, it is well defined even at zero and unity. The arcsine transformation is applied to the direct ratio estimators of the M imputed data sets:

$$\hat{\theta}_{d,m}^{Dir^*} = sin^{-1}\left(\sqrt{\hat{\theta}_{d,m}^{Dir}}\right) \quad \text{with variances} \quad \sigma_{e_{d,m}}^{2^*} = \sigma_{e_d}^{2^*} = \frac{1}{4\tilde{n}_d} \quad \text{for} \quad m = 1, .., M.$$

The effective sample size of domain d is denoted by $\tilde{n}_d$, which takes into account the sampling design effect (Jiang et al. 2001). The approximation of the sampling error variance on the transformed scale is based on a Taylor expansion for moments like in Jiang et al. (2001). The combined point estimator $\hat{\theta}_d^{Dir.RR^*}$ and its variance $\hat{\sigma}_{e_d}^{2RR*}$ are calculated by applying Rubin's rules presented in Equations (8) and (9). The components of the regression-synthetic part of the model are calculated as described in Subsection 3.1 with the pooled direct components on the transformed scale. Afterwards $\hat{\theta}_d^{FH.MI^*}$ can be calculated as in Equation (12). The resulting estimator $\hat{\theta}_d^{FH.MI^*}$ is on a $sin^{-1}(\sqrt{\ })$-scale and needs to be transferred to the original scale. A naive back-transformation is the inverse h^{-1}, which introduces a bias for non-linear h. For this reason, for common transformations bias-corrected back-transformations are proposed, such as in Hadam et al. (2023) for the arcsine transformation which is a special case of Sugasawa and Kubokawa (2017), who present an asymptotically unbiased back-transformation for a general parametric transformation. We apply the bias-corrected back-transformation following Hadam et al. (2023), using the normal distribution of the transformed estimator and the expected value (E) of a transformed variable:

$$\hat{\theta}_d^{FH.MI} = \mathrm{E}\left[sin^2\left(\hat{\theta}_d^{FH.MI^*}\right)\right] = \int_{-\infty}^{\infty} sin^2(t) f_{\hat{\theta}_d^{FH.MI^*}}(t)dt$$

$$= \int_{-\infty}^{\infty} sin^2(t) \frac{1}{\sqrt{2\pi \frac{\hat{\sigma}_v^{2^{RR}} \sigma_{e_d}^{2^{RR^*}}}{\hat{\sigma}_v^{2^{RR}} + \sigma_{e_d}^{2^{RR^*}}}}} exp\left\{ - \frac{\left(t - \hat{\theta}_d^{FH.MI^*}\right)^2}{2\frac{\hat{\sigma}_v^{2^{RR}} \sigma_{e_d}^{2^{RR^*}}}{\hat{\sigma}_v^{2^{RR}} + \sigma_{e_d}^{2^{RR^*}}}} \right\} dt. \tag{19}$$

The integral in Equation (19) must be solved by numerical integration methods. The MSE of $\hat{\theta}_d^{FH.MI}$ is approximated with a parametric bootstrap procedure analogue to Hadam et al. (2023) based on Gonzalez-Manteiga et al. (2005). The bootstrap procedure comprises the following steps:

1. Estimate the regression-synthetic components $\hat{\beta}$ and $\hat{\sigma}_v^{2^{RR}}$ analogously to Subsection 3.1 using the pooled direct components $\hat{\theta}_d^{Dir.RR^*}$ and $\hat{\sigma}_{e_d}^{2^{RR^*}}$ on the arcsine scale.
2. For b = 1,..., B
   (a) Generate sampling errors $e_d^{(b)} \overset{ind}{\sim} N\left(0, \hat{\sigma}_{e_d}^{2^{RR^*}}\right)$ and random effects $v_d^{(b)} \overset{ind}{\sim} N\left(0, \hat{\sigma}_v^{2^{RR}}\right)$,
   (b) Simulate a bootstrap sample $\hat{\theta}_d^{Dir^*(b)} = x_d^T\hat{\beta} + v_d^{(b)} + e_d^{(b)}$,
   (c) Calculate the true bootstrap population indicator $\theta_d^{*(b)} = x_d^T\hat{\beta} + v_d^{(b)}$ on the transformed scale and back-transform with $\theta_d^{(b)} = sin^2\left(\theta_d^{*(b)}\right)$,
   (d) Calculate the bootstrap estimator of the model variance $\hat{\sigma}_v^{2(b)}$ using $\hat{\theta}_d^{Dir^*(b)}$ and $\hat{\sigma}_{e_d}^{2^{RR^*}}$,
   (e) Using $\hat{\sigma}_v^{2(b)}$ and $\hat{\theta}_d^{Dir^*(b)}$, calculate bootstrap estimators of the regression coefficients $\hat{\beta}^{(b)}$and estimate the random effects $\hat{v}_d^{(b)}$, and
   (f) Determine the bootstrap estimator $\hat{\theta}_d^{FH.MI^*(b)}$ with Equation (12) by using the estimates from the step before and back-transform to the original scale applying (19) to obtain $\hat{\theta}_d^{FH.MI^*(b)}$.
3. Estimate the MSE:

$$\widehat{\mathrm{MSE}}\left(\hat{\theta}_d^{FH.MI}\right) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\theta}_d^{FH.MI(b)} - \theta_d^{(b)}\right)^2.$$

The pooled sampling and model variances, which account for the additional uncertainty about the missing values, are used in the initialization of the bootstrap method. Hence, the extra uncertainty induced by the missing data is accounted for by the bootstrap MSE estimator.

5. Simulation Study

In this section, we investigate the behaviour of the estimators proposed in Sections 3 and 4 by simulation studies with suitable data models. The population is repeatedly generated according to an underlying model. With each simulation run, a sample is taken from the generated population, to which the methods are then applied. We evaluate the performance in terms of bias and RMSE of the proposed point estimators and the inflation of RMSE arising from MI.

5.1. Data Generation

The simulation setup and data models are chosen to be consistent with those of Kreutzmann et al. (2022). For the simulations, finite populations of size $N = 60{,}000$ with $D = 100$ domains are generated so that in each domain the population size N_d is between 200 and 1,000 for $d = 1,\ldots, D$. The samples were drawn via stratified random sampling, where the strata represent the domains. To have rather small and large domains in the samples, sample sizes n_d lie within a range of 8 and 145, so that the total sample size is $n = 5{,}961$. To apply the transformations discussed in the special cases in Section 4, appropriate data models are chosen. In the standard case, a normal data model is used, where no transformation to the direct estimator of a mean value is necessary. Right-skewed log-normal data is generated when investigating the proposed method with a log transformation like in Subsection 4.2. In many applications, the indicator of interest is a ratio. In order to construct a ratio that is used in real data applications, a wealth ratio is calculated. In publications of the Federal Statistical Office (see e.g., Destatis 2018) it is derived by taking the percentage of households with a household income above the 200% median household income. As data model for the ratio the log-scale data is also used. The unit-level data models and scenarios are described in detail in Table 1. The shapes of the distribution for one selected population can be found in Figure 5 in the Appendix (Subsection 8.2). With a sample at the unit-level, the missing data is generated.

