Dipartimento di Statistica
"Giuseppe Parenti"

WORKING PAPER 2010/13

Small area estimation in presence of nonresponse

Caterina Giusti,
Emilia Rocco

Università degli Studi
di Firenze

Dipartimento di Statistica "G. Parenti" – Viale Morgagni 59 – 50134 Firenze - www.ds.unifi.it

Small area estimation in presence of nonresponse

Caterina Giusti[1], Emilia Rocco[2]

Abstract. In standard survey estimation the problem of nonresponse is well known and a variety of methods exist to adjust for this phenomenon. Less well understood are the effects of nonresponse on small area estimation. In this paper we propose a probability weighted estimation procedure that adjusts for the effect of an informative nonresponse mechanism on the small area mean predictor when a small area model at unit level is adopted. We follow the approach suggested by Pfefferman *et al*. (1998) to compensate for the effect of unequal sample selection probabilities in multilevel models. Since the survey sampler has no control over the response mechanism, our situation is further complicated by the fact that the response probabilities are unknown and need to be estimated. To analyse the performance of the suggested weighted estimation procedure we present the results of several Monte Carlo experiments implemented under different scenarios. These results show that the proposed procedure is effective if the response probabilities are "properly" estimated and, above all, that the nonresponse and small area estimation problems, if both present in a survey, need to be addressed simultaneously.

Keywords: Informative nonresponse; Multilevel pseudo maximum likelihood; Survey weights; Unit level random effects models; Weights scaling.

1. Introduction

Small area estimation has received a lot of attention in recent years due to a growing demand for reliable small area statistics (see Rao 2003 for a review of available methods). Traditional area-specific direct estimators do not provide adequate precision because sample sizes in small areas are seldom large enough. This makes it necessary to employ indirect estimators

[1] Department of Statistics and Mathematics Applied to Economics, University of Pisa, via Ridolfi 10, 56124 Pisa, Italy. Email: caterina.giusti@ec.unipi.it
[2] Department of Statistics "G. Parenti", University of Florence, viale Morgagni 59, 50134 Firenze, Italy. Email: rocco@ds.unifi.it

that borrow strength from related areas, in particular model-based indirect estimators. In addition to the problem of small sample sizes, a further complication may be that not all sampled units respond to the survey, and the probability that a sampled unit responds may be related to the study variable. The work presented here addresses the problem of estimating small area means in the presence of possibly non-random nonresponse using a unit level random effects model.

When multilevel models are estimated from survey data, typically estimators do not make use of the survey weights. However, if the sampling design is "informative" in the sense that the outcome variable is correlated with design variables not included in the model, even after conditioning on the model covariates, standard estimates of the model parameters, and consequently small area estimates, can be severely biased. Standard inference may be biased even when the original sample design is not informative but the presence of nonresponse results in de facto different posterior inclusion probabilities. Assuming that the nonresponse may be viewed as an additional phase of sampling that follows the first phase of sampling determined by the original sample design, the inclusion probabilities are actually the product of the sample selection probabilities and the response probabilities and the survey weights are the reciprocals of these products.

The use of probability weighting procedures that adjust for the effect of an informative sample design on multilevel model parameter estimators has been suggested and studied in depth in the relatively recent literature by Pfefferman *et al*. (1998), Korn and Graubard (2003), Grilli and Pratesi (2004), Asparouhov (2006), Rabe-Hesketh and Skrondal (2006), Carle (2009). In the context of small area estimation, You and Rao (2002) proposed a pseudo-empirical best linear unbiased prediction (pseudo-EBLUP) estimator to estimate small area means by combining a nested error linear regression model and the survey weights. This last solution produces an estimator that is design consistent and that automatically satisfies the benchmarking property without any adjustment, however it assumes that the weights are ignorable for estimation of the model variance components and uses them only to estimate the regression parameters. In this paper we adopt the former approach, in which the survey weights are used for estimating all the model parameters. Thus, we extend the framework developed to adjust for the effect of an informative sample design on multilevel model parameter estimators to the case of unequal response probabilities. Via simulation studies we investigate:

1) the effects of some informative nonresponse mechanisms on the traditional unweighted small area mean predictor;
2) the potential of multilevel pseudo maximum likelihood (MPML) estimation procedure to adjust for the effect of informative nonresponse;
3) how the MPML small area mean predictor works when the true response probabilities (usually unknown) are replaced with estimated response probabilities.

The results of our simulation studies indicate that the MPML estimation procedure may be effective to adjust for the effect of informative nonresponse on small area estimates based on unit level random effects models if the unknown response probabilities included in the weights are "properly" estimated; moreover, our simulation results suggest that in a survey it is opportune to deal simultaneously with the nonresponse and small area estimation problems.

The paper is organized as follows. Section 2 reviews the use of the MPML estimation for unit-level small area models and the nonresponse framework, and develops a small area predictor in presence of nonresponse. The performance of this small area mean predictor is then evaluated in Section 3 through simulation studies considering different scenarios. Concluding remarks are set out in Section 4.

2. Methodological framework

Unit level random effects models are often used in small area estimation to obtain efficient model-based estimators of small area means. Such model-based estimators typically do not make use of the survey weights. In Section 2.1 we briefly summarize the unit-level small area model and show how the survey weights can be used in the estimation process. We also show that the use of a weighted estimation procedure may be induced by nonresponse, and, in section 2.2, we briefly review the weighting adjustment method that is usually applied to compensate for unit nonresponse. In section 2.3 we clarify the concept of weights scaling and finally, in section 2.4, we show the expression of the small area mean estimator suggested in this study.

2.1 Unit-level small area model and MPML estimation procedure

In a general small area model at unit level it is assumed that, given a two-level population with M level 2 units (areas) and N_i level 1 units within the i^{th} area ($i = 1,...M$), the value of the response variable associated with the unit j within the area i, y_{ij}, is generated by a onefold nested error linear regression model $y_{ij} = \mathbf{x}_{ij}^T \mathbf{b} + v_i + e_{ij}$, $j = 1,..., N_i$, $i = 1,...M$, where $\mathbf{x}_{ij} = \left(x_{ij1},...,x_{ijp}\right)^T$ is a fixed covariate vector of dimension p, $\mathbf{b}$ is a fixed $p \times 1$ vector of parameters, v_i and e_{ij} are normally distributed mutually independent terms of error respectively at area and unit level with means zero and variances σ_v^2 and σ_e^2. Small area mean estimation under this model is typically based on the empirical best linear unbiased predictor (EBLUP) computed by substituting maximum likelihood or restricted maximum likelihood (ML or REML) estimates for the unknown model parameters under the hypothesis that sample values obey the assumed population model (Rao, 2003). This assumption is satisfied under not informative sample selection. When it is not satisfied, the estimators of the model parameters and consequently the EBLUP small area mean predictor are biased.

When nonresponse occurs and the response probabilities are related to the target variable even after conditioning on the covariates, the hypothesis that the observed values obey the assumed population model may be violated even if the sample selection mechanism is noninformative. In such cases, ignoring the effect of nonresponse on the distribution of the observed values may bias the inference very severely. The Pseudo Maximum Likelihood (PML) approach developed by Skinner (1989) is the solution usually suggested for informative sample designs: in this paper this approach is extended to the case of informative nonresponse mechanisms.

