International Journal of Approximate Reasoning 115 (2019) 134–143

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Estimation of classification probabilities in small domains accounting for nonresponse relying on imprecise probability

Aziz Omar [a,b,c,*], Thomas Augustin [b]

[a] Department of Psychology, University of the German Federal Armed Forces in Munich, Germany
[b] Department of Statistics, LMU Munich, Munich, Germany
[c] Department of Mathematics, Insurance and Applied Statistics, Helwan University, Cairo, Egypt

ARTICLE INFO

Article history:
Received 24 February 2019
Received in revised form 13 September 2019
Accepted 18 September 2019
Available online 24 September 2019

Keywords:
Small area estimation
Missing data
Nonrandom nonresponse
Imprecise Beta model
Generalized empirical Bayes estimation
American National Crime Survey

ABSTRACT

In this paper, we propose two generalized Bayesian imprecise probability approaches for estimation of proportions under potentially nonignorable nonresponse using data from small domains. Our approaches produce reliable inference, refraining from strong assumptions on the response process that are typically not testable. Hence, our estimates avoid the possibly severe bias arising from fallaciously imposing such assumptions. Specifically, we generalize the imprecise Beta model to the small area estimation setting, first treating the missing values in a radically cautious way and then deriving a method that allows incorporating powerfully weak knowledge on the missingness process. Additionally, we extend the empirical Bayes small area estimation approach applied by Stasny [27] through considering a set of priors arising from neighborhood of maximum likelihood estimates of the hyperparameters. As an illustration, we reanalyze data from the American National Crime Survey to estimate the probability of victimization in domains formed by cross-classification of certain characteristics.

1. Background

Censuses and sample surveys are the main sources of data for both official authorities and private sector entities. Since censuses are very time and resources consuming, sample surveys are usually favored to provide estimates for the parameters of interest. A major problem with estimates pertaining to a small domain is that direct estimators based only on domain data usually yield unacceptably large standard errors due to the corresponding small domain sample size. One main reason for the small sample sizes is that survey resources do not allow for accepted accuracy at all levels of domains. To overcome this problem, a set of statistical methods has been developed to estimate reliable domain-specific parameters and is known as Small Area Estimation (SAE) with the terms area and domain being used interchangeably. The main approach of SAE is to compensate insufficient domain sample size through the use of data from other domains in a practice known as *borrowing strength* that is carried out under the assumption of domains similarity. Ghosh and Rao [9], Pfeffermann [19] and Pfeffermann [20] gave timely reviews of SAE techniques and their developments. Rao and Molina [26] gave a comprehensive account of theoretical and practical aspects of SAE.

* Corresponding author at: Department of Statistics, LMU Munich, Munich, Germany.
E-mail addresses: aziz.omar@stat.uni-muenchen.de (A. Omar), augustin@stat.uni-muenchen.de (T. Augustin).

https://doi.org/10.1016/j.ijar.2019.09.006
0888-613X/

Technically, SAE is a field of research where Bayesian approaches are very naturally justified. Assuming that domain-specific parameters follow a common (prior) distribution is a very natural way of "borrowing strength from the ensemble" (Morris [14, p. 47]). Practically, empirical Bayes and hierarchical Bayes are two of the widely used SAE techniques.

In the field of official statistics, the use of SAE techniques was promoted by the US Census Bureau as a means to overcome the problem of undercounting, (cf. Platek et al. [24]). Vividly discussed large scale applications of SAE include also the 2011 population censuses in Germany (cf. Münnich et al. [15,16]), and England and Wales (cf. Baffour et al. [2]). See also Tzavidis et al. [28] for a comprehensive production framework for small area official statistics.

Many surveys suffer from severe nonresponse. In a SAE setting, nonresponse has even a worse consequence, as it may severely reduce the amount of information in the – already small – domain sample, leading to very high variance of the estimators and a severe systematic bias. Classical remedies for nonresponse depend mainly on the implied missingness mechanism defined in the sense of Little and Rubin [10] where assumptions regarding the missingness mechanism and other – usually accompanied – distributional assumptions are required, and even enforced, to achieve point identifiability. For example, weighting and imputations techniques, which are among the common practices, assume nonresponse to occur (conditionally) randomly, i.e. independently of the data generating process given observed covariates. Additionally, modeling the response process along with the target variable through a precise model is the usual approach under the assumption of nonrandom nonresponse. Problematically, such assumptions are usually not testable and fallaciously imposing them may lead to severely biased estimates.[1] Manski [11], among others, argued in favor of an approach that is reluctant towards strong assumptions and thus, capable of producing credible results. To establish such approach, one needs to abandon the point-identifiability mindset towards a partial identifiability mindset, where the sought true value is located within a set or an interval whose limits are attainable without imposing unjustified assumptions. This *cautious* approach to deal with missing data has already been applied to produce credible estimates in the field of official statistics, for example by Manski [12]. In the field of SAE, Plass et al. [23] considered data from the German General Social Survey that suffer from nonresponse and produced cautious versions of some prominent small area estimators based on a generalized likelihood approach.

This criticism of the devastating effects spurious information from wrongly imposed assumptions on the missingness process may have on the analysis is also valid for most Bayesian approaches. It applies not only to direct Bayesian estimation of proportions under assumed missing at random, but ultimately also extends to all approaches where – on whatever level of a hierarchical model – prior probabilities are unjustifiably treated as known in advance. An approach where the underlying prior parameters are not fixed in advance has been proposed by Stasny [27] who developed a hierarchical model connecting the classification process and the response process. Stasny estimated the values of the prior hyperparameters by an empirical Bayes approach. These values are fixed in the subsequent analysis and regarded as the "true" values. Hence, this approach, too, has to be judged as overprecise, eventually.

Against this background, we propose in this article two approaches for a cautious treatment of nonresponse in data from small domains, generalizing the traditional Bayes approaches. Specifically, we estimate the classification probability of a binary variable in small domains using data that suffer from nonresponse. First, we communicate the uncertainty implied through nonresponse as well as prior ignorance by utilizing an appropriately generalized version of Walley [30] imprecise Beta model for a direct generalized Bayesian analysis of our data situation. In this framework, we also apply an imprecise Beta model to the response process itself, which provides us with the opportunity to incorporate additional information on the missingness process derived from the subject matter context. Secondly, we extend Stasny [27] empirical Bayes approach by considering a set of priors consisting of all priors in the neighborhood of the maximum likelihood estimator of the hyperparameters. To illustrate our methods and to compare our results with previous studies, we reanalyze data from the National Crime Survey regarding the victimization status of the residents in certain domains within the United States. These data have been previously analyzed by Stasny [27] and Nandram and Choi [17].