As mentioned in Section 1, MAR is often plausible and assumed in most programs for handling missing data. Therefore, in the simulation, missing values are generated using the fully observed additional variable x, from the data models in Table 1. The MAR mechanism is implemented as follows:

$$y_{id} = \begin{cases} y_{\text{missing}}, & x_{id} \le x_q \\ y_{id}, & \text{otherwise.} \end{cases} \tag{20}$$

x_q is the q-quantile of the auxiliary information x from the sample. This results in a non-response rate of $q \cdot 100\%$ by definition of the q-quantile. For the selected data models, the implemented MAR mechanism leads to missing values in the upper ends of the distribution. When it comes to sensible data as wealth related data, item nonresponse rates can be very high. For example, the Household Finance and Consumption Network (HFCN) reports for 2017 (HFCN 2020a) nonresponse rates for the value of savings account between 18% in Belgium and 64% in Finland. Therefore, it is reasonable to investigate the proposed methods under varying $q \in \{0.1, 0.3, 0.5\}$ to obtain nonresponse rates of 10%, 30% and 50%. A two-level normal model is used as an imputation model for the missing y_{id} values, which is implemented in the R-package `mice` (Van Buuren and Groothius-Oudshoorn 2011). The x serve as covariate information and v_d as area-specific random effects, so that

Table 1. Overview of unit-level data models in model-based simulation, $i = 1,\ldots, N$, $d = 1,\ldots, D$.

Setting	y_{id}	x_{id}	μ_d	v_d	e_{id}
Mean	$250000-400x_{id} + v_d + e_{id}$	$N(\mu_d; 150^2)$	$U[-150, 150]$	$N(0, 25000^2)$	$N(0, 50000^2)$
Logmean	$exp(15-x_{id} + v_d + e_{id})$	$N(\mu_d; 1)$	$U[3, 5]$	$N(0, 0.4^2)$	$N(0, 0.6^2)$
Ratio	$exp(15-x_{id} + v_d + e_{id})$	$N(\mu_d; 1)$	$U[3, 5]$	$N(0, 0.4^2)$	$N(0, 0.6^2)$

the clustering is incorporated in the imputation model. According to Van Buuren (2018), between five and 20 imputed values are often sufficient for each missing observation. The HFCN delivers five imputed values per missing observation, hence in the simulation we set $M = 5$. In the log-scale setting the data was log transformed prior to the imputation to achieve normality and back transformed with the inverse afterwards. After imputation, the data is still on a unit-level and has to be aggregated on an area-level according to the indicator of interest of the setting. Then the appropriate FH.MI estimators given in Section 3 with the special cases in Section 4 are calculated. Table 2 provides an overview showing for each setting the direct estimator, the transformation used, and the section of the corresponding FH.MI model for the special case. In Table 2, I denotes an indicator function that is 1 if the condition is true and 0 otherwise; $\tilde{Y}$ denotes the population median of y.

Each setting, including the generation of the population according to the data model, the sampling, the missing data generating process, the multiple imputation and the application of the MI adjusted FH estimators is repeated $R = 500$ times. The steps of the simulation can be summarized as follows: We generate the population according to a data model in Table 1. Next a stratified random sample is selected. Then missing values are generated according to Equation (20) and imputed to create M copies of the data. Using the M data sets the direct estimators are calculated according to Table 2 and x_{id} are aggregated to a domain level by taking the mean per domain. Afterwards the indicator of interest and its MSE are estimated by applying the methods described in Sections 3 and 4.

5.2. *Performance of Point Estimators*

In the simulation we assess the performance of six point estimators in the *mean* and *log mean* setting and five in the *ratio* setting. For each setting direct, (Direct) and Fay-Herriot (FH) estimators are calculated before deletion on the aggregated sample, that is, the steps of deleting and imputing are omitted. In the case of the FH estimator, the transformation corresponding to the setting is applied so that the Fay-Herriot estimator introduced in Section 2 is calculated. The FH estimator before deletion serves as the gold standard in this simulation. In addition, we compare the performance of the proposed FH.MI estimators with the pooled Fay-Herriot estimator (FH.RR) mentioned in Section 3 and with the estimator proposed by Kreutzmann et al. (2022) denoted by FH.DirectRR. They consider the estimator under a normal and log-normal setting for a mean value, and so we also examine this estimator only under these settings. Furthermore, with Rubin's rule combined direct estimators (Direct.RR) are calculated to show the efficiency gain of the Fay-Herriot estimators with good covariate information after MI. All estimators are implemented in the statistical programming language R (R Core Team 2020) and for the

Table 2. Overview of settings.

Setting	$\hat{\theta}_d^{Dir}$	$h\left(\hat{\theta}_d^{Dir}\right)$	FH.MI model
Mean	$\frac{1}{n_d}\sum_{i=1}^{n_d} y_{id}$	$\hat{\theta}_d^{Dir}$	4.1
Log mean	$\frac{1}{n_d}\sum_{i=1}^{n_d} y_{id}$	$log\left(\hat{\theta}_d^{Dir}\right)$	4.2
Ratio	$\frac{1}{n_d}\sum_{i=1}^{n_d} I(y_{id} > 2\cdot\tilde{Y})$	$sin^{-1}\left(\sqrt{\hat{\theta}_d^{Dir}}\right)$	4.3

standard area-level models and its components the package emdi (Kreutzmann et al. 2019) was used. The code can be obtained from the authors on request. To evaluate and compare the performance of the estimators, the following quality measures are calculated using the R Monte-Carlo replications. $\hat{\theta}_{d_r}$ denotes the estimator of the target indicator in domain d and replication r, θ_{d_r} is the true value of the indicator:

$$\mathrm{Bias}\left(\hat{\theta}_d\right) = \frac{1}{R}\sum_{r=1}^{R}\left(\hat{\theta}_{d_r} - \theta_{d_r}\right), \quad \mathrm{rel.\ Bias}\left(\hat{\theta}_d\right) = \frac{1}{R}\sum_{r=1}^{R}\left(\frac{\hat{\theta}_{d_r} - \theta_{d_r}}{\theta_{d_r}}\right),$$
$$\mathrm{RMSE}\left(\hat{\theta}_d\right) = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(\hat{\theta}_{d_r} - \theta_{d_r}\right)^2}, \quad \mathrm{RRMSE}\left(\hat{\theta}_d\right) = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(\frac{\hat{\theta}_{d_r} - \theta_{d_r}}{\theta_{d_r}}\right)^2}. \tag{21}$$