In the context of multilevel models the implementation of the PML approach is complicated by the fact that the population log-likelihood is not a simple sum of elementary unit contributions, but rather a function of sums across level-2 and level-1 units. The usual marginal population log-likelihood expression is:

$$
\begin{aligned}
l(\theta_1, \theta_2) &= \log\left(\prod_i \int \left\{ \prod_j f\left(y_{ij} | x_{ij}, v_i, \theta_1\right) \right\} \phi(v_i \mid x_i, \theta_2) d(\theta_2) \right) = \\
&= \sum_i \log\left(\int \exp\left\{ \sum_j \log\left(f\left(y_{ij} | x_{ij}, v_i, \theta_1\right)\right) \right\} \phi(v_i \mid x_i, \theta_2) d(\theta_2) \right)
\end{aligned} \qquad (1)
$$

where $f\left(y_{ij} | x_{ij}, v_i, \theta_1\right)$ and $\phi(v_i \mid x_i, \theta_2)$ are the density functions respectively for y_{ij}, and v_i; θ_1 and θ_2 are the parameters to be estimated. Therefore the application of the PML approach requires the knowledge of the survey weights at every level of the population structure, since each sum over the level 2 population units i need to be replaced by a weighted sum using the level 2 weights w_i, and each sum over the level 1 units j need to be replaced by a weighted sum using the level 1 weights $w_{j|i}$. In absence of nonresponse the survey weights coincide with the reciprocal of the sample selection probabilities at the corresponding level ($w_i = 1/\pi_i$ and $w_{j|i} = 1/\pi_{j|i}$).

When nonresponse occurs we may assume that it corresponds to an additional phase of sampling that follows the first phase of sampling determined by the original sample design; then, the survey weights are equal to the reciprocal of the product of the sample selection probabilities and the response probabilities. As we are interested in investigating the effects of nonresponse, we assume here that the sample design is self-weighting and thus that the weights to be used in MPML procedure are function only of the response probabilities. The extension to the case of not self-weighting sample design is straightforward.

Moreover, we assume here that nonresponse concerns only level-1 units. In many situations in which level-2 units are small areas, they usually are institutional units for which the participation to the survey is a duty, or they may be geographic areas or socio-demographic groups that cannot decide, as a whole, to participate or not to the survey. Therefore, we believe that in the context of the small area estimation problems nonresponse usually concerns only the level-1 units, even if the individual response probabilities may depend not only on individual characteristics but also on characteristics of the small areas (level-2 units) which the units belong to. Under these assumptions, as formalized in the next subsection, the weights for the MPML procedure become $w_i = 1$ and $w_{j|i} = 1/p_{ij}$ where p_{ij} denotes the response probability of unit j belonging to area i.

2.2 Response probabilities

Unit nonresponse, a common problem in sample surveys, appears when for a part of the sampled units the data are not observed. Weighting adjustments are usually applied to compensate for unit nonresponse. Treating nonresponse as a second phase sampling from the original sample, weighting adjustment methods operate by increasing the sampling weights of the respondents in the sample, multiplying them by the reciprocal of their response probabilities. Formally, let U be the target population, s a sample drawn from U according to a probabilistic sampling design $p(s)$, and π_j the inclusion probability of unit j, for all $j \in U$. When nonresponse occurs, we only observe the values of the study variable for the units in a subset $r \subseteq s$. According to the two-phase sampling theory, the survey data may be considered as the results of a two-step process: (1) selection of the sample using a sampling design $p(s)$ having the usual properties $p(s) \geq 0$ for all $s \in S$ and $\sum_S p(s) = 1$, where *S* is the set of all possible samples *s*; (2) given *s*, selection of a sub-sample of respondents *r* through the response probability $p(r|s)$. Unlike the sampling selection, the survey sampler has no control over the response mechanism. Nevertheless, it is usually assumed that $p(r|s) \geq 0$ for all $r \in R_s$ and $\sum_{Rs} p(r|s) = 1$, where R_s is the set of all the possible sub-samples of respondent units given sample *s*. Another common hypothesis is that the units respond independently from each other and from *s*, that is:

$$p(r|s) = \prod_{j \in s} p_j^{\delta_j} (1 - p_j)^{(1-\delta_j)}$$

where p_j and δ_j are respectively the individual response probability and the individual response indicator for unit *j*.

Given these assumptions, the survey weight of unit j, for all $j \in U$, may be defined as $w_j = \frac{1}{\pi_j p_j}$. Specifically, in our situation where the nonresponse concerns only the level-1 units, the first level survey weights become $w_{j|i} = \frac{1}{\pi_{j|i} p_{ij}}$ where p_{ij} denotes the response probability of unit j belonging to area i. As already underlined in the previous subsection, when the sample design is self-weighting the first level weights for the MPML procedure reduce to $w_{j|i} = 1/p_{ij}$.

Since response probabilities are usually unknown, they must be estimated using the available information. The simplest and maybe most diffuse way to estimate individual response probabilities consists in partitioning the sampled units in "weighting classes", assumed homogeneous with respect to the mechanism of response, and then in estimating response probabilities as rates of respondent units within each class. Another common way to estimate individual response probabilities is by expressing them as a logit function of a set of known variables. Many other response probabilities estimation methods are present in the literature (among others see Lessler and Kalsbeek, 1992; Lundström and Särndal, 2005); a detailed discussion of these methods is beyond the scope of the current paper.

2.3 Weights scaling

One of the key issues in the multilevel weighted estimation literature is that the parameter estimators are usually only approximately unbiased, i.e. they are unbiased for sufficiently large cluster (level 2 units) sample sizes, but can be severely biased when the cluster sample sizes are small. For this reason, different scaling methods of the weights have been proposed (Pfefferman *et al*., 1998; Stapleton, 2002; Asparouhov, 2006; Korn and Graubard, 2003).

In multilevel modelling scaling of the weights consists in multiplying the weights by cluster specific scaling constants so that the sum of the weights at the cluster level is equal to some cluster characteristic. Different scaling methods, however, may have different effects on the estimation technique and model parameters. Until now there have been no theoretical results to support one scaling method over another. In this paper we do not wish to discuss the relative merits of the various scaling methods; therefore, in the simulation studies discussed in section 3, we consider the two most popular scaling methods:

Scaling method 1. $w_{j|i}^{scaling1} = \frac{w_{j|i}}{\overline{w}_i}$ with $\overline{w}_i = \left(\sum_{j \in r_i} w_{j|i}\right) / n_{ri}$, where r_i is the set of responding units in area *i*, so that for the i^{th} area the sum of the scaled weights equals the number on respondents n_{ri} in the area;

Scaling method 2. $w_{j|i}^{scaling2} = \frac{w_{j|i}}{\tilde{w}_i}$ where $\tilde{w}_i = \left(\sum_{j \in r_i} w^2{}_{j|i}\right) / \sum_{j \in r_i} w_{j|i}$ so that in each area the sum of the scaled weights equals the "effective respondent set size" $n_{r0i} = \left(\sum_{j \in r_i} w_{j|i}\right)^2 / \left(\sum_{j \in r_i} w^2{}_{j|i}\right)$, which correspond to the "effective sample size" defined by Potthoff et al. (1992) in the context of informative sampling.