The rest of this article is organized as follows: Section 2 presents the data and outlines previous analyses run on it. Section 3 is devoted to our generalization of the imprecise Beta model to handle nonrandom nonresponse in the SAE setting, where we briefly recall the basic model and discuss a radically cautious treatment of missing data. In Section 4, we derive a cautious direct estimator of the ratio of response probabilities. Combining results from the two previous sections enables us in Section 5 to incorporate powerfully weak knowledge about the response process into our analysis. Section 6 introduces our extension of the empirical Bayes estimation framework, where sets of prior probabilities forming neighborhoods of hyperparameters around Stasny [27] estimator are considered. Section 7 gives some concluding remarks.

2. Notation and basic model

The National Crime Survey (NCS) is a large-scale household survey conducted by the U.S. Census Bureau. In the NCS, members of selected households are asked about crimes committed against them or their properties in the previous six months. Stasny [27] created a subset of the NCS data pertaining to the first half of 1975. The data are post-stratified into 10 domains according to three domain characteristics: (1) urban (U) or rural (R); (2) central city (C), other incorporated place

[1] See for instance the simulation results given in Plass et al. [21, Figure 2] pertaining estimation of proportions.

Table 1
National Crime Survey Data (Stasny [27]) and θ estimates under the naive, empirical Bayes (Stasny [27]) and full Bayes (Nandram and Choi [17]) estimation schemes.

Domain	$\boldsymbol{\mathfrak{y}}$	$\boldsymbol{r}$	$\boldsymbol{n}$	$\hat{\boldsymbol{\theta}}^{(N)}$	$\hat{\boldsymbol{\theta}}^{(E)}$	$\hat{\boldsymbol{\theta}}^{(F)}$
UCL	156	711	815	0.219	0.272	0.270
UCH	95	459	532	0.207	0.265	0.261
UIL	162	719	820	0.225	0.276	0.263
UIH	72	334	370	0.216	0.254	0.244
UNL	92	389	468	0.237	0.305	0.300
UNH	15	55	64	0.273	0.287	0.270
RIL	11	47	54	0.234	0.265	0.256
RIH	10	115	135	0.087	0.185	0.210
RNL	35	309	341	0.113	0.166	0.184
RNH	79	492	556	0.161	0.213	0.217

(I) or unincorporated/not a place (N) and (3) low poverty level (L) (less than 10% of the families are below the poverty level) or high poverty level (H) (at least 10% of the families are below the poverty level). The data are presented in Table 1.

Due to the sensitive nature of certain crimes, it is suspected that not all victims of such crimes would openly report them during the survey. Hence, nonresponse should not be naively treated as occurring randomly. To model this, a two-step model will be considered later.

Let the binary variables Y_{ij} and R_{ij} represent, respectively, the victimization status and the response indicator of household j in domain i, $j = 1, \cdots, n_i$, $i = 1, \cdots, m$, where n_i is the ith domain sample size and m is the number of domains; and y_{ij}, r_{ij} are their observed counterparts. It is of interest to estimate θ_i, the probability that at least one victimization occurred to a household member in domain i during the previous six months, where

$$\theta_i = \mathrm{P}(Y_{ij} = 1), \quad i = 1, \cdots, m, \tag{1}$$

hereafter the victimization probability in domain i. Additionally, let $r_i = \sum_{j=1}^{n_i} r_{ij}$ and $\mathfrak{y}_i = \sum_{j=1}^{n_i} y_{ij}$ represent, respectively, the number of respondents and the number of observed victimization (successes) in the ith domain sample. Finally, $\boldsymbol{n}$, $\boldsymbol{r}$ and $\boldsymbol{\mathfrak{y}}$ represent vectors of their counterparts n_i, r_i and $\mathfrak{y}_i$, respectively.

Stasny [27] considers a two-step model, with a first model for the victimization, and a second model for the response that is based on the true victimization status. Specifically, Y_{ij} is assumed to be independently distributed Bernoulli(θ_i) random variables and nonresponse is modeled to be nonrandom by defining the response probabilities

$$\pi_{it} = \mathrm{P}(R_{ij}|Y_{ij} = t), \quad t = 0, 1. \tag{2}$$

To allow borrowing of strength across different domains, prior similarity among domains regarding both the victimization process and the response process is assumed via the common priors

$$\theta_i \overset{iid}{\sim} \mathrm{Beta}(a, b), \quad \pi_{it} \overset{iid}{\sim} \mathrm{Beta}(\alpha_t, \beta_t). \tag{3}$$

Following the empirical Bayes framework (see, e.g., Casella [5] for a tutorial), the marginal likelihood is first optimized w.r.t. $\boldsymbol{\phi} = (a, b, \alpha_0, \beta_0, \alpha_1, \beta_1)^\intercal$, the vector of hyperparameters, which means, informally speaking, that the space of the hyperparameters is searched for the values that would have best predicted the actually observed sample. The maximum likelihood estimator $\hat{\boldsymbol{\phi}}$ is then received and treated as fixed thereafter to obtain the following empirical Bayes estimator (Stasny [27, p. 299])

$$\hat{\boldsymbol{\theta}}^{(E)} = \left(\sum_{k=0}^{n_i - r_i} T_{ik} \left(\mathfrak{y}_i + a + k / n_i + a + b \right) \right) \left(\sum_{k=0}^{n_i - r_i} T_{ik} \right)^{-1}, \tag{4}$$

where $T_{ik} = \binom{n_i - r_i}{k} \cdot B(\mathfrak{y}_i + a + k, n_i - \mathfrak{y}_i + b - k) \cdot B(\mathfrak{y}_i + \alpha_1, \beta_1 + k) \cdot B(r_i - \mathfrak{y}_i + \alpha_0, n_i - r_i + \beta_0 - k)$.

Nandram and Choi [17] analyzed the same data using similar priors, but they followed the classical precise full Bayesian framework by imposing further prior distributions on $\boldsymbol{\phi}$ assuming further parameters that are fixed on a higher level. They used an MCMC algorithm to sample from the posterior distribution, hence, obtaining the full Bayes estimator $\hat{\boldsymbol{\theta}}^{(F)}$. Table 1 shows a comparison between the naive estimator of $\hat{\boldsymbol{\theta}}^{(N)}$ (obtained using only complete observed data) and the posterior estimators $\hat{\boldsymbol{\theta}}^{(E)}$ and $\hat{\boldsymbol{\theta}}^{(F)}$.

A common characteristic of both previous analyses is that they follow traditional Bayesian frameworks that treat the hyperparameters of the final models as fixed. Additionally, both imposed strict distributional assumptions while modeling the nonresponse. In contrast, we propose two less restrictive and cautious treatments that result in set-valued estimates instead of the usual single-valued estimates. Nevertheless, these set-valued estimates entertain a high degree of credibility as they do not depend on untestable assumptions. Each of the following two sections represents the proposed treatments, respectively.