We want to evaluate the performance of the introduced methodology in terms of bias and RMSE. For the *mean* and *log mean* setting we consider the relative bias and the RRMSE. For the *ratio* setting the bias and RMSE are taken into account since the indicator itself is already on a relative scale. The median and mean values over domains of the bias and RMSE values for different nonresponse rates are presented in Table 3. The direct estimators (Direct.RR) remain unbiased after multiple imputation in the *mean* and *ratio* setting as before deletion (Direct) and almost unbiased in the *log mean* setting. The small bias could be introduced by the inverse back-transformation after applying the imputation model. Compared to the combined direct estimators (Direct.RR) and the model-based estimators before deletion (FH), the model-based estimators FH.MI, FH.RR and FH.DirectRR remain also unbiased in the *mean* and *ratio* setting and the results of the model-based estimators are comparable. Only in the *log mean* setting does the FH.MI estimator, like the other two model-based estimators, suffer from a small bias that increases slightly with higher nonresponse rates. Again this bias could be due to the inverse back-transformation in the imputation process. In terms of efficiency, we see that the RRMSE/RMSE are the smallest before deletion and increase with higher nonresponse rates for each estimator in each setting, reflecting the additional uncertainty about missing values. Within each setting and nonresponse rate the order of the RRMSE/RMSE is as expected: the RRMSE/RMSE of the direct estimators is always higher than that of the proposed FH.MI estimator, which shows that the introduced methodology behaves the same way as in cases without missing values (i.e., before deletion). The RRMSE/RMSE of the FH.MI and the FH.RR are almost identical, which indicates that the proposed methodology leads to reasonable results and is similar to the more straightforward approach of combining the Fay-Herriot estimators. The proposed FH.MI estimator is at least as efficient as the FH.DirectRR estimator. In the *log mean* setting, the *superefficiency* of imputation, when more information is used than in the analysis model (Rubin 1996), can be observed. At a nonresponse rate of 10%, Direct.RR is slightly more efficient than the direct estimator before deletion (Direct). All summed up, the results confirm our expectations. The presented FH.MI estimators lead to plausible results regarding bias and efficiency in the investigated settings, in which the imputation models follow the data structure of the generated population and thus fit the data.

Table 3. Relative bias and RRMSE for mean and log mean, bias and RMSE for ratio.

Nonresponse rate		Before deletion		10%		30%		50%	
	Estimator	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Mean									
(rel.) Bias [%]	Direct	0.0464	0.0149						
	Direct.RR			0.0254	0.0198	0.0390	0.0092	0.0862	0.0290
	FH	0.2390	0.1812						
	FH.Direct.RR			0.2291	0.1691	0.2536	0.1872	0.3082	0.2583
	FH.MI			0.2245	0.1615	0.2355	0.1761	0.2704	0.2186
	FH.RR			0.2171	0.1568	0.2195	0.1639	0.2554	0.1840
Log mean									
(rel.) Bias [%]	Direct	-0.2191	-0.0318						
	Direct.RR			0.1548	0.0342	1.1479	0.8566	2.8100	2.2903
	FH	-0.8797	-0.6057						
	FH.Direct.RR			0.0191	0.2091	1.4284	1.4639	3.1864	2.9609
	FH.MI			-0.2772	-0.1096	0.8383	0.8272	2.4169	2.3568
	FH.RR			-0.6948	-0.4258	0.0216	0.2115	1.3747	1.4485
Ratio									
Bias	Direct	-0.0004	0.0000						
	Direct RR			-0.0003	0.0000	-0.0000	0.0005	0.0009	0.0007
	FH	-0.0027	-0.0022						
	FH.MI			-0.0016	-0.0010	0.0012	0.0012	0.0011	0.0016
	FH.RR			-0.0026	-0.0021	-0.0024	-0.0018	-0.0015	-0.0009
Mean									
RRMSE [%]	Direct	5.0318	4.2722						
	Direct.RR			5.1345	4.4849	5.5337	4.7889	6.1003	5.4419
	FH	4.4300	3.9609						
	FH.Direct.RR			4.5470	4.1570	4.9845	4.5694	5.6775	5.3471
	FH.MI			4.5444	4.1524	4.9643	4.5509	5.6018	5.2498
	FH.RR			4.5386	4.1385	4.9517	4.5388	5.5741	5.1978
Log mean									
RRMSE [%]	Direct	25.5219	23.0001						
	Direct.RR			24.8037	22.0991	26.3014	23.1315	29.1076	26.1128
	FH	20.7739	20.0316						
	FH.Direct.RR			21.916	21.4101	23.8789	22.5175	27.0243	25.9548
	FH.MI			21.3353	20.6552	22.7919	21.3174	25.4294	23.9328
	FH.RR			20.7741	19.9455	22.1078	20.6367	24.7177	23.3957
Ratio									
RMSE	Direct	0.0655	0.0563						
	Direct.RR			0.0655	0.0565	0.0663	0.0565	0.0702	0.0617
	FH	0.0539	0.0506						
	FH.MI			0.0544	0.0510	0.0572	0.0533	0.0636	0.0607
	FH.RR			0.0541	0.0507	0.0564	0.0524	0.0624	0.0590

5.3. *Performance of Uncertainty Measures*

We now move on to the performance of the three proposed MSE estimators of the FH.MI estimator, each corresponding to one setting. In the case of the *mean* and *log mean* setting, we evaluate the adapted analytical Prasad-Rao estimator as described in Subsections 4.1 and 4.2 with a back-transformation when the log transformation is used. In the *ratio* setting the parametric bootstrap estimator from Subsection 4.3 with $B = 500$ replications is evaluated. Performance is evaluated by looking at the relative bias of the MSE estimator defined as followed:

$$\mathrm{RBRMSE}(\hat{\theta}_d) = \frac{\sqrt{\frac{1}{R}\sum_{r=1}^{R}\widehat{\mathrm{MSE}}_{dr} - RMSE\left(\hat{\theta}_d\right)}}{\mathrm{RMSE}\left(\hat{\theta}_d\right)}.$$

Table 4 shows the median and mean values over the domains of the RBRMSE. We see a slight underestimation in the *mean* setting with an increasing effect at higher nonresponse rates. On the other hand, in the *log mean* setting the true RMSE is slightly overestimated at a lower nonresponse rate of 10% and minimally underestimated at a higher nonresponse rate of 50%. Nevertheless, the values are all close to zero. In the *ratio* setting, the bias of the bootstrap RMSE estimator is close to zero at 10% nonresponse rate. At 30% and 50% it increases and reaches almost identical values, but still at a tolerable level. In all three settings the additional uncertainty of the FH.MI estimator can be satisfactorily addressed and the bias is within an acceptable range. To have a closer look on the performance of the adapted Prasad-Rao MSE estimator the estimated and true RMSE values per domain are plotted in Figure 1 for the *mean* setting. First we observe that within each nonresponse rate the estimated RMSE decreases with higher sample size, which is in line with the behaviour of the true RMSE. Secondly, we see that per domain the estimated RMSE values increase with increasing nonresponse rates, which is consistent with the expected behaviour. At a nonresponse rate of 10% and 30%, the estimated RMSE tracks very well the behaviour of the true RMSE. With a higher nonresponse rate of 50% we see that there are underestimations in some areas, but overall the uncertainty is well accounted for. The proposed methods are good at capturing the additional variation due to the missing observations and imputation and also provide a realistic estimate of the uncertainty of the FH.MI estimator in our settings.