2.4 Proposed small area mean estimator

The expression of the small area mean estimator used in this study is:

$$\hat{\bar{y}}_i = N_i^{-1}\left[\sum_{j \in r_i} y_{ij} + \sum_{j \in U_i - r_i} \hat{y}_{ij}\right] \tag{2}$$

where r_i and $U_i - r_i$ are respectively the set of respondents and the set of non-sampled plus nonrespondent units in area *i*, and $\hat{y}_{ij} = x_{ij}^T \hat{\mathbf{b}} + \hat{v}_i$ with $\hat{\mathbf{b}}$ and $\hat{v}_i$ obtained using the MPML estimation procedure with weights $w_i = 1$ and $w_{j|i} = 1/p_{ij}$ where the $w_{j|i}$ can be unscaled or scaled, and the response probabilities p_{ij} may be true or estimated.

3. Simulation studies

3.1 Description

To illustrate the bias of the small area estimators that can occur when ignoring an informative response mechanism and to assess the performance of the MPML estimation procedure based on the true or estimated response probabilities, we designed three simulation studies hereafter called study A, study B, and study C. Each simulation study consists of the following steps:

1. Generate area indexes $i = 1,\dots, M = 30$, and population sizes $N_i = \text{int}(70\exp(0.7 + 0.22\tilde{u}_i))$, with $\tilde{u}_i$ generated from $N(0,\omega^2)$ truncated below by -1.5ω and above by 1.5ω; for $\omega = 10$ the N_i lie in the range [70, 126].
2. Generate the population random area effects, $u_i \sim N(0,\sigma_u^2), i = 1,\dots,30, \sigma_v^2 = 200$ and the covariates $\mathbf{x}_{ij} = (1, 10\times\text{int}[i - 30/3 \times \text{int}(3/30 \times ii)]/10 + 20\varepsilon_{ij})^T$, assuming $\varepsilon_{ij} \sim U(0,1)$. This rather complicated formula for generating the auxiliary variables follows the one used by Pfefferman and Sverchkov (2007) and guarantees that the covariates are the same in each of the three groups of areas, except for the random disturbances ε_{ij}. The three groups consist respectively in areas $1 \le i \le 10$, areas $11 \le i \le 20$ and areas $21 \le i \le 30$.
3. Generate the y values according to the model defined in section 2, with $\beta = (10,30)$ and $e_{ij} \sim N(0,\sigma_e^2), \sigma_e^2 = 200$.
4. Associate to each level-1 unit a response probability as follows: in study A for each unit in each area the response probability is obtained through an exponential function of $z_{ij} = y_{ij} - \mathbf{x}_{ij}^T\boldsymbol{\beta}$; in study B we split the areas into 4 groups using the quartiles of the random area effects distribution, and in each group the response probabilities are generated through an exponential function of the z_{ij} values but the parameters of this function change from a group to another; in study C we proceed as in study B but the exponential function used to generate the nonresponse is assumed to depend only from the individual random effects e_{ij}. In all the studies the parameters of the nonresponse generating function are chosen to produce an expected overall population response rate of about 0.7.
5. Select a stratified sample of the first level units with strata equal to the second level units

and a sampling fraction equal to 0.1 in each stratum.

6. Classify each level-1 unit in the sample as respondent or not respondent carrying out for each of them a Bernoulli experiment.
7. Repeat steps 2-6 1000 times.

In study A for each set of respondents, we computed the following six predictors of the area means:

a) the standard (unweighted) EBLUP estimator calculated on the set of respondents;
b) the MPML predictor (2) with weights computed using the true response probabilities;
c) the MPML predictor (2) with weights computed using response probabilities estimated with the weighting within cells method and using the z_{ij} [3] values to define the cells;
d) the MPML predictor (2) with weights computed using response probabilities estimated with a logit model function of the z_{ij};
e) the MPML predictor (2) with weights computed using response probabilities estimated as in point (d), but assuming as explicative in the logit model $z_{ij} + \eta_{ij}$, with $\eta_{ij} \sim N(0,100)$;
f) the standard (unweighted) EBLUP estimator calculated on the entire sample.

For study B and study C we compute the same predictors with the exception that the estimator described at point (e) is replaced by:

g) the MPML predictor (2) with weights computed using response probabilities estimated with a logit model assuming as covariate not only z_{ij} but also a categorical variable that identifies the groups of areas with different response mechanisms.

In all the three studies and for all the probability weighting estimation procedures, unscaled and scaled weights (method 1 and method 2) are used.

For each case (response mechanism × estimation procedure × scaling method) and each area we consider the following performance indexes. The percentage relative bias (RB):

$$RB_i = \frac{\sum_{h=1}^{1000} \left(\hat{\bar{y}}_{hi} - \bar{y}_{hi} \right) \Big/ 1000}{\sum_{h=1}^{1000} \bar{y}_{hi} \Big/ 1000} \times 100$$

[3] We are aware that in real situations the z_{ij} values are unknown and the response probabilities are estimated using auxiliary information that approximate them. Thus, the use of the z_{ij} values here may be viewed as a perfect approximation. Note however that in our analyses we also show what happens when the z_{ij} values are perturbed, that is when the available auxiliary information is not perfectly predictive of the response mechanism.

where $\hat{\bar{y}}_{hi}$ and $\bar{y}_{hi}$ are respectively the predicted and the true mean of y in area i for the h replica, and the Relative Root Mean Squared Error (RRMSE), that is

$$RRMSE_i = \frac{\sqrt{\sum_{h=1}^{1000}\left(\hat{\bar{y}}_{hi} - \bar{y}_{hi}\right)^2 \Big/ 1000}}{\sum_{h=1}^{1000}\bar{y}_{hi} \Big/ 1000} \times 100$$

3.2 Results

To compute the estimates of interest we used the SAS NLMIXED Procedure. Table 1 shows, for each weighted estimation procedure, the mean[4] of the RB_i values across the areas.

Table 1. Mean of the percentage relative biases RB_i of the small area mean predictor across the areas.

Predictor	***Unscaled***	***Scaled 1***	***Scaled 2***
		Study A	
b	-0.281	-0.426	-0.454
c	-0.304	-0.454	-0.480
d	-0.278	-0.425	-0.450
e	-0.451	-0.587	-0.605
		Study B	
b	-0.306	-0.330	-0.348
c	-0.701	-0.762	-0.776
d	-0.700	-0.763	-0.768
g	-0.298	-0.330	-0.344
		Study C	
b	-0.315	-0.259	-0.270
c	-0.906	-0.893	-0.891
d	-0.908	-0.896	-0.893
g	-0.358	-0.305	-0.303

A first evident result is that the performance of scaling method 2 is always worse or at the most equivalent to the performance of scaling method 1: therefore, hereafter all

[4] The mean of the relative biases is an appropriate performance index here since all the RB values are negative.

comparisons will address only the other two methods. In study A the mean predictors based on unscaled weights are less biased than those based on scaled weights (method 1). This happens also in study B, but the differences are smaller. In study C the order is reversed. We think that these results are related to the role of the area random effects on the response mechanism: whereas in Study A the function used to generate the response mechanism is continuous respect to the v_i values, in study C it is a step function of the v_i values. In study B, for which the results are intermediate, the response mechanism is a piecewise continuous function of the v_i values.