3. Direct cautious modeling based on the imprecise Beta model

In this section, we discuss how to express both the uncertainty associated with nonresponse and the uncertainty resulting from prior ignorance. We give a brief account of the usual Bayesian estimation of proportion. Then, we elaborate on our first proposed treatment.

Consider for a moment the ideal situation where no nonresponse occurs in the ith domain sample. In this case, $\mathfrak{y}_i$ is equivalent to y_i, the "true" number of successes. Under the traditional Bayesian framework, it is natural to consider the conjugate Beta-Binomial model for θ_i receiving the conjugate posterior mean estimator

$$\hat{\theta}_i|y_{ij} = \frac{y_i + a}{n_i + a + b}. \tag{5}$$

To express uncertainty associated with prior ignorance about the probability of success in a Binomial experiment, Walley [30] introduced the imprecise Beta model (IBM).[2] Unlike the traditional Bayes approach, where a single (vague) prior is utilized, the IBM expresses prior uncertainty by defining a set of Beta prior distributions, connected through a common hyperparameter. This set of priors is to be updated into an equivalent set of Beta posteriors after obtaining the sample.

Adopting the IBM to the SAE setting, we can express uncertainty about each θ_i by defining $\mathcal{M}_\nu$, the set of all Beta distributions that have a fixed value, ν, as the summation of their parameters. After observing y_i successes in the ith domain sample, each prior in $\mathcal{M}_\nu$ is updated into a corresponding posterior that is in $\mathcal{M}_{n_i+\nu}$, the set of all Beta distributions that have $n_i + \nu$ as the summation of their parameters.

Inference about certain events and parameters defined on the convex space of θ_i is attainable in the form of set-valued estimates whose minimum and maximum values are found by optimizing (5) w.r.t. a constraining for $\nu = a + b$. It is straightforward to deduce that

$$\hat{\theta}_i|y_{ij} \in \left(\frac{y_i}{n_i + \nu}, \frac{y_i + \nu}{n_i + \nu}\right). \tag{6}$$

Now, consider the case of nonresponse, where y_i is no longer available. Without forcing any assumptions pertaining the nonresponse process, it is certain that

$$y_i \in [\mathfrak{y}_i, \mathfrak{y}_i + n_i - r_i]. \tag{7}$$

Incorporating (7) in (6), the following set-valued estimator $\hat{\Theta}_i$ can be obtained

$$\hat{\Theta}_i := \left(\underline{\hat{\Theta}}_i, \overline{\hat{\Theta}}_i\right) := \left(\frac{\mathfrak{y}_i}{n_i + \nu}, \frac{\mathfrak{y}_i + n_i - r_i + \nu}{n_i + \nu}\right), i = 1, \cdots, m. \tag{8}$$

The borrowing of strength typical for the SAE context is expressed by the common parametrization of the Beta distribution with parameters (a, b) and confines the vector $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_m)^\intercal$ to be within $\hat{\boldsymbol{\Theta}}$, a proper subset of the cube $\times_{i=1}^m \hat{\Theta}_i$ produced from (8), where

$$\hat{\boldsymbol{\Theta}} := \left\{ \left(\frac{\mathfrak{y}_1 + a}{n_1 + \nu}, \cdots, \frac{\mathfrak{y}_m + n_m - r_m + a}{n_m + \nu}\right)^\intercal \middle| a \in (0, \nu) \right\}. \tag{9}$$

Since, however, $\hat{\boldsymbol{\Theta}}$ cannot be fully described component-wise, in the left panel of Table 2 we only show the component-wise projections, i.e. domain specific set-valued estimates $\hat{\Theta}_i$ for the NCS data (using $\nu = 1$).

Our procedure resembles the approaches of Ramoni and Sebastiani [25], De Cooman and Zaffalon [8] and Utkin and Augustin [29]. It can also be seen as a natural application of cautious data completion discussed in Augustin et al. [1, Section 7.8.2] to adjust for missingness while refraining from any distributional assumptions on the missingness process. It may be noted that the set-valued estimates $\hat{\Theta}_i$ contain, for each domain, the naive estimates, the empirical Bayes estimates and the full Bayes estimates shown in Table 1. This confirms the cautious nature of our approach depicted in its similarity to the approaches mentioned above, where cautious set-valued estimates contain single-valued estimates produced by traditional precise approaches.

4. Cautious estimation of the response ratio

The treatment developed above is radically cautious in the sense that it imposes no assumptions regarding the missingness process and considers all possible values for the missing data. A direct implication is that the set-valued estimates resulting from applying this treatment are characterized by having broad intervals that may not convey insights regarding θ_i that are judged informative enough from the perspective of the intended application. The results shown in Table 2

[2] For a detailed justification of the approach see also the intensive discussion of its multinomial counterpart, the imprecise Dirichlet model, in Walley [31] and Bernard and Ruggeri [4].

Table 2
Interval limits of the posterior expectations $\hat{\Theta}_i$ for the NCS data under the extension of the IBM, range of the response rate $\mathcal{R}_{n_i,r_i,\mathfrak{y}_i}$ and $\hat{\Theta}_i$ assuming a specific $\mathcal{R}^{\star}$.

Domain	$\nu = 1$		$\hat{\mathcal{R}}_i^{\star}(n_i, r_i, \mathfrak{y}_i)$		$\mathcal{R}_i^{\star} < 1$	
	$\underline{\hat{\Theta}}_i$	$\overline{\hat{\Theta}}_i$	$\underline{\mathcal{R}}_i$	$\overline{\mathcal{R}}_i$	$\underline{\hat{\Theta}}_i$	$\overline{\hat{\Theta}}_i$
UCL	0.191	0.320	0.598	1.19	0.220	0.320
UCH	0.178	0.317	0.562	1.20	0.207	0.317
UIL	0.197	0.322	0.614	1.18	0.225	0.322
UIH	0.194	0.294	0.661	1.14	0.216	0.294
UNL	0.196	0.367	0.535	1.27	0.237	0.367
UNH	0.231	0.385	0.600	1.25	0.275	0.385
RIL	0.200	0.345	0.579	1.22	0.236	0.345
RIH	0.074	0.228	0.323	1.20	0.088	0.228
RNL	0.102	0.199	0.515	1.12	0.114	0.199
RNH	0.142	0.259	0.549	1.16	0.161	0.259

confirm this notice. It is then desirable to reach meaningful (i.e. narrower) set-valued estimates. The sensible way to reach them would be by implying some sort of premises regarding the missingness process, in particular as often some weak background knowledge is available which could not be used in a traditional, precise approach, but proves to be powerful in our setting. It seems, then, that a trade off between an assumption-free treatment and informative estimates is inevitable. One can argue, however, in favor of well-grounded assumptions. The incorporation of such tenable assumptions regarding the missingness process to obtain narrower set-valued estimates has been studied by Plass et al. [22] in the likelihood framework.