Table 4. Relative bias (%) of estimated RMSE (RBRMSE) of FH.MI.

Nonresponse rate	10%		30%		50%	
	Mean	Median	Mean	Median	Mean	Median
Mean	-1.4198	-1.8291	-3.4427	-3.1477	-6.9352	-6.9390
Log mean	2.5119	2.5719	1.8185	2.6527	-4.0788	-3.2214
Ratio	2.9396	3.1787	8.7815	9.0866	8.1231	8.278

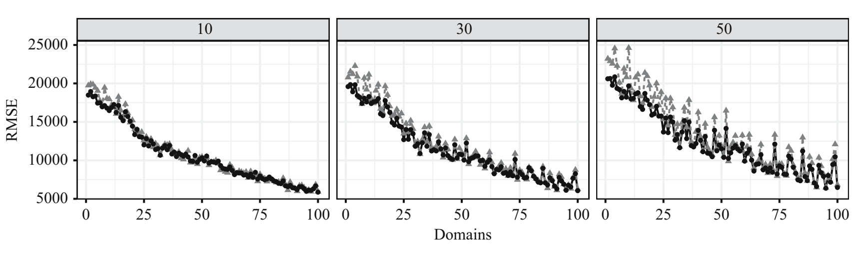

Fig. 1. RMSE of FH.MI estimator per domain for mean setting and varying nonresponse rates. Domains are ordered by increasing sample size.

6. Application to Eurosystem's HFCS

In the following, we provide an example of how the proposed framework can be used for surveys with multiply imputed data in combination with small area methods. The purpose is to show a possible application with the HFCS data for scientists or institutions from relevant research areas rather than to discuss the estimates for each country. The HFCS is a large-scale survey of the financial and consumption situation of European households. The first wave was carried out in 2010 in 15 countries of the European Union (EU). The HFCS contains household data on both economic and demographic variables such as income, wealth, private pension, employment and consumption characteristics (HFCN 2020a). So far three waves have been carried out, the last of which was collected in 2017 and released in March 2020. For the application the third wave is considered. The sample contains about 91,200 households in 22 countries of the EU, between 1,000 and 14,000 households per country. The HFCS is a joint project of several national statistical institutes, Eurosystem national central banks (NCB) and three noneuro area NCBs (Poland, Hungary, Croatia). For these countries, all values are converted into euros by the HFCN (HFCN 2020a). The HFCN asked very sensitive questions, so the item nonresponse rate is high. Missing values in the HFCS data were iteratively and sequentially imputed. The variables are imputed along a path of imputation models. Each model is run several times, and the imputed values from the previous round are treated as given in the subsequent iteration (HFCN 2020a). For each missing observation the HFCS data set contains $M = 5$ imputed values. For more information on the imputation method see HFCN (2020a). Of interest for this application is the value of the household's bonds, which is part of the household's assets and therefore relevant when considering the distribution of wealth. The HFCN reports conditional medians for the value of bonds per EU country (HFCN 2020b). The values are calculated conditioned on households that have bonds; households with no bonds are discarded from the analysis. This results in partly very small sample sizes even on a country level, so that for some countries with fewer than 25 observations direct estimates are not reported by the HFCN. Furthermore, the rate of collected values differs between the countries. Since some households do not even indicate whether they own bonds or not, these values are also imputed by the HFCN. Therefore, the sample size per country, that is, the number of households with bonds and the collected rate for these households, may differ slightly among the five imputed data sets provided by the HFCN. We calculate the sample sizes and collection rates based on the first imputed data sets. An overview of the sample sizes per country and the collected rates are given in Table 5. As dependent variable we choose the

Table 5. Summary of EU-countries sample sizes, collected rates and auxiliary variables.

	Min	1stQ	Median	Mean	3rdQ	Max
Sample size	2.00	12.25	61.50	148.73	209.50	832.00
Collected rate	0.04	0.49	0.66	0.61	0.81	1.00
Total receipts from taxes and social contributions (% of GDP)	23.20	32.83	36.90	36.84	41.85	48.10
Final consumption expenditure (Current prices, EUR per capita)	5630	11710	17170	20424	29258	48140

mean value of bonds in thousand of euros (TEUR) on a country level, resulting in $D = 22$ domains. In 2017 the EU consisted of 28 member states. Six EU members are not included in the HFCS as their noneuro area NCBs do not participate. These domains are considered as out-of-sample (OOS) and model-based estimates are provided in the application. The direct estimators of the mean value of bonds for each imputed data set $\hat{\theta}^{Dir}_{d,m}, d = 1,\ldots,22, m = 1,\ldots,5$ are calculated according to Equation (13) using the sampling weights provided by the data provider, which corrects for potential bias due the sampling design and unit nonresponse. The variances $\sigma^2_{e_{d,m}}$ are estimated with a bootstrap method following the instructions by HFCN (2020a) using the provided replicate weights derived by the Rao-Wu rescaled bootstrap method. As a result we obtain $M = 5$ replicates of direct estimators and their variances, which are then pooled according to Sections 3 and 4.

6.1. Model Selection and Validation

To obtain auxiliary information from additional sources needed for the Fay-Herriot models, country-level data were collected from Eurostat, the statistical office of the EU and the European Commission. Within this set, data such as real estate data, unemployment rates, age dependency ratios, national accounts and tax aggregates from 2011 and 2017 were collected. The sources and years of this supplemental information are shown in Table 6 in the Appendix (Subsection 8.2). Due to the small number of domains, variables that were not available for the entire set of domains were excluded. The remaining auxiliary information includes variables such as the old, youth and age dependency ratio, the unemployment rate, the ratio of taxes to GDP, final consumption expenditure, the share of consumption expenditure on GDP, GDP at market prices and a variable indicating whether the country has a wealth tax. In addition, the number of covariates in the model is severely limited by the small number of domains, which is why we restricted the model to two possible auxiliary variables. In the context of area-level data, Han (2013) transferred the conditional Akaike information in linear mixed models from Vaida and Blanchard (2005) to a conditional Akaike information criterion for Fay-Herriot models. Marhuenda et al. (2014) examine this criterion among Kullback symmetric divergence criterion (KIC) and propose a bootstrap variant of the KIC (KICb2) especially developed for FH models. They conclude that KICb2 criterion is one of the best model selection criteria for Fay-Herriot models. Therefore, in this application the preselection of variables was performed using the KICb2 criterion. Model selection was carried out for each of the five imputed data sets, with no particular difference in the results. A union of two auxiliary variables was selected for the final model, as shown in Table 5. To obtain a model-based estimator of the mean value of household bonds, the estimator from Subsection 4.1 is calculated with the auxiliary information in Table 5. The model variances $\sigma^{2^{RR}}_v$ are calculated for the MI-adjusted Fay-Herriot model on the original scale using the REML method. The distributional assumptions of the model presented in Section 3 are checked by the Shapiro-Wilk test applied to the residuals and the random effects. For the MI-adjusted Fay-Herriot model for a normal mean, the p-values of the tests for the standardized residuals and the random effect are 0.223 and 0.965, respectively. Therefore, the normality assumptions for both error terms cannot be rejected at a 5% significance level. Consequently, all further considerations and results are based on the