To further investigate the performance of the small area mean estimators, Figure 1, Figure 2 and Figure 3 show respectively for study A, study B, and study C the areas percentage relative biases of the six considered predictors ((a), (b), (c), (d), (e) or (g), (f)). In all the figures the predictor (f), which corresponds to the hypothesis of complete responses, is considered and shown as benchmark.

From all the figures it is evident that an informative response mechanism may induce a significant bias in the estimation of the small area means if the hierarchical regression model is fitted using the standard ML estimation method (case a). The bias can be reduced in an effective manner by the probability-weighted estimation procedure (MPML), assuming unrealistically that the response probabilities are known (case b).

Figure 1 provides evidence also of the reduction of bias that occurs if auxiliary variables predictive of the response behaviour are available and the unknown response probabilities are estimated through a logit model (cases d and e). Obviously, the more predictive of the response mechanism are the available auxiliary variables, the greater is the bias reduction (case d versus e). In particular, when the response mechanism (conditionally to the auxiliary variables) becomes fully ignorable, the estimated response probabilities produce a bias reduction equivalent to that obtained with the true response probabilities. The performance of the weighting within cells method (case c) is equivalent to the performance of true response probabilities when based on auxiliary variables predictive of the response behaviour.

The two response probabilities estimation methods using as covariates only the z_{ij}, parametric the first, non-parametric the second (weighting class method), appear equivalent not only in study A, in which they have a good performance, but also in studies B and C, in which they do not reduce in a significant manner the bias of the traditional EBLUP estimator.

Figure 1 – Percentage Relative Bias for all estimators and small areas: Study A – unscaled (top), scaling 1 (center) and scaling 2 (bottom) weighting method.

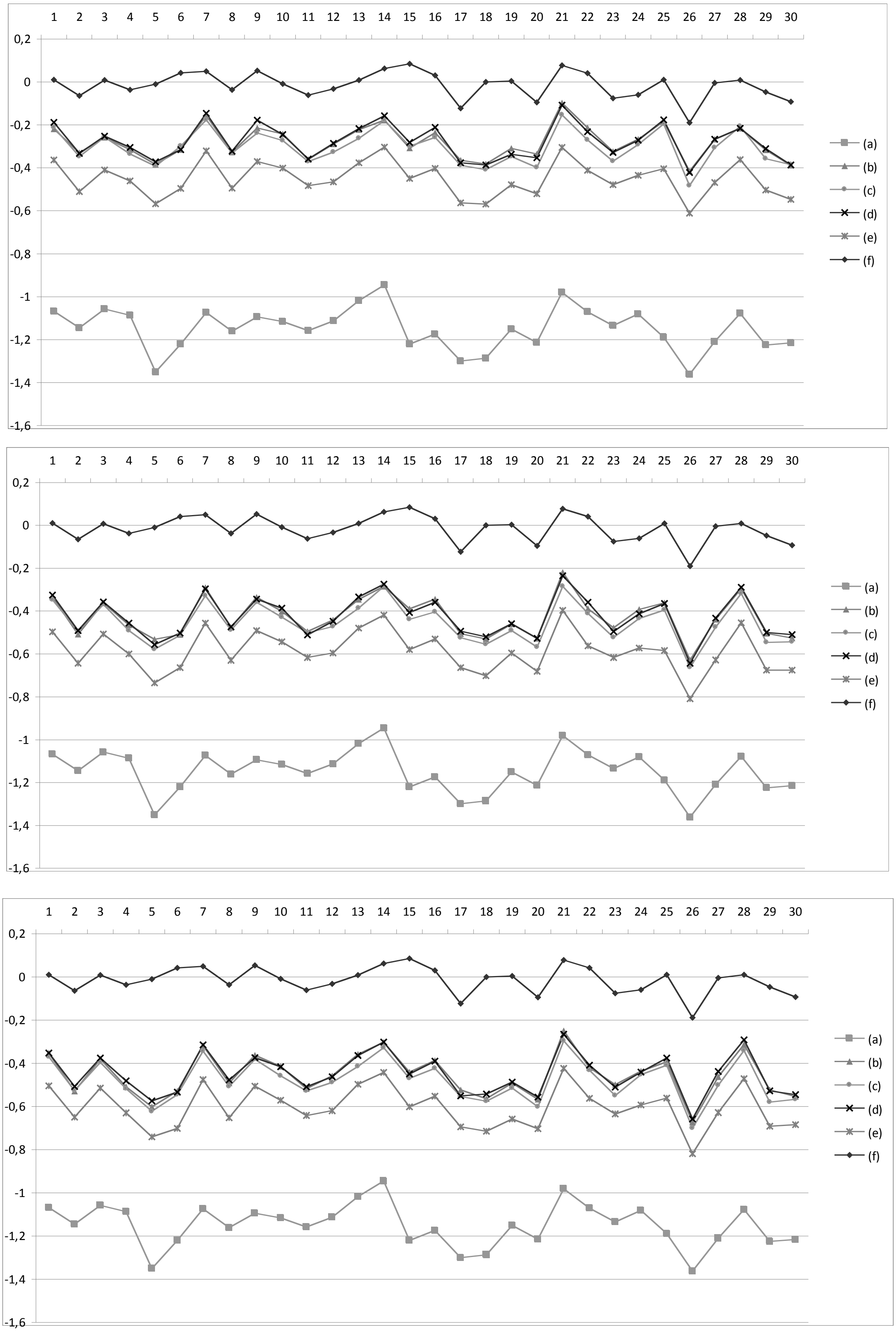


Figure 2 – Percentage Relative Bias for all estimators and small areas: Study B – unscaled (top), scaling 1 (center) and scaling 2 (bottom) weighting method.