To make use of such assumptions in our proposed treatment we need to establish a relationship between the unknown frequency y_i and the missingness process. To do so, let $\mathcal{R}_i$ be the response ratio by which households with positive victimization status, i.e. $(Y = 1)$, respond comparing to households with negative victimization status, i.e. $(Y = 0)$, in the ith domain, $i = 1, \cdots, m$. That is,

$$\mathcal{R}_i = \frac{\pi_{i1}}{\pi_{i0}}, \tag{10}$$

where π_{it}, $t = 0, 1$ are the response probabilities defined in (2). As long as the data set contains nonresponse, we are sure that $\mathcal{R}_i > 0$. Furthermore, the case where $\mathcal{R}_i = 1$ refers to the assumption of missing (completely) at random that promotes the use of only the complete data leading to the naive estimates $\hat{\boldsymbol{\theta}}^{(N)}$ in Table 1.

Since the response processes for the victimized and the non-victimized households can be thought of as separate Binomial experiments where $\mathfrak{y}_i \sim \text{Bin}(y_i, \pi_{i1})$ and $r_i - \mathfrak{y}_i \sim \text{Bin}(n_i - y_i, \pi_{i0})$, uncertainty associated with prior ignorance of π_{it} can be expressed through the IBM as was shown previously for θ_i.

Hence, we deduce (in the notation of Stasny [27] with $\nu_1 = \alpha_1 + \beta_1$ and $\nu_0 = \alpha_0 + \beta_0$) the following set-valued estimators for π_{it}, $t = 0, 1$

$$\hat{\pi}_{i1} | y_{ij}, r_{ij} \in \left(\frac{\mathfrak{y}_i}{y_i + \nu_1}, \frac{\mathfrak{y}_i + \nu_1}{y_i + \nu_1} \right) \tag{11}$$

and

$$\hat{\pi}_{i0} | y_{ij}, r_{ij} \in \left(\frac{r_i - \mathfrak{y}_i}{n_i - y_i + \nu_0}, \frac{r_i - \mathfrak{y}_i + \nu_0}{n_i - y_i + \nu_0} \right). \tag{12}$$

The sought relationship between $\mathcal{R}_i$ and y_i can be established by substituting the suitable upper and lower limits of $\hat{\pi}_{i1}$ and $\hat{\pi}_{i0}$ from (11) and (12), respectively, into (10) to obtain the following set-valued estimator of $\mathcal{R}_i$

$$\hat{\mathcal{R}}_i \in \left(\underline{\hat{\mathcal{R}}}_i, \overline{\hat{\mathcal{R}}}_i \right) := \left(\frac{\mathfrak{y}_i}{r_i - \mathfrak{y}_i + \nu_0} \cdot \frac{n_i - y_i + \nu_0}{y_i + \nu_1}, \frac{\mathfrak{y}_i + \nu_1}{r_i - \mathfrak{y}_i} \cdot \frac{n_i - y_i + \nu_0}{y_i + \nu_1} \right), \tag{13}$$

that shows that both $\underline{\hat{\mathcal{R}}}_i$ and $\overline{\hat{\mathcal{R}}}_i$ are strictly decreasing functions of y_i.[3] It is then certain that the limits of $\hat{\mathcal{R}}_i$ are attained at $\underline{\hat{\mathcal{R}}}_i$ and $\overline{\hat{\mathcal{R}}}_i$ shown in (13).

It is worth noting that as long as $\nu_1, \nu_0 > 0$, the following relationship is always satisfied

$$\frac{\mathfrak{y}_i}{r_i - \mathfrak{y}_i + \nu_0} < \frac{\mathfrak{y}_i + \nu_1}{r_i - \mathfrak{y}_i},$$

[3] Unless we let $\nu_1 + \nu_0 = -n_i$, which is an unreasonable choice for ν_1 and ν_0.

which implies that $\hat{\underline{\mathcal{R}}}_i < \hat{\overline{\mathcal{R}}}_i$ for any given value of the unknown frequency y_i. This note, however, does not necessarily apply if we substitute two different values $y_{(1)i} < y_{(2)i}$ of y_i into (13).

Indeed, we are interested in exploiting the strict negative relationship between $\hat{\underline{\mathcal{R}}}_i$ and $\hat{\overline{\mathcal{R}}}_i$ and y_i by substituting $\underline{y}_i = \mathfrak{y}_i$ and $\overline{y}_i = \mathfrak{y}_i + n_i - r_i$ from (7) into $\hat{\underline{\mathcal{R}}}_i$ and $\hat{\overline{\mathcal{R}}}_i$, respectively, to obtain a cautious version of the estimator $\hat{\mathcal{R}}_i$. Applying this substitution results in receiving $\hat{\mathcal{R}}^\star_i(n_i, r_i, \mathfrak{y}_i)$, the cautious version of $\mathcal{R}_i$, that takes the form

$$\hat{\mathcal{R}}^\star_i(n_i, r_i, \mathfrak{y}_i) \in \left(\hat{\underline{\mathcal{R}}^\star}_i, \hat{\overline{\mathcal{R}^\star}}_i\right) := \left(\frac{\mathfrak{y}_i}{\mathfrak{y}_i + n_i - r_i + \nu_1}, \frac{n_i - \mathfrak{y}_i + \nu_0}{r_i - \mathfrak{y}_i}\right). \tag{14}$$

It is notable from (14) that both $\hat{\underline{\mathcal{R}^\star}}_i$ and $\hat{\overline{\mathcal{R}^\star}}_i$ are strictly non decreasing functions in the observed frequency $\mathfrak{y}_i$.[4] Moreover, since

$$\frac{\mathfrak{y}_i}{\mathfrak{y}_i + n_i - r_i + \nu_1} < 1 < \frac{n_i - \mathfrak{y}_i + \nu_0}{r_i - \mathfrak{y}_i}, \quad 0 \le \mathfrak{y}_i \le r_i \le n_i, \quad \nu_1, \nu_0 > 0, \tag{15}$$

we will always have $\hat{\underline{\mathcal{R}^\star}}_i < 1 < \hat{\overline{\mathcal{R}^\star}}_i$, which confirms that the cautious estimator $\hat{\mathcal{R}}^\star_i$ does not exclude the case of missing at random.