MI-adjusted Fay-Herriot model for a mean value as presented in Subsection 4.1. The explanatory power of the model is assessed using the modified R^2 for Fay-Herriot models according to Lahiri and Suntornchost (2015) and we obtain a value of 45%. Due to the low number of domains, it is not possible to include more auxiliary variables to potentially increase explanatory power. We obtain positive estimated regression coefficients for both auxiliary variables. The impact on the tax-to-GDP ratio seems reasonable, given that tax contributions include taxes on wealth (at least in some countries) and that high tax revenues from income could indicate a high level of capital assets. The relationship between consumption and wealth is not independent of income, because if income is higher than consumption, the rest can be invested, and if consumption cannot be covered by income, there is nothing left to invest. Nevertheless, with the given data, the model also shows a positive effect for consumption.

6.2. Small Area Estimates

The estimates of the mean value of bonds on a country level are calculated using the FH.MI estimator for a mean value and to estimate the MSE the MI adapted Prasad-Rao estimator is applied as described in Subsection 4.1. To compare the model-based estimators with a direct estimator, the direct estimators and their variance estimates are computed for each imputed data set as described above and pooled using Rubin's rule in Equation (8) (Direct.RR). The point estimates of the model-based estimators (FH.MI) should be consistent with the unbiased estimates of the direct estimator, but be more precise. Figure 2 compares the direct and the model-based point estimates for the 22 in-sample domains and additionally reports the estimates for the six OOS EU countries. Due to the guidelines of the data provider, the direct estimates for domains with less than 25 observations are not reported. We observe that, for countries with large sample sizes, the direct and model-based estimates are almost identical, consistent with the expectation that high weight is given to the direct estimator when precision is high. An exception is Belgium (BE), where the sample size is rather high, but the shrinkage to the mean quite strong. For most of the direct estimates, which tend to be high, we see that the model-based estimates are smaller, showing the shrinkage effect to the mean of the model-based estimates. (see summary statistics of point estimates in Table 7 in the Appendix (Subsection 8.2)). Possibly due to the low number of covariates very little shrinkage takes place for some countries with small sample sizes (GR, SI, LI). The model-based point

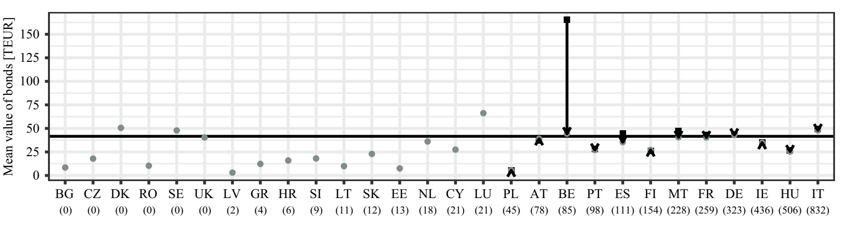

Fig. 2. Direct and model-based estimates for the mean value of bonds, own estimations. Domains are ordered by increasing sample size, sample sizes in brackets. Direct estimates for domains with less than 25 observations are not reported.

estimators are furthermore reported in the map in Figure 3. The highest values are estimated for Luxembourg (LU), followed by Denmark (DK) (OOS) and Sweden (SE) (OOS). For eastern European countries, the estimates are rather low, followed by southern European countries. The estimated model-based values range from EUR 3,000 to EUR 66,000 (cf. Table 7 in the Appendix (Subsection 8.2)), which seems plausible given the median values reported by the HFCN (HFCN 2020b) between EUR 2,000 and EUR 25,000, considering that the distribution at the household level tends to be right skewed and therefore the mean values should be higher than the median values. Figure 4 shows the coefficients of variation (CV) for the direct and model-based estimates. We see that the model-based estimator is at least as efficient as the direct estimator. The CVs of the model-based estimators are mostly significantly smaller than those of the direct estimators, with the effect decreasing with increasing sample size. For large sample sizes, the gain is barely noticeable, but this is consistent with the expected behavior that the direct estimator is sufficiently accurate in this case. For some domains, such as Croatia (HR) and Cyprus (CY), the CV is almost halved. Due to the relatively small domain size of $D = 22$ and hence the limitation to the number of covariates in the model, the efficiency gain is limited.

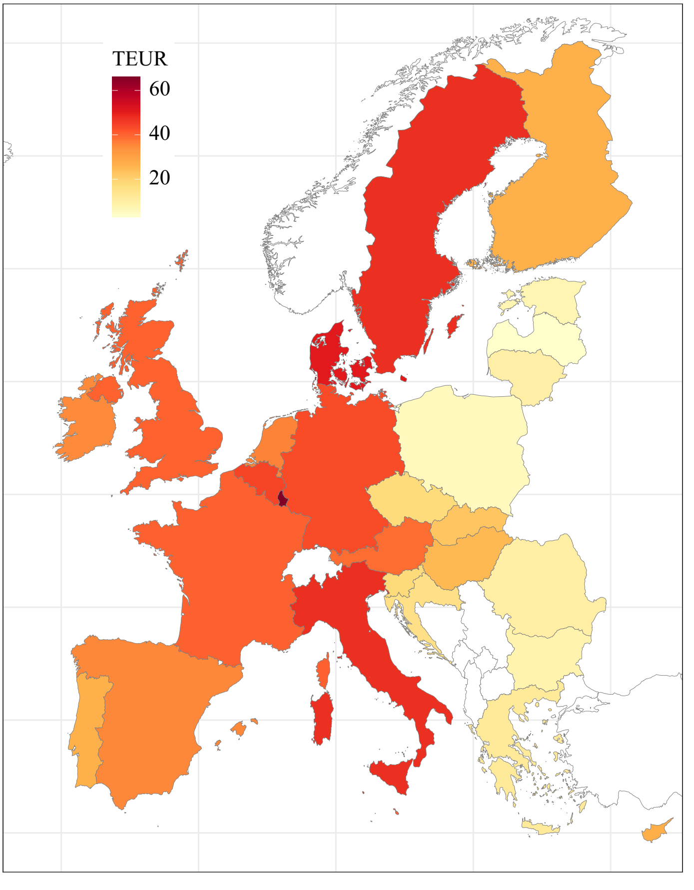

Fig. 3. Map of model-based FH.MI estimates for mean value of bonds, own estimations. Non-EU countries in 2017 are colored in white.