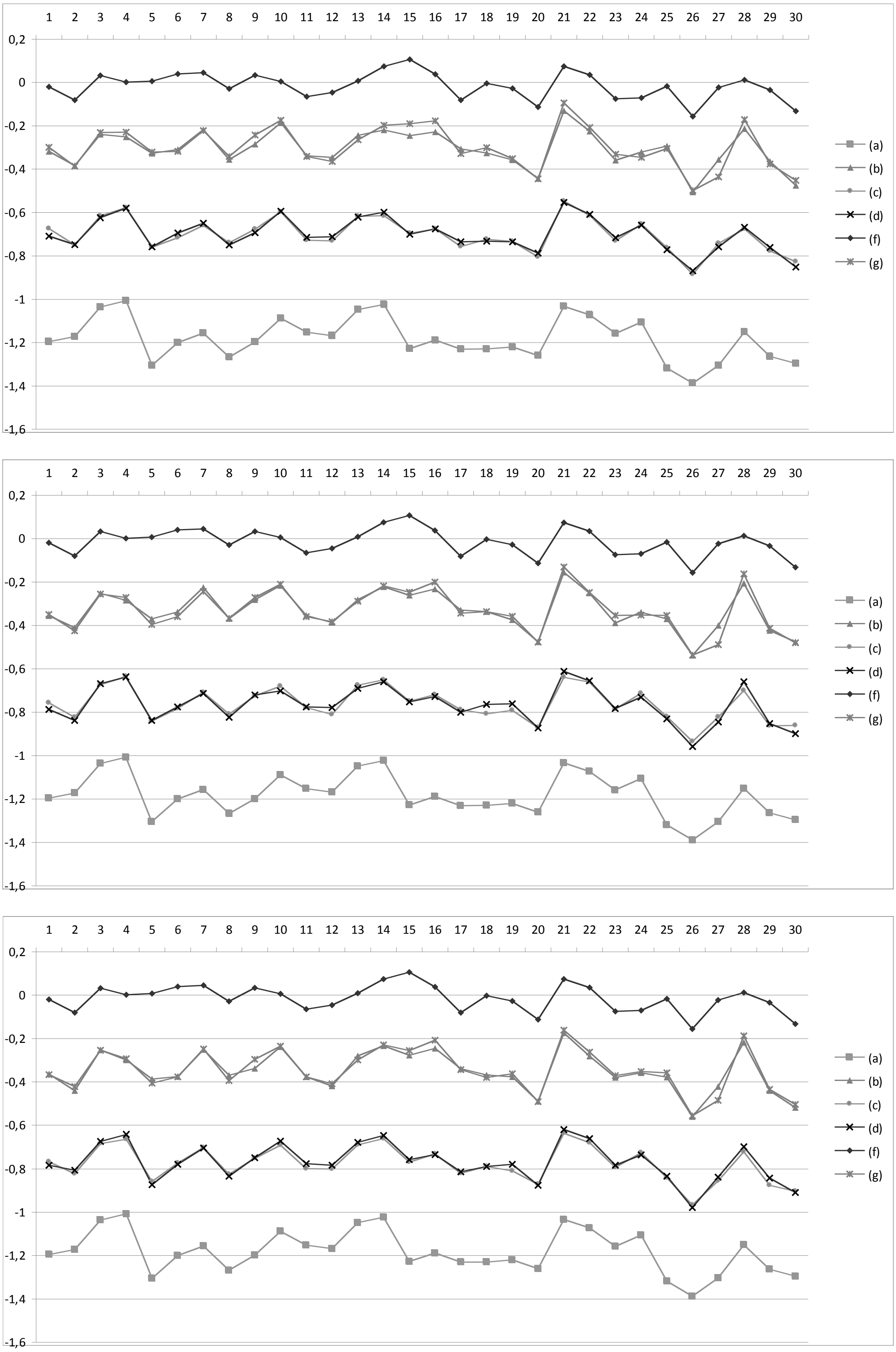
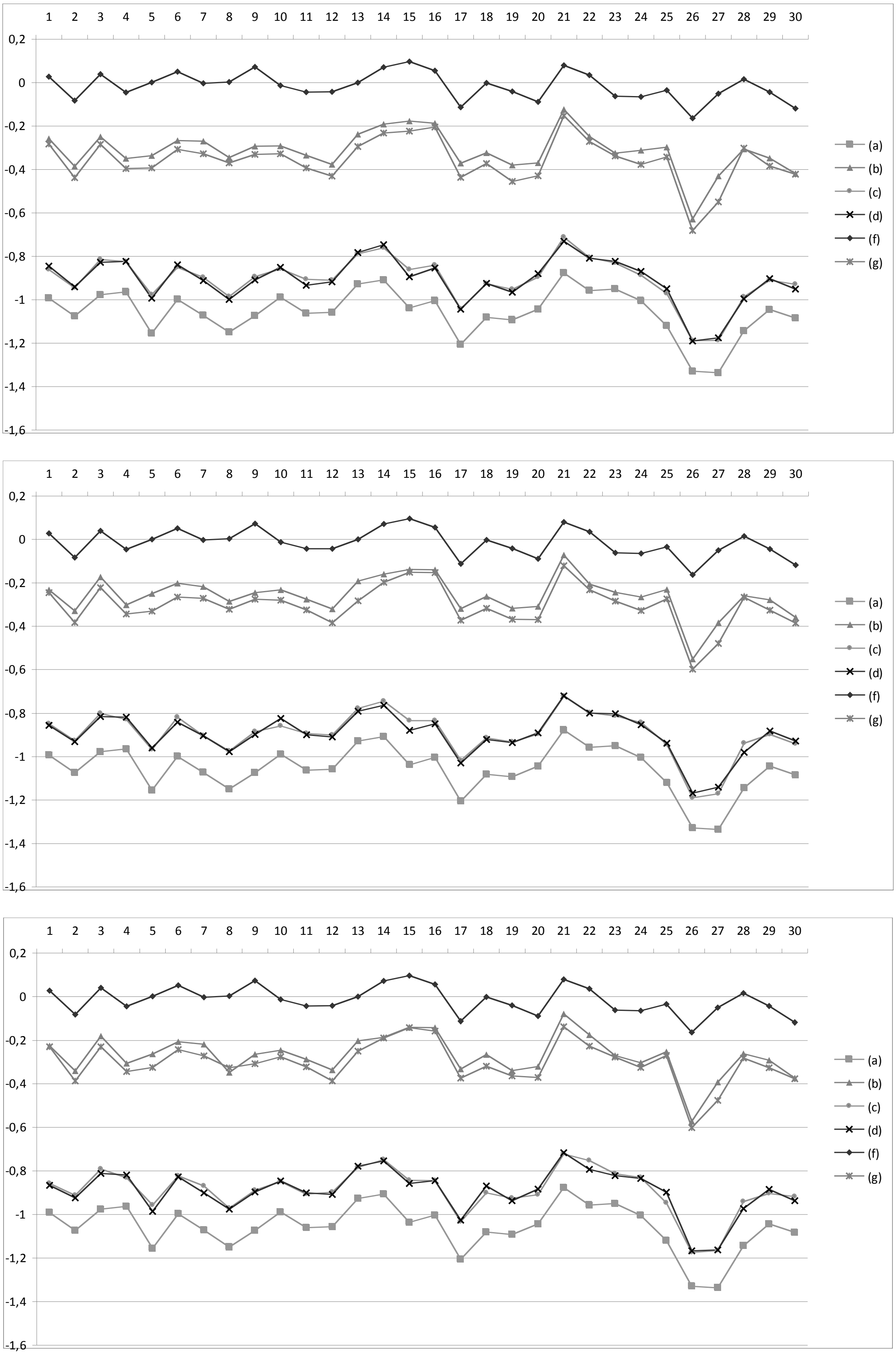


Figure 3 – Percentage Relative Bias for all estimators and small areas: Study C – unscaled (top), scaling 1 (center) and scaling 2 (bottom) weighting method.



In these two studies, to have a good performance of the suggested MPML predictor it is necessary to include in the estimation process of the response probabilities a categorical variable that identifies the groups of areas with different response mechanisms. The advantage of introducing this categorical variable in the procedure to estimate the response probabilities in studies B and C is obvious. The point in question is the following: both the response estimation procedures that use only the z_{ij} values almost remove the bias of the whole population mean direct estimator $\hat{\bar{y}} = \frac{\sum_r y_{ij} / \hat{p}_{ij}}{\sum_r 1/\hat{p}_{ij}}$ [5] (see Table 3).