The middle panel of Table 2 shows the limits of $\hat{\mathcal{R}}^\star_i$ for all domains using the NCS data (taking $\nu_1 = \nu_0 = 1$). Collectively for all domains, the cautious response rate is limited to the range $(0.661, 1.12)$. Although this range shows an obvious tendency of victimized households to respond less frequently than the non-victimized ones, the assumption of random nonresponse should not be excluded as the value 1 is contained within the interval of $\hat{\mathcal{R}}^\star_i$.

5. Incorporating assumptions regarding the missingness process

The relationship between $\hat{\underline{\mathcal{R}}}_i$ and $\hat{\overline{\mathcal{R}}}_i$, and y_i indicated by (13) can be further exploited to incorporate assumptions regarding the missingness process. To do so, we obtain the following inverse relations between $\hat{\underline{\mathcal{R}}}_i$ and $\hat{\underline{\mathcal{R}}}_i$, and y_i:

$$y_i = \frac{\mathfrak{y}_i(n_i + \nu_0) - \hat{\underline{\mathcal{R}}}_i \nu_1 (r_i - \mathfrak{y}_i + \nu_0)}{\mathfrak{y}_i + \hat{\underline{\mathcal{R}}}_i (r_i - \mathfrak{y}_i + \nu_0)}, \tag{16}$$

$$y_i = \frac{(\mathfrak{y}_i + \nu_1)(n_i + \nu_0) - \hat{\overline{\mathcal{R}}}_i \nu_1 (r_i - \mathfrak{y}_i)}{\mathfrak{y}_i + \nu_1 + \hat{\overline{\mathcal{R}}}_i (r_i - \mathfrak{y}_i)}. \tag{17}$$

Then, utilizing the relationship between $\hat{\underline{\mathcal{R}}}_i$ and $\hat{\overline{\mathcal{R}}}_i$ and y_i, we can obtain (guessed) limits of the unknown frequency y_i under certain assumed values, $\underline{\mathcal{R}}^{(A)}_i$ and $\overline{\mathcal{R}}^{(A)}_i$, of $\hat{\underline{\mathcal{R}}}_i$ and $\hat{\overline{\mathcal{R}}}_i$, respectively. Appendix A shows that for two different y_i guessed values, $y^\star_{(1)i}$ and $y^\star_{(2)i}$ resulting from substituting $\hat{\underline{\mathcal{R}}}_i$ and $\hat{\overline{\mathcal{R}}}_i$ in (16) and (17), respectively, to satisfy $y^\star_{(1)i} < y^\star_{(2)i}$, we should choose $\underline{\mathcal{R}}^{(A)}_i$ and $\overline{\mathcal{R}}^{(A)}_i$ such that

$$\frac{\overline{\mathcal{R}}^{(A)}_i}{\underline{\mathcal{R}}^{(A)}_i} < \frac{(\mathfrak{y}_i + \nu_1)(r_i - \mathfrak{y}_i + \nu_0)}{\mathfrak{y}_i (r_i - \mathfrak{y}_i)}. \tag{18}$$

For $\underline{\mathcal{R}}^{(A)}_i$ and $\overline{\mathcal{R}}^{(A)}_i$ that do not necessarily satisfy (18), the interval of y_i resulting from using (16) and (17) can be represented as $\left(\left(y^\star_{(1)i}, y^\star_{(2)i}\right)\right)$, where the notation $((l, k))$ denotes an interval whose lower (upper) limit is the minimum (maximum) of l and k, $\forall\, l, k \in \mathbb{R}$.

It is possible to choose $\underline{\mathcal{R}}^{(A)}_i$ and $\overline{\mathcal{R}}^{(A)}_i$ to represent a specific assumption regarding the missingness process. Hence, the interval $\left(\left(y^\star_{(1)i}, y^\star_{(2)i}\right)\right)$ will represent y_i values under that assumption.

In Appendix B we show that the limits $y_{(1)i}$ and $y_{(2)i}$ obtained by constraining $\mathcal{R}_i$ to be within $\hat{\underline{\mathcal{R}^\star}}_i$ and $\hat{\overline{\mathcal{R}^\star}}_i$, the cautious limits of $\hat{\mathcal{R}}_i$, will always be within the cautious limits of y_i shown in (7). That is, $\mathfrak{y}_i < y^\star_{(1)i}, y^\star_{(2)i} < \mathfrak{y}_i + n_i - r_i \quad \forall \quad 0 \le \mathfrak{y}_i \le r_i \le n_i$. This means, it is possible to obtain a narrower cautious interval for the unknown frequency y_i under assumptions on the missingness process. Hence, we obtain a narrower cautious interval for $\hat{\Theta}_i$ utilizing the following equivalent form of (8)

[4] Unless we let $\nu_1 = \nu_0 = -(n_i - r_i)$, which is an unreasonable choice for ν_1 and ν_0.

$$\hat{\Theta}_i = \left(\underline{\hat{\Theta}}_i, \overline{\hat{\Theta}}_i\right) := \left(\frac{\operatorname{Min}\left(y^{\star}_{(1)i}, y^{\star}_{(2)i}\right)}{n_i + \nu}, \frac{\operatorname{Max}\left(y^{\star}_{(1)i}, y^{\star}_{(2)i}\right) + \nu}{n_i + \nu} \right). \tag{19}$$

In order to give a concrete application of these results, we propose to utilize the intuitive assumption that victimized households tend to respond less frequently comparing to their non-victimized counterparts, i.e. $\mathcal{R}_i < 1$. Incorporating this assumption does not contradict to the cautious nature of our treatment since it does not impose any hard-to-be-proven restrictions on the data generating process.

The right panel of Table 2 shows the interval limits of the posterior expectations $\hat{\boldsymbol{\Theta}}$ for the NCS data under the extension of the IBM assuming the aforementioned assumption, $\mathcal{R}_i < 1$ (taking $\nu = \nu_1 = \nu_0 = 1$). These estimates have lower limits equivalent to the naive estimates – thus reflecting the increase in the estimated victimization probability under the assumption of nonrandom nonresponse – and upper limits equivalent to the cautious estimates obtained without any assumptions regarding $\mathcal{R}_i$ – reflecting the still cautious nature of the assumption $\mathcal{R}_i < 1$, and hence, have narrower intervals than the cautious assumption-free estimates.

6. Imprecise probability extension of the empirical Bayes approach

As shown in Section 2, traditional empirical Bayes inference founds the analysis on a single fixed value $\hat{\boldsymbol{\phi}}^{\star}$ of a high dimensional hyperparameters $\boldsymbol{\phi}$. We regard this to be over-optimistic and argue that considering a class of prior distributions, instead, is more suitable. Therefore, we propose in this section another framework that can be seen as an extension to the usual empirical Bayes approach.