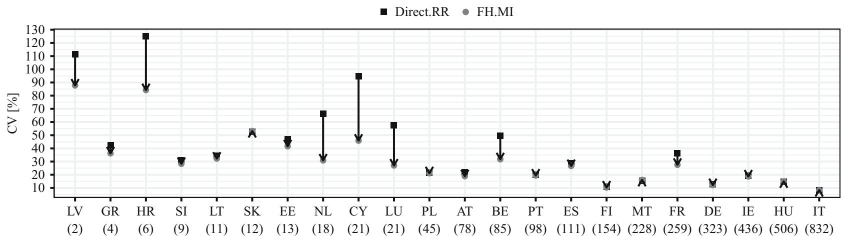

Fig. 4. CVs of direct and model-based estimates, own estimations. Domains are ordered by increasing sample size, sample sizes in brackets.

A summary of the distribution of the point estimators and CVs from Figures 2 and 4 can be found in Table 7 in the Appendix (Subsection 8.2).

7. Concluding Remarks

In this article, we derive small area indicators based on multiply imputed survey data and present uncertainty measures for common cases that capture the additional uncertainty. We present the transformed Fay-Herriot model calculated on each imputed data set. We then combine the components into a MI adjusted Fay-Herriot model that retains the model structure of the Fay-Herriot model. With this approach, results that exist for the Fay-Herriot model regarding transformations, back-transformations and MSE estimators can be extended. It is a general approach that can be applied to any indicator with a given transformation and an appropriate back-transformation. We discuss common special cases of the model (mean, log mean, arcsine ratio). For these special cases we propose MSE estimators. For the mean and logarithmic mean, we present an analytical adaption of the Prasad-Rao estimator and, for the arcsine ratio, we use a bootstrap estimator. We demonstrate in simulation studies that the resulting FH.MI point estimators lead to valid results in terms of bias and RMSE in the given settings and under different nonresponse rates and that the proposed MSE estimators are able to capture the additional imputation uncertainty and lead to good uncertainty measures. We carried out an application using the proposed framework to obtain estimates for European household assets.

A limitation of the proposed approach is that it is not as straightforward for the user as it would be if only the Fay-Herriot estimators were estimated for each imputed data set and the mean value calculated. But, as mentioned above, it is not clear how the variance pooling rules can be applied to the MSE. This could be part of further research. To facilitate the application, it is planned to provide an R-package with the methodology presented. Other open research questions are the extension from a cross-sectional to a longitudinal analysis to provide stable estimates across panel waves (i.e., over time) when multiple imputations are performed and sample sizes are small. If the underlying data structure is a panel survey and individuals or households are observed over multiple time periods, the Fay-Herriot model can be adapted to consider the correlation of the same observations over time. To borrow strength for domain estimates, Rao and Yu (1994) propose a model with auto-correlated random effects and assume an autoregressive process of first order. In addition to the temporal Fay-Herriot models, a multivariate approach could serve the requirement to consider the temporal dimension in the data.

In the multivariate Fay-Herriot model (Benavent and Morales 2016) the domain indicators are estimated simultaneously for the different panel waves. In this way, correlations for both error terms can be considered. These models have not yet been investigated in combination with multiple imputation. The approach in this article could be extended to include correlations over time to ensure reliable estimates over time based on multiply imputed survey data. Since asset values are usually highly skewed, more robust indicators such as the median or other quantiles could be estimated instead of the mean. Therefore, the estimation of small area medians using the Fay-Herriot model would be interesting for future research.

8. Appendix

8.1. MSE Back-Transformation for a Log Mean

Let $\mu = exp(\theta)$ be the true indicator value and $\hat{\theta}$ be an estimate for θ. Furthermore, $\hat{\mu}$ is an estimator for μ with $\hat{\mu} = g(\hat{\theta})$, where g is a continuously differentiable function. For

$$g(\hat{\theta}) = exp\{\hat{\theta} + 0.5\widehat{\text{MSE}}(\hat{\theta})\}$$

an approximation of $\text{MSE}(\hat{\mu})$ using a Taylor expansion can be derived as follows:

$$\begin{aligned}
\text{MSE}(g(\hat{\theta})) &= \text{Var}(g(\hat{\theta})) + \text{Bias}^2(g(\hat{\theta})) \\
&= \text{E}[g(\hat{\theta})^2] - \text{E}[g(\hat{\theta})]^2 + \text{E}[g(\hat{\theta}) - g(\theta)]^2 \\
&\approx \text{E}[\{g(\theta) + g'(\theta)(\hat{\theta} - \theta)\}^2] - \text{E}[\{g(\theta) + g'(\theta)(\hat{\theta} - \theta)\}]^2 + \text{E}[g'(\theta)(\hat{\theta} - \theta)]^2 \\
&= g'(\theta)^2\{\text{E}[\hat{\theta}^2] - \text{E}[\hat{\theta}]^2\} + g'(\theta)^2\text{E}[\hat{\theta} - \theta]^2 \\
&= g'(\theta)^2\{\text{Var}(\hat{\theta}) + \text{Bias}^2(\hat{\theta})\} = g'(\theta)^2\text{MSE}(\hat{\theta}).
\end{aligned}$$

A estimator of $\text{MSE}(\hat{\mu})$ is then obtained by

$$\widehat{\text{MSE}}(\hat{\mu}) = \widehat{\text{MSE}}(g(\hat{\theta})) = g'(\hat{\theta})^2\widehat{\text{MSE}}(\hat{\theta}) = exp\{\hat{\theta} + 0.5\widehat{\text{MSE}}(\hat{\theta})\}^2\widehat{\text{MSE}}(\hat{\theta})\}.$$

8.2. Plots and Tables (Figure 5 and Tables 6–7)

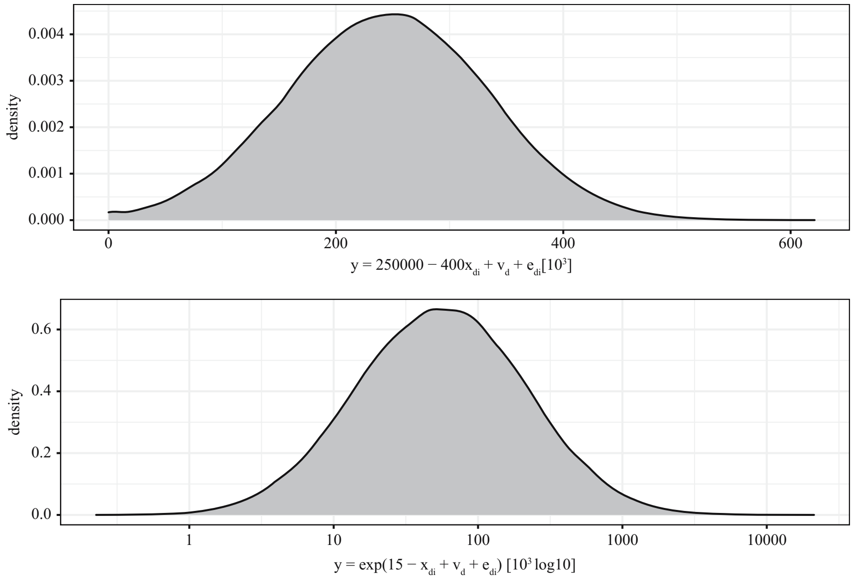

Fig. 5. Density of population target variable of one replication.