Table 3. Percentage Relative Bias of the whole population mean direct estimator.

	Response estimation method		
Study	$\hat{p}_{ij} = 1$	Logit model	Weighting within cell
A	-1.833	-0.140	-0.045
B	-1.205	-0.153	-0.182
C	-0.473	-0.075	-0.038

Thus, if the researcher who calculates the survey weights is not interested in the small area estimation problem, he may not realize this advantage. In other words, compensating for nonresponse using a method that works well for the estimation of the overall population mean without considering the estimation at the small area level may reduce or not reduce the bias of the small area mean predictions (predictors c and d in the different studies). This depends on the compensation method but also on the response mechanism.

From Table 3 it is also evident that in Study C the bias of the whole population direct estimator with $\hat{p}_{ij} = 1$ (not adjusted for nonresponse) is less than half of the bias of the corresponding small area mean unweighted predictor (see Figure 3). This result indicates that an informative response mechanism may have a modest effect on population estimators, having at the same time a significant effect on small areas estimators.

Figure 4, Figure 5 and Figure 6 show for each area the Relative Root Mean Squared Error of the six small area mean predictors for the same settings (studies and weighting methods) considered respectively in Figure 1, Figure 2 and Figure 3.

[5] The weights in the population mean direct estimator are equal to $1/\pi_{ij}\hat{p}_{ij}$ with $\pi_{ij} = \pi_{j|i}\pi_i$; in this expression we have only $1/\hat{p}_{ij}$ due to the sample self-weighting hypothesis.

Figure 4 – Percentage Relative Root Mean Squared Error for all estimators and small areas: Study A – unscaled (top), scaling 1 (center) and scaling 2 (bottom) weighting method.

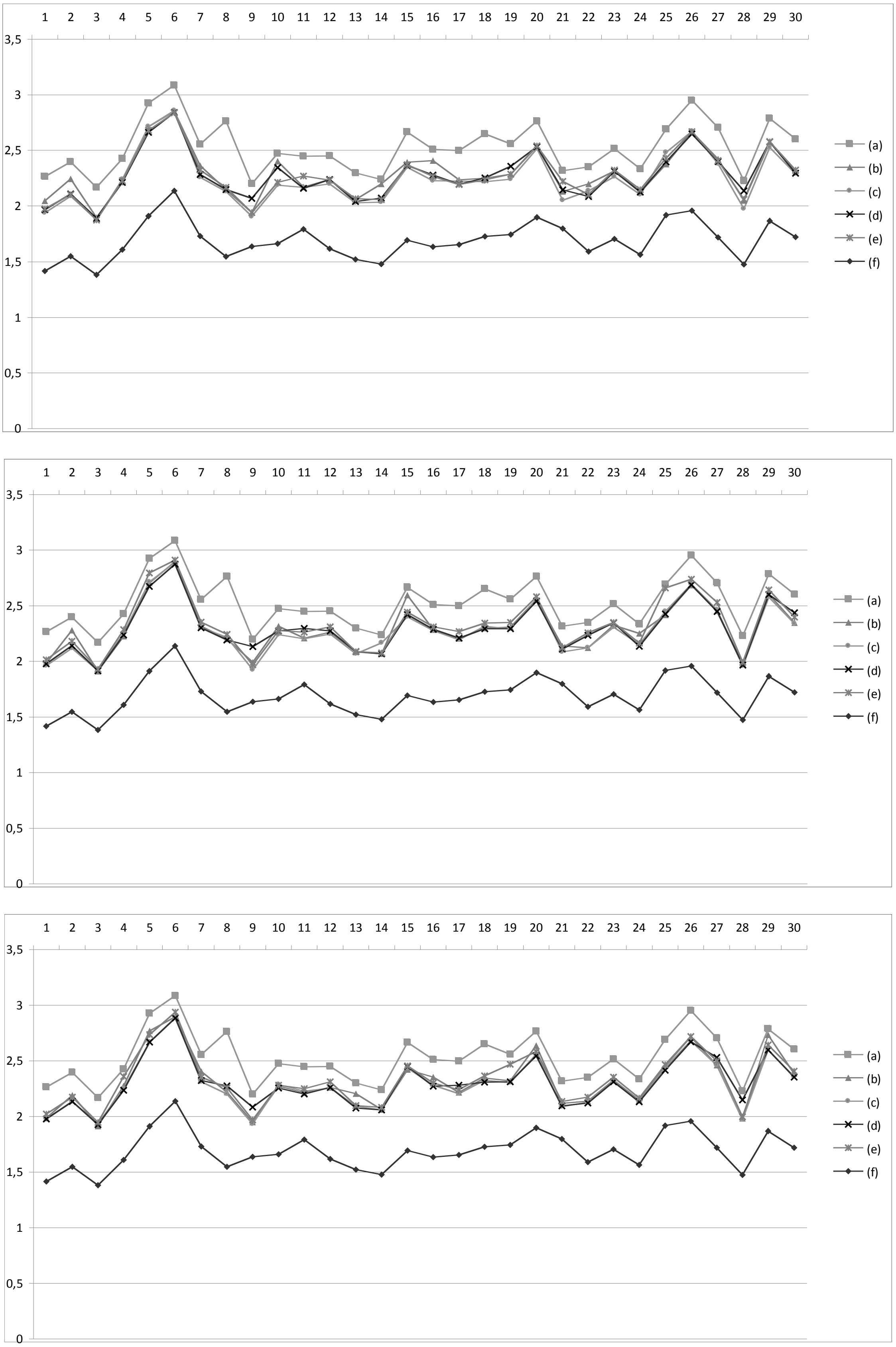
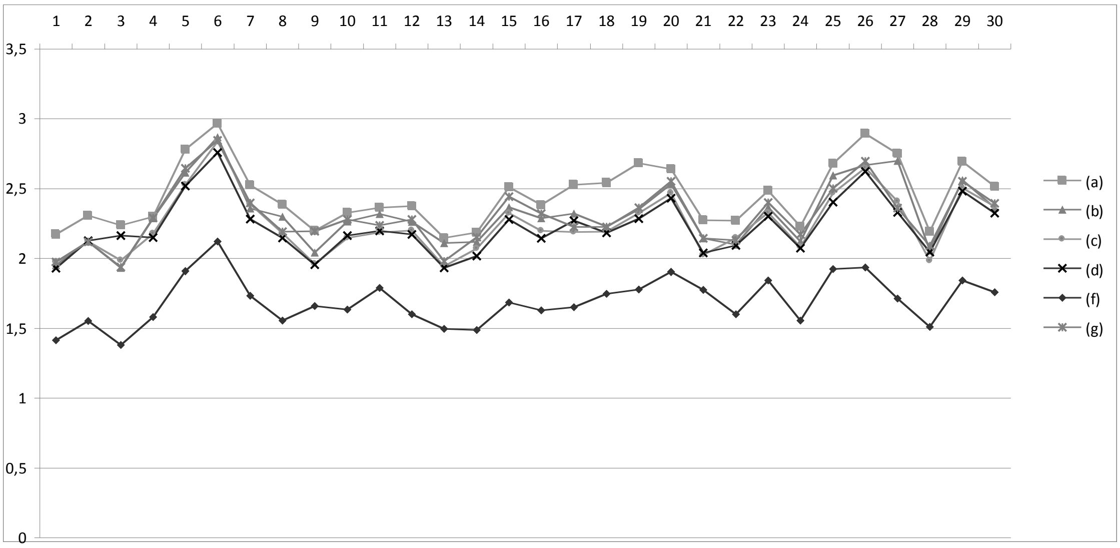
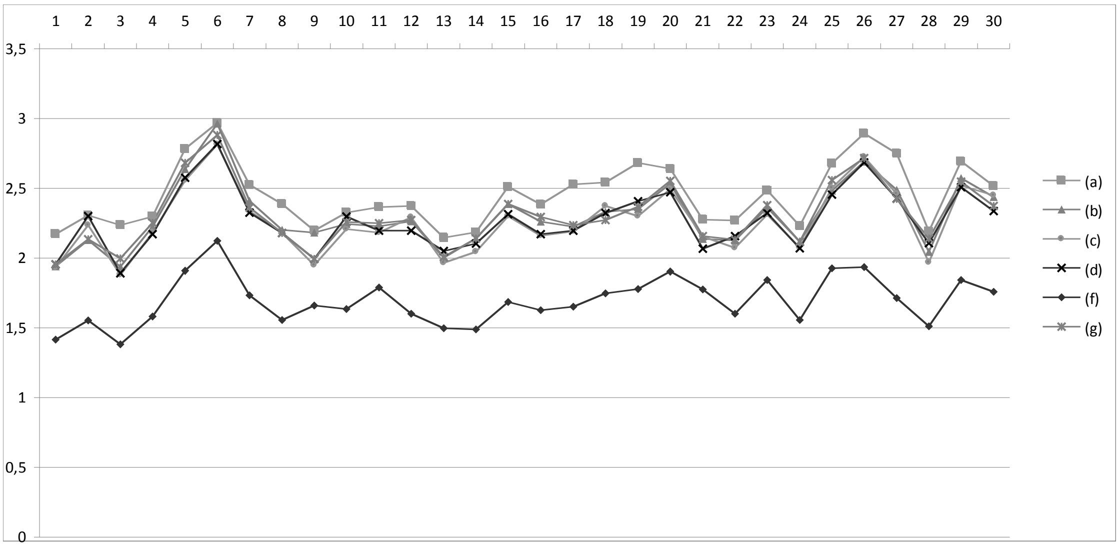