Recalling the nonrandom nonresponse model of Stasny [27] introduced in Section 2, let the prior distributions of θ_i and π_{it} be characterized by having a compact, convex set of hyperparameters $\boldsymbol{\Phi}$. Let $\hat{\theta}_i(\boldsymbol{\phi})$ be the domain estimator resulting from the direct application of the empirical Bayes approach for a specific value $\boldsymbol{\phi}$ of $\boldsymbol{\Phi}$. With $\hat{\boldsymbol{\phi}}$ as the maximum likelihood estimator of $\boldsymbol{\phi}$ the estimator $\hat{\boldsymbol{\theta}}^{(E)}$ in (4) corresponds to $\hat{\theta}_i(\hat{\boldsymbol{\phi}})$.

Instead of the estimator $\hat{\theta}_i(\hat{\boldsymbol{\phi}})$, we propose the following more credible, set-valued estimator

$$\hat{\Theta}_i(\hat{\boldsymbol{\Phi}}) := \left[\underline{\hat{\Theta}}_i(\hat{\boldsymbol{\Phi}}), \overline{\hat{\Theta}}_i(\hat{\boldsymbol{\Phi}})\right] := \left[\inf_{\boldsymbol{\phi}\in\hat{\boldsymbol{\Phi}}} \hat{\theta}_i(\boldsymbol{\phi}), \sup_{\boldsymbol{\phi}\in\hat{\boldsymbol{\Phi}}} \hat{\theta}_i(\boldsymbol{\phi})\right], \; i = 1, \cdots, m, \tag{20}$$

where $\hat{\boldsymbol{\Phi}}$ is an appropriately specified subset of $\boldsymbol{\Phi}$.

Preserving the principle of borrowing strength in the SAE setting at hand, a simultaneous optimization of $\hat{\theta}_i(\boldsymbol{\phi})$ for all domains w.r.t. $\boldsymbol{\phi}$ results in the set-valued estimator $\hat{\boldsymbol{\Theta}}(\hat{\boldsymbol{\Phi}})$, where

$$\hat{\boldsymbol{\Theta}}(\hat{\boldsymbol{\Phi}}) := \left\{ \left(\hat{\theta}_1(\boldsymbol{\phi}), \cdots, \hat{\theta}_m(\boldsymbol{\phi})\right)^{\mathsf{T}} \middle| \boldsymbol{\phi} \in \hat{\boldsymbol{\Phi}} \right\}. \tag{21}$$

Fundamentally, there are several ways to specify $\hat{\boldsymbol{\Phi}}$. We suggest, as a kind of a generalized empirical Bayes approach, to let $\hat{\boldsymbol{\Phi}}$ be data-dependent by taking $\hat{\boldsymbol{\Phi}}$ to be a defined neighborhood of $\hat{\boldsymbol{\phi}}$. A natural definition of such neighborhood is to rely on a certain threshold δ of the relative marginal likelihood, and thus considering all prior probabilities with a relatively high predictive power for the actually observed sample. That is, let $(\boldsymbol{Y}, \boldsymbol{R})$ represent the data from all domains and take $\hat{\boldsymbol{\Phi}}$ as the convex hull of all values $\boldsymbol{\phi}$ with $\mathrm{f}(\boldsymbol{Y}, \boldsymbol{R}|\boldsymbol{\phi})/\mathrm{f}(\boldsymbol{Y}, \boldsymbol{R}|\hat{\boldsymbol{\phi}}) \geq \delta$. This rule to select a set of hyper priors fits naturally to the general inference framework recently developed by Moral [13], see also Cattaneo [6] already discussing this rule in the context of (dis)continuity of generalized Bayesian updating.

Defining a neighborhood along these lines requires complete knowledge of the behavior of $\mathrm{f}(\boldsymbol{Y}, \boldsymbol{R}|\boldsymbol{\phi})$ over all possible $\boldsymbol{\phi}$. Such knowledge is not straightforwardly attainable, though, especially for complex likelihood functions.

To explain our approach, we base our illustration on the likelihood of the nonrandom nonresponse model of Stasny [27] and we define two neighborhoods with radii in relative terms of the components of $\hat{\boldsymbol{\phi}}$: the first is defined with radius $\epsilon = 0.9 * \hat{\boldsymbol{\phi}}$ and the second is defined such that $\hat{\boldsymbol{\Phi}} := \left\{\boldsymbol{\phi} \middle| 0 < \boldsymbol{\phi} \leq 3 * \hat{\boldsymbol{\phi}}\right\}$.

As a surrogate for (21), the right panel of Table 3 shows for the NCS data the domain specific estimates $\hat{\Theta}_i(\hat{\boldsymbol{\Phi}})$ for each neighborhood. It is worth noting that, by construction of $\hat{\boldsymbol{\Phi}}$ as a neighborhood of $\hat{\boldsymbol{\phi}}$, the estimate $\hat{\Theta}_i(\hat{\boldsymbol{\Phi}})$ contains the empirical Bayes estimates $\hat{\theta}_i(\hat{\boldsymbol{\phi}})$. Observing that domains with small samples (e.g. RIL, UNH and RIH) have set-valued estimates with large interval widths may suggest an effect of the domain sample size. Such effect suggests limitation of borrowing strength using only the response variable, therefore motivates inclusion of auxiliary covariates through an appropriate model. This approach is known as model-based small area estimation (cf. e.g. Datta [7]).

7. Concluding remarks

We discussed cautious treatment of nonresponse using data pertaining to small domains with the aim of estimating proportions utilizing an imprecise probability perspective. First, we extended the IBM through a generalization that allows

Table 3
Interval limits of the posterior expectations $\hat{\Theta}(\phi)$ for the NCS data under the extension of the empirical Bayes approach.

Domain	$\epsilon = 0.9 * \hat{\phi}$		$\hat{\Phi}^{\dagger}$	
	$\underline{\hat{\Theta}}_i(\hat{\Phi})$	$\overline{\hat{\Theta}}_i(\hat{\Phi})$	$\underline{\hat{\Theta}}_i(\hat{\Phi})$	$\overline{\hat{\Theta}}_i(\hat{\Phi})$
UCL	0.269	0.275	0.232	0.317
UCH	0.261	0.269	0.210	0.331
UIL	0.273	0.279	0.236	0.321
UIH	0.248	0.260	0.182	0.345
UNL	0.300	0.310	0.240	0.380
UNH	0.268	0.308	0.102	0.626
RIL	0.244	0.287	0.078	0.633
RIH	0.174	0.197	0.073	0.386
RNL	0.160	0.172	0.103	0.261
RNH	0.209	0.217	0.165	0.275

$^{\dagger}\hat{\Phi} := \left\{\phi | 0 < \phi \leq 3 * \hat{\phi}\right\}$.

handling missing data by cautious data completion and allowing for borrowing strength across domains. Furthermore, by utilizing an IBM also for the response process, we proposed a framework obtaining more informative set-valued estimates by incorporating tenable assumptions like the assumption that victimized households have a lower response probability than the non-victimized ones. Then, we proposed an imprecise probability extension of the empirical Bayes estimation of the nonrandom nonresponse model of Stasny [27] through considering a neighborhood of the hyperparameters, instead of depending on the single-valued empirical Bayes estimate.