Table 6. Source and year of auxiliary information.

	Year	Source
Private households by type, tenure status (Real estate)	2011	Eurostat (2011b)
Dwellings by occupancy status, type of building (Real estate)	2011	Eurostat (2011a)
Age, Old, Young-age dependency ratios	2017	Eurostat (2017d)
Unemployment rate	2017	Eurostat (2017a)
Tax to GDP ratio	2017	Eurostat (2017c)
Final consumption expenditure	2017	Eurostat (2017b)
GDP at market prices	2017	Eurostat (2017b)
Share of consumption expenditure on GDP	2017	Eurostat (2017b)
Indicator for presence of wealth tax	2017	European Commission (2017)

Table 7. Summary of point estimators and CVs for mean value of bonds (TEUR).

Estimator		Min	1stQ	Median	Mean	3rdQ	Max
Direct.RR	Point est.	2.5	19.6	36.2	41.6	49.0	165.5
FH.MI		3.1	16.4	27.3	28.6	40.0	66.2
Direct.RR	CV [%]	8.3	19.3	32.8	41.9	51.7	125.0
FH.MI		8.3	18.8	27.3	31.5	35.3	87.8

9. References

Battese, G., Harter, R., and W. Fuller. 1988. "An error-components model for prediction of county crop areas using survey and satellite data." *Journal of The American Statistical Association* 83: 28–36. DOI: https://doi.org/10.1080/01621459.1988.10478561.

Benavent, R., and D. Morales. 2016. " Multivariate Fay Herriot models for small area estimation." *Computational Statistics and Data Analysis* 94: 372–390. DOI: https://doi.org/10.1016/j.csda.2015.07.013.

Bijlsma, I., Brakel, J., R. van der Velden, and J. Allen. 2020. "Literacy Levels at a Detailed Regional Level: An Application Using Dutch Data." *Journal of Official Statistics* 36: 251–274. DOI: https://doi.org/10.10.2478/jos-2020-0014.

Casas-Cordero, C., J. Encina, and P. Lahiri. 2016. *Poverty Mapping for the Chilean Comunas*: 379–404. DOI: https://doi.org/10.10.1002/9781118814963.ch20.

Chandra, H., K. Aditya, and S. Kumar. 2017. "Small area estimation under a log transformed area level model." *Journal of Statistical Theory and Practice* 12: 497–505. DOI: https://doi.org/10.10.1080/15598608.2017.1415174.

Citro, C.F., and G. Kalton. 2000. *Small-Area Estimates of School-Age Children in Poverty: Evaluation of Current Methodology*. Washington, DC: The National Academies Press. DOI: https://doi.org/10.10.17226/10046.

Destatis. 2018. *Wirtschaftsrechnungen Einkommens-und Verbrauchsstichsprobe Einkommensverteilung in Deutschland 2013*. Statistisches Bundesamt (Destatis). Available at: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Einkommen-Konsum-Lebensbedingungen/Einkommen-Einnahmen-Ausgaben/Publikationen/Downloads-Einkommen/einkommensverteilung-2152606139004.pdf?__blob=publicationFile.

European Commission. 2017. *Taxation and customs, taxes in europe database v3*. Available at: https://ec.europa.eu/taxation_customs/tedb/taxSearch.html (accessed February 2023).

Eurostat. 2011a. *Conventional dwellings by occupancy status, type of building and nuts 3 region*. Available at: http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset = cens_11dwob_r3&lang = en (accessed April 2021).

Eurostat. 2011b. *Private households by type, tenure status and nuts 2 region*. Available at: http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset = cens_11htts_r2&lang = en (accessed April 2021).

Eurostat. 2017a. *Harmonised unemployment rates*. Available at: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset = ei_lmhr_m&lang = en (accessed April 2021).

Eurostat. 2017b. *Main gdp aggregates per capita*. Available at: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset = nama_10_pc&lang = en (accessed April 2021).

Eurostat. 2017c. *Main national accounts tax aggregates*. Available at: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset = gov_10a_taxag&lang = en (accessed April 2021).

Eurostat. 2017d. *Population: Structure indicators*. Available at: http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset = demo_pjanind (accessed April 2021).

Fay, R.E., and R.A. Herriot. 1979. "Estimates of income for small places: An application of James-Stein procedures to census data." *Journal of the American Statistical Association* 74: 269–277. DOI: https://doi.org/10.10.2307/2286322.

Fisher, J. 2006. *Income imputation and the analysis of expenditure data in the consumer expenditure survey*. U.S. Bureau of Labor Statistics, Working Articles. Available at: https://www.bls.gov/opub/mlr/2006/11/art2full.pdf (accessed October 2020).

Gonzalez-Manteiga, W., Lombardia, M.J., I. Molina, D. Morales, and L. Santamaria. 2005. "Analytic and bootstrap approximations of prediction errors under a multivariate Fay-Herriot model." *Computational Statistics and Data Analysis* 52: 5242–5252. DOI: https://doi.org/10.10.1016/j.csda.2008.04.031.

Hadam, S., N. Würz., A.-K. Kreutzmann, and T. Schmid. 2023. "Estimating regional unemployment with mobile network data for functional urban areas in Germany." *Statistical Methods & Applications* DOI: https://doi.org/10.1007/s10260-023-00722-0.

Han, B. 2013. "Conditional Akaike information criterion in the Fay-Herriot model." *Statistical Methodology* 11: 53–67. DOI: https://doi.org/10.10.1016/j.stamet.2012.09.002.

HFCN. 2020a. *The household finance and consumption survey: Methodological report for the 2017 wave*. Household Finance and Consumption Network. Available at: https://www.ecb.europa.eu/pub/pdf/scpsps/ecb.sps35~b9b07dc66d.en.pdf (accessed April 2020).

HFCN. 2020b. *The household finance and consumption survey wave 2017 statistical tables*. Household Finance and Consumption Network. Available at: https://www.ecb.europa.eu/home/pdf/research/hfcn/HFCS_Statistical_Tables_Wave_2017_May2021.pdf?ca15e575b6b7765dad1147e7a3dba728 (accessed April 2020).