Figure 5 – Percentage Relative Root Mean Squared Error for all estimators and small areas: Study B – unscaled (top), scaling 1 (center) and scaling 2 (bottom) weighting method.





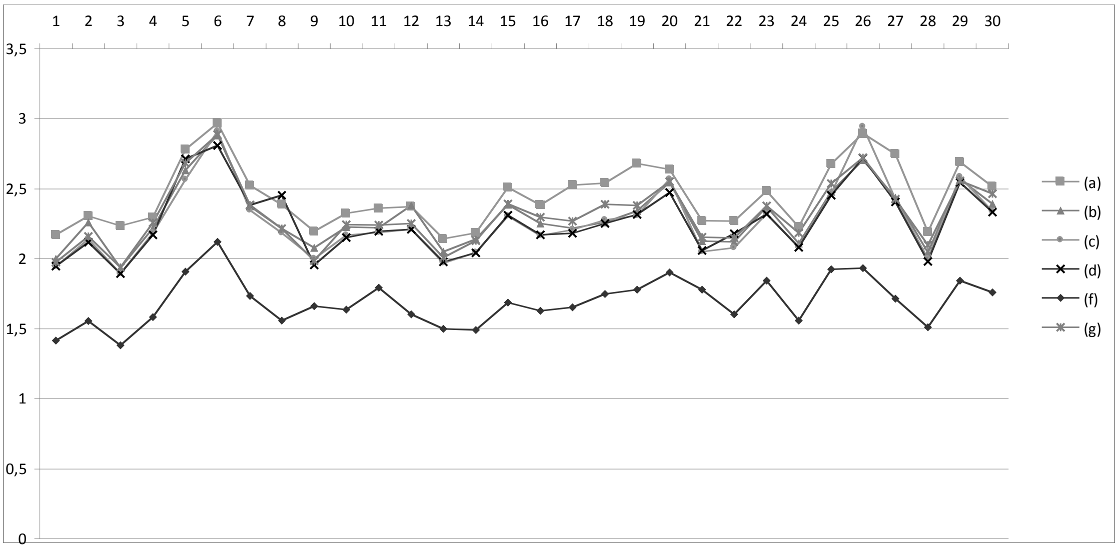
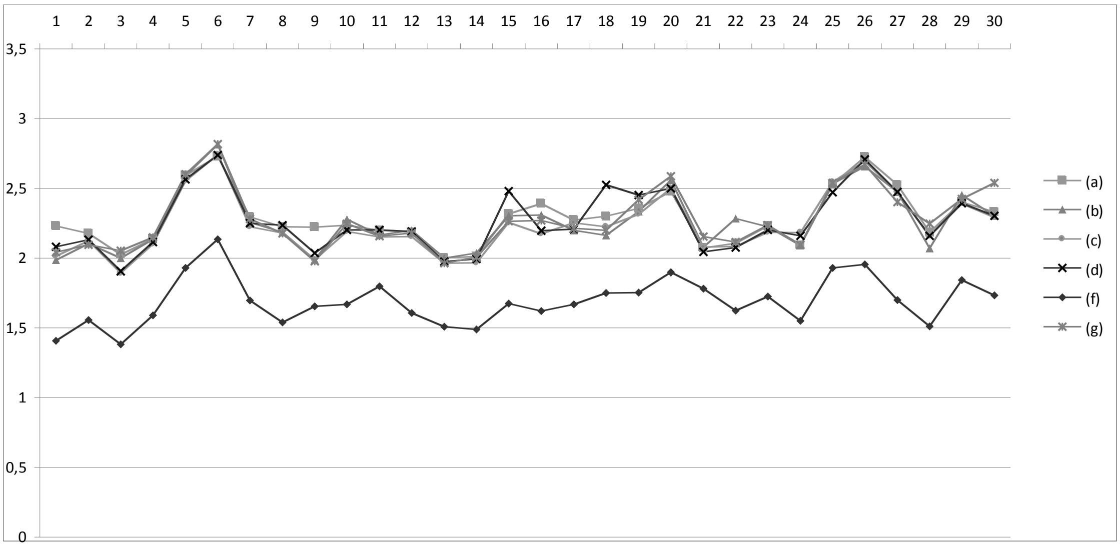
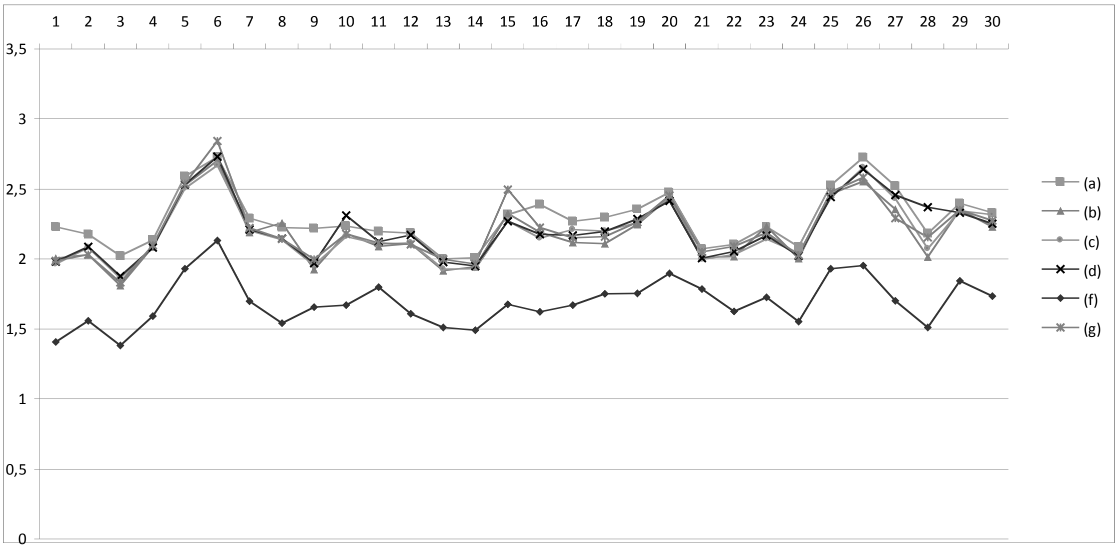
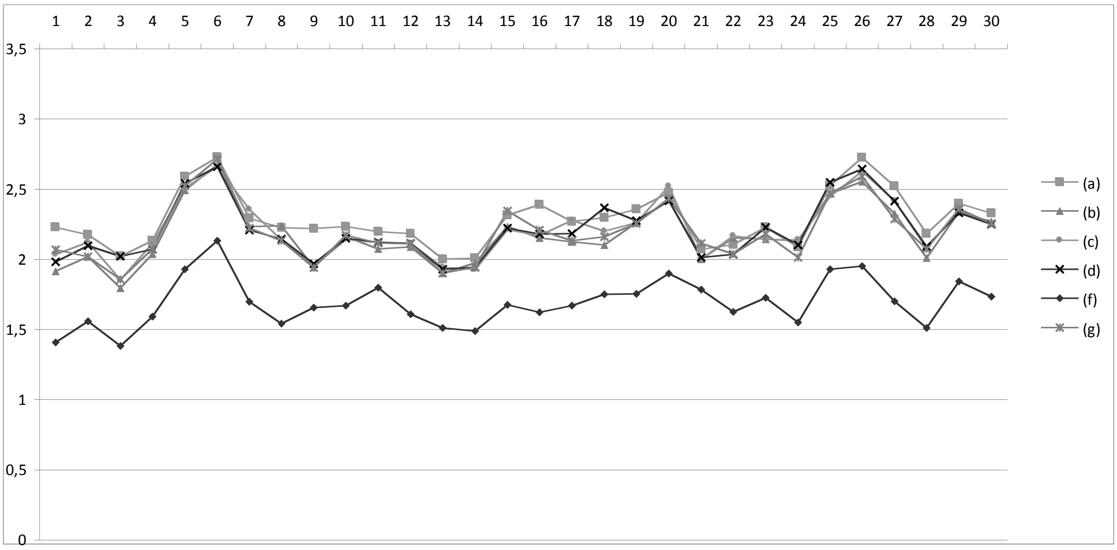