Both our extensions can readily form bases for further development to nonresponse treatments refraining from strong assumptions that are hardly testable. From a practical perspective, it is of great interest to consider not only binary but multi-categorical responses. The results developed in Sections 3, 4 and 5 promise to be generalizable indeed, starting from the imprecise Dirichlet model instead of the imprecise Beta model.

A referee kindly pointed us to the Imprecise Hierarchical Dirichlet Model (IHDM) proposed by Benavoli [3], which produces by imprecise hyper models a shrinkage effect. This indeed makes the IHDM readily suitable for the SAE setting and provides an interesting alternative to the starting point of our approach in Section 3. Since, as can be seen in Benavoli [3, Corallary 1], the main inference is still linear in the domain-specific responses, it looks promising to try to extend not only the techniques described in Sections 3, but even our developments from Sections 4 and 5 to the IHDM setting.

For the generalized empirical Bayes approach, an extension to multi-categorical responses seems technically possible as well, but it must not be forgotten that the dimension of the hyperparameters then becomes even substantially larger, further increasing the computational challenges. From a foundational perspective, quite in accordance with the general framework recently proposed by Moral [13], our second approach demonstrated the attractiveness of looking at a combination of traditional empirical Bayes estimation and imprecise probability concepts, aiming at a sort of an imprecise empirical Bayes method. In some sense, depending on the size of the neighborhood, this method lies on a continuum between the overprecise traditional empirical Bayes rule on the one extreme and, on the other extreme, Walley's generalized Bayes rule, sometimes criticized as being rather conservative in updating. As a first step towards a more in-depth study of the merits and properties of this approach, the different ways to construct the different types of neighborhood of empirical Bayes estimates have to be investigated further, theoretically as well as from the computational point of view.

Declaration of competing interest

The authors wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Acknowledgements

This work is a substantially extended version of the proceedings contribution [18], presented at the SMPS/Belief Conference 2018. We thank the reviewers of the SMPS contribution as well as the reviewers of the earlier version of this article for their very helpful remarks and suggestions.

Appendix A

Let $y^{\star}_{(1)i}$ and $y^{\star}_{(2)i}$ be two guessed values of y_i resulting from substituting $\underline{\mathcal{R}}_i^{(A)}$ and $\overline{\mathcal{R}}_i^{(A)}$ using (16) and (17), respectively, such that $y^{\star}_{(1)i} < y^{\star}_{(2)i}$. Then we have

$$\frac{\mathfrak{y}_i(n_i + \nu_0) - \underline{\mathcal{R}}_i^{(A)}\nu_1(r_i - \mathfrak{y}_i + \nu_0)}{\mathfrak{y}_i + \underline{\mathcal{R}}_i^{(A)}(r_i - \mathfrak{y}_i + \nu_0)} < \frac{(\mathfrak{y}_i + \nu_1)(n_i + \nu_0) - \overline{\mathcal{R}}_i^{(A)}\nu_1(r_i - \mathfrak{y}_i)}{\mathfrak{y}_i + \nu_1 + \overline{\mathcal{R}}_i^{(A)}(r_i - \mathfrak{y}_i)},$$

that implies

$$\Big(\mathfrak{y}_i(n_i+\nu_0)-\underline{\mathcal{R}}_i^{(A)}\nu_1(r_i-\mathfrak{y}_i+\nu_0)\Big)\times\Big((\mathfrak{y}_i+\nu_1)+\overline{\mathcal{R}}_i^{(A)}(r_i-\mathfrak{y}_i)\Big)$$
$$<\Big(\mathfrak{y}_i+\underline{\mathcal{R}}_i^{(A)}(r_i-\mathfrak{y}_i+\nu_0)\Big)\times\Big((\mathfrak{y}_i+\nu_1)(n_i+\nu_0)-\overline{\mathcal{R}}_i^{(A)}\nu_1(r_i-\mathfrak{y}_i)\Big),$$

which can be simplified to obtain the condition represented in (18).

Appendix B

Let $\underline{\mathcal{R}}_i^{(A)}$ and $\overline{\mathcal{R}}_i^{(A)}$ be chosen such that

$$\underline{\mathcal{R}}_i^{(A)}=\underline{\hat{\mathcal{R}}^{\star}}_i=\frac{\mathfrak{y}_i}{\mathfrak{y}_i+n_i-r_i+\nu_1},\quad \overline{\mathcal{R}}_i^{(A)}=\overline{\hat{\mathcal{R}}^{\star}}_i=\frac{n_i-\mathfrak{y}_i+\nu_0}{r_i-\mathfrak{y}_i},$$

and $y^{\star}_{(1)i}$ and $y^{\star}_{(2)i}$ be the two guessed values of y_i resulting from substituting $\underline{\mathcal{R}}_i^{(A)}$ and $\overline{\mathcal{R}}_i^{(A)}$ using (16) and (17), respectively. Then we have

$$y^{\star}_{(1)i}=\frac{\mathfrak{y}_i(n_i+\nu_0)-\left(\frac{\mathfrak{y}_i}{\mathfrak{y}_i+n_i-r_i+\nu_1}\right)\nu_1(r_i-\mathfrak{y}_i+\nu_0)}{\mathfrak{y}_i+\left(\frac{\mathfrak{y}_i}{\mathfrak{y}_i+n_i-r_i+\nu_1}\right)(r_i-\mathfrak{y}_i+\nu_0)}=\mathfrak{y}_i+n_i-r_i,$$

and

$$y^{\star}_{(2)i}=\frac{(\mathfrak{y}_i+\nu_1)(n_i+\nu_0)-\left(\frac{n_i-\mathfrak{y}_i+\nu_0}{r_i-\mathfrak{y}_i}\right)\nu_1(r_i-\mathfrak{y}_i)}{\mathfrak{y}_i+\nu_1+\left(\frac{n_i-\mathfrak{y}_i+\nu_0}{r_i-\mathfrak{y}_i}\right)(r_i-\mathfrak{y}_i)}=\mathfrak{y}_i.$$

Since both $y_{(1)i}$ and $y_{(2)i}$ are decreasing functions in $\underline{\mathcal{R}}_i^{(A)}$ and $\overline{\mathcal{R}}_i^{(A)}$, respectively, then choosing $\underline{\mathcal{R}}_i^{(A)}$ such that $\underline{\mathcal{R}}_i^{(A)}>\underline{\hat{\mathcal{R}}^{\star}}_i$ will result in $y^{\star}_{(1)i}$ that satisfies $y^{\star}_{(1)i}<\mathfrak{y}_i+n_i-r_i$. Similarly, choosing $\overline{\mathcal{R}}_i^{(A)}$ that satisfies $\overline{\mathcal{R}}_i^{(A)}<\overline{\hat{\mathcal{R}}^{\star}}_i$ will result in $y^{\star}_{(2)i}$ that satisfies $\mathfrak{y}_i<y^{\star}_{(2)i}$.