Jiang, J., Lahiri, P., S.-M. Wan, and C.-H. Wu. 2001. "Jackknifing in the Fay-Herriot model with an example. "In Proceedings og the seminar." Funding Opportunity in Survey Research, Washington D.C., U.S. Bureau of Labor Statistics. Available at: https://citeseerx.ist.psu.edu/document?repid = rep1&type = pdf&doi = 70c03a8b572a00931735409d7173b087645006b9 (accessed April 2020).

Jiang, J., and J. Rao. 2020. "Robust small area estimation: An overview." *Annual Review of Statistics and Its Application* 7: 337–360. DOI: https://doi.org/10.10.1146/annurev-statistics-031219-041212.

Kreutzmann, A.-K., P. Marek, M. Runge, N. Salvati, and T. Schmid. 2022. "The Fay-Herriot model for multiply imputed data with an application to regional wealth estimation in Germany." *Journal of Applied Statistics* 49: 3278–3299. DOI: https://doi.org/10.10.1080/02664763.2021.1941805.

Kreutzmann, A.-K., Pannier, S., N. Rojas-Perilla, T. Schmid, M. Templ, and N. Tzavidis. 2019. "The R package emdi for estimating and mapping regionally disaggregated indicators." *Journal of Statistical Software* 91: 1–33. DOI: https://doi.org/10.10.18637/jss.v091.i07.

Lahiri, P. and J.B. Suntornchost. 2015. "Variable selection for linear mixed models with applications in small area estimation." *The Indian Journal of Statistics* 77: 312–320. DOI: https://doi.org/10.10.1007/s13571-015-0096-0.

Longford, N. 2004. "Missing data and small area estimation in the uk labour force survey." *Journal of the Royal Statistical Society* 167: 341–373. DOI: https://doi.org/10.10.1046/j.1467-985X.2003.00728.x.

Longford, N.T. 2005. *Missing data and small-area estimation*. Springer, London.

Maiti, T., H. Ren, and S. Sinha. 2014. "Prediction error of small area predictors shrinking both means and variances." *Scandinavian Journal of Statistics* 41: 775–790. DOI: https://doi.org/10.10.1111/sjos.12061.

Marhuenda, Y., D. Morales, and M. Pardo. 2014."Information criteria for Fay-Herriot model selection." *Computational Statistics and Data Analysis* 70: 268–280. DOI: https://doi.org/10.10.1016/j.csda.2013.09.016.

Neves, A., D. Silva, and S. Correa. 2013. "Small domain estimation for the brazilian service sector survey." *Estadistica* 65: 13–37. Available at: https://www.statistics.gov.hk/wsc/CPS003-P7-S.pdf (accessed November 2019).

Pfeffermann, D. 2013. "New important developments in small area estimation." *Statistical Science* 28: 40–68. DOI: https://doi.org/10.10.1214/12-STS395.

Prasad, N., and J. Rao. 1990. "The estimation of the mean squared error of small-area estimators." *Journal of the American Statistical Association* 85: 163–171. DOI: https://doi.org/10.10.2307/2289539.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/.

Rao, J., and M. Yu. 1994. "Small area estimation by combining time series and cross sectional data." *Canadian Journal of Statistics* 22: 511–528. DOI: https://doi.org/10.10.2307/3315407.

Rao, J.N.K. and Molina. 2015. *Small Area Estimation*. John Wiley and Sons, Hoboken.

Riphahn, R. and O. Serfling. 2005. "Item non-response on income and wealth questions." *Empirical Economics* 30: 521–538. DOI: https://doi.org/10.10.1007/s00181-005-0247-7.

Rivest, L.-P., and N. Vandal. 2002. "Mean squared error estimation for small areas when the small area variances are estimated." In Proceedings of International Conference of Recent Advanced Survey Sampling, edited by J. Rao, July 10–13, Ottawa, Canada: 197–206. Available at: https://www.mat.ulaval.ca/fileadmin/mat/documents/lrivest/-Publications/64-RivestVandal2003.pdf (accessed October 2020).

Rubin, D.B. 1976. "Inference and missing data." *Biometrika* 63: 163–171. DOI: https://doi.org/10.10.2307/2335739.

Rubin, D.B. 1987. *Multiple imputation for nonresponse in surveys*. John Wiley and Sons, Hoboken.

Rubin, D.B. 1996. "Multiple imputation after 18 + years." *Journal of the American Statistical Society* 91: 473–489. DOI: https://doi.org/10.10.2307/2291635.

Schmid, T., F. Bruckschen, N. Salvati, and T. Zbiranski. 2017. "Constructing sociodemographic indicators for national statistical institutes by using mobile phone data: estimating literacy rates in Senegal." *Journal of the Royal Statistical Society* 180: 1163–1190. DOI: https://doi.org/10.10.1111/rssa.12305.

Slud, E., and T. Maiti. 2006. "Mean-squared error estimation in transformed Fay-Herriot models." *Journal of the Royal Statistical Society* 68: 239–257. DOI: https://doi.org/10.10.1111/j.1467-9868.2006.00542.x.

Sugasawa, S., and T. Kubokawa. 2017. "Transforming response values in small area prediction." *Computational Statistics and Data Analysis* 114: 47–60. DOI: https://doi.org/10.10.1016/j.csda.2017.03.017.

Sugasawa, S., H. Tamae, and T. Kubokawa. 2017. "Bayesian estimators for small area models shrinking both means and variances." *Scandinavian Journal of Statistics* 44: 150–167. DOI: https://doi.org/10.10.1111/sjos.12246.

Sverchkov, M., and D. Pfaeffermann. 2018. "Small area estimation under informative sampling and not missing at random non-response." *Journal of the Royal Statistical Society*: 981–1008. DOI: https://doi.org/10.10.1111/rssa.12362.

Tzavidis, N., l.-C. Zhang, A. Luna, T. Schmid, and N. Rojas-Perilla. 2018. "From start to finish: a framework for the production of small area official statistics." *Journal of the Royal Statistical Society* 181: 927–979. DOI: https://doi.org/10.10.1111/rssa.12364.

Vaida, F., and S. Blanchard. 2005. "Conditional akaike information for mixed-effects models." *Biometrika* 92: 351–370.

Van Buuren, S. 2018. *Flexible Imputation of Missing Data*. Second Edition. Chapman and Hall/CRC.

Van Buuren, S., and K. Groothius-Oudshoorn. 2011. "mice: Multivariate imputation by chained equation in R." *Journal of Statistical Software* 45: 1–67. DOI: https://doi.org/10.10.18637/jss.v045.i03.

Wang, J., and W.A. Fuller. 2003. "The mean squared error of small area predictors constructed with estimated area variances." *Journal of the American Statistical Association* 98: 716–723. DOI: https://doi.org/10.10.1198/016214503000000620.

You, Y., and B. Chapman. 2006. "Small area estimation using area level models and estimated sampling variances." *Survey Methodology* 32: 97–103. Available at: https://www150.statcan.gc.ca/n1/en/catalogue/12-001-X20060019263 (accessed March 2020).

Received November 2022
Revised March 2023
Accepted June 2023