References

[1] T. Augustin, G. Walter, F.P.A. Coolen, Statistical inference, in: T. Augustin, F.P.A. Coolen, G. de Cooman, M.C.M. Troffaes (Eds.), Introduction to Imprecise Probabilities, Wiley, 2014, pp. 135–189.

[2] B. Baffour, D. Silva, A. Veiga, C. Sexton, J.J. Brown, Small area estimation strategy for the 2011 census in England and Wales, Stat. J. IAOS 34 (3) (2018) 395–407.

[3] A. Benavoli, Imprecise hierarchical Dirichlet model with applications, in: W. Koch (Ed.), Proceedings of the 16th International Conference on Information Fusion, IEEE, 2013, pp. 1918–1925.

[4] J.-M. Bernard, F. Ruggeri (Eds.), Special Section on the Imprecise Dirichlet Model and Special Section on Bayesian Robustness, Issues in Imprecise Probability, vol. 50, Elsevier, 2009.

[5] G. Casella, Illustrating empirical Bayes methods, Chemom. Intell. Lab. Syst. 16 (2) (1992) 107–125.

[6] M.E.G.V. Cattaneo, A continuous updating rule for imprecise probabilities, in: A. Laurent, O. Strauss, B. Bouchon-Meunier, R.R. Yager (Eds.), International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Springer, 2014, pp. 426–435.

[7] G.S. Datta, Model-based approach to small area estimation, in: C. Rao (Ed.), Handbook of Statistics – Sample Surveys: Inference and Analysis, in: Handbook of Statistics, vol. 29(B), Elsevier, 2009, pp. 251–288.

[8] G. De Cooman, M. Zaffalon, Updating beliefs with incomplete observations, Artif. Intell. 159 (1–2) (2004) 75–125.

[9] M. Ghosh, J. Rao, Small area estimation: an appraisal, Stat. Sci. (1994) 55–76.

[10] R. Little, D. Rubin, Statistical Analysis with Missing Data, 2nd edition, Wiley, 2014.

[11] C.F. Manski, Partial Identification of Probability Distributions, Springer, 2003.

[12] C.F. Manski, Credible interval estimates for official statistics with survey nonresponse, J. Econom. 191 (2) (2016) 293–301.

[13] S. Moral, Learning with imprecise probabilities as model selection and averaging, Int. J. Approx. Reason. 109 (2019) 111–124.

[14] C.N. Morris, Parametric empirical Bayes inference: theory and applications, J. Am. Stat. Assoc. 78 (381) (1983) 47–55.

[15] R. Münnich, J.P. Burgard, M. Vogt, Small Area-Statistik: Methoden und Anwendungen, AStA Wirtsch. Sozialstat. Arch. 6 (3–4) (2013) 149–191.

[16] R. Münnich, S. Gabler, M. Ganninger, J.P. Burgard, J.-P. Kolb, Stichprobenoptimierung und Schätzung im Zensus 2011, Stat. Bundesamt (2012).

[17] B. Nandram, J. Choi, Hierarchical Bayesian nonresponse models for binary data from small areas with uncertainty about ignorability, J. Am. Stat. Assoc. 97 (2002) 381–388.

[18] A. Omar, T. Augustin, Estimation of classification probabilities in small domains accounting for nonresponse relying on imprecise probability, in: S. Destercke, T. Denoeux, M.A. Gil, P. Grzegorzewski, O. Hryniewicz (Eds.), SMPS 2018: Uncertainty Modelling in Data Science, in: Advances in Intelligent Systems and Computing, vol. 832, Springer, 2018, pp. 175–182.

[19] D. Pfeffermann, Small area estimation—new developments and directions, Int. Stat. Rev. 70 (2002) 125–143.

[20] D. Pfeffermann, New important developments in small area estimation, Stat. Sci. 28 (1) (2013) 40–68.

[21] J. Plass, T. Augustin, M. Cattaneo, G. Schollmeyer, Statistical modelling under epistemic data imprecision: some results on estimating multinomial distributions and logistic regression for coarse categorical data, in: T. Augustin, S. Doria, E. Miranda, E. Quaeghebeur (Eds.), ISIPTA '15, in: Proceedings of the Ninth International Symposium on Imprecise Probability: Theories and Applications, Aracne, Rome, 2015, pp. 247–256.

[22] J. Plass, M.E. Cattaneo, T. Augustin, G. Schollmeyer, C. Heumann, Reliable inference in categorical regression analysis for non-randomly coarsened observations, Int. Stat. Rev. (2019).

[23] J. Plass, A. Omar, T. Augustin, Towards a cautious modelling of missing data in small area estimation, in: A. Antonucci, G. Corani, I. Couso, S. Destercke (Eds.), Proceedings of the Tenth International Symposium on Imprecise Probability: Theories and Applications, in: Proceedings of Machine Learning Research, vol. 62, 2017, pp. 253–264.

[24] R. Platek, J.N.K. Rao, C.E. Särndal, M.P. Singh (Eds.), Small Area Statistics: An International Symposium, Wiley, 1987.

[25] M. Ramoni, P. Sebastiani, Robust learning with missing data, Mach. Learn. 45 (2) (2001) 147–170.

[26] J.N. Rao, I. Molina, Small Area Estimation, 2nd edition, Wiley, 2015.

[27] E. Stasny, Hierarchical models for the probabilities of a survey classification and nonresponse: an example from the National Crime Survey, J. Am. Stat. Assoc. 86 (1991) 296–303.

[28] N. Tzavidis, L.-C. Zhang, A. Luna, T. Schmid, N. Rojas-Perilla, From start to finish: a framework for the production of small area official statistics, J. R. Stat. Soc., Ser. A, Stat. Soc. 181 (4) (2018) 927–979.

[29] L.V. Utkin, T. Augustin, Decision making under imperfect measurement using the imprecise Dirichlet model, Int. J. Approx. Reason. 44 (3) (2007) 322–338.

[30] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman and Hall, 1991.

[31] P. Walley, Inferences from multinomial data: learning about a bag of marbles, J. R. Stat. Soc., Ser. B, Methodol. 58 (1) (1996) 3–57.