Figure 6 – Percentage Relative Root Mean Squared Error for all estimators and small areas: Study C – unscaled (top), scaling 1 (center) and scaling 2 (bottom) weighting method.







In all the three studies the RRMSE of the unweighted estimator is higher than that of all the other predictors even if this gap is less pronounced in study C. The differences among the weighted predictors are instead less evident: thus, the bias reduction of some of these estimators does not correspond to an equivalent reduction in variability. This result probably depends on the variability of the response probabilities (and consequently of the weights) among the areas. For example in study B and C the two worst estimators in term of bias are those not able to adequately incorporate in the estimated response probabilities the variability of the true probabilities among the areas. On the other hand, this situation corresponds to a lower variability respect to the other weighted estimators, but in this case the lower variability is not a good reason to prefer an estimator respect to another.

4. Final remarks

The growing demand for reliable small area estimates on one hand and the consciousness that nonresponse is a common problem in sample surveys on the other hand have motivated our study.

The analysis of the effect of nonresponse on small area estimates is a wide issue. This article presents a first attempt of analyzing this problem and suggesting possible solutions only for the case in which small area estimates are produced using a unit level random effects model and nonresponse concerns only the level-1 units.

The results of our simulations show that when informative nonresponse occurs it should not be ignored. The use of the MPML estimation procedure with weights function of response probabilities properly estimated may reduce the bias in a significant manner. Another important message standing from our simulations is the necessity to address together the nonresponse problem and the small area estimation problem, since a nonresponse adjustment method that effectively reduces the nonresponse bias of estimators referred to the whole population may not be able to also reduce the bias of small area estimators. This issue is relevant whenever some characteristics of the areas affect the response mechanism in a different way from an area to another, or from a group of areas to another.

We think that the necessity to address together the two problems is important in many real surveys. Often the producer of data is interested in producing estimates for the whole population or at most for some large subpopulations; therefore, he computes the survey weights adopting only these goals. Secondary users of data instead may be interested in producing small area estimates. However, these subjects often do not have access to all the original data and consequently it is impossible for them to calculate weights different from those provided by the producer of data.

The nature of the available information induces to another consideration. In theory the effect of nonresponse on unit level random effects models can be controlled including the variables that affect the response process among the model covariates. However, this is less practical than using these variables to estimate response probabilities because in this second case they only need to be known for the sampled units and not for all population units. Another reason is that these variables may not be available at the inference stage (e.g. sensible data) but may be used by the producer of data to calculate the survey weights.

The analyses performed in this work are based on the following assumptions: available auxiliary information allows fitting a small area model at unit level; all areas are represented in the set of level-1 responding units. These assumptions may appear as a limitation. We think that in many real situations they may be valid; nevertheless, in next developments of our work we plan to extend the theory presented here to cases where in some areas all the units do not respond to the survey, further evaluating also the role of the different scaling methods of the weights.

Another direction in which we plan to extend our work is the analysis of the effect of nonresponse and of weighting adjustment methods to the case in which the small area mean estimates are produced fitting a small area model at the area level. A useful starting point in this direction may be the extension to the nonresponse context of the procedure suggested by Pfefferman and Sverchkov (2007) in order to predict small area means under informative sampling.

In all these frameworks we also plan to propose an estimator of the MSE of the small area mean predictors proposed in this work, also considering, if necessary, bootstrap specifications.

References

Asparouhov, T. (2006) General Multi-level modeling with sampling weights, *Communications in Statistics – Theory and Methods*, 35, 439-460.

Carle, A. C. (2009) Fitting multilevel models in complex survey data with design weights: recommendations, *BMC Medical Research Methodology*, 9,1-13.

Chambers, R., and Dorfman, A.H. (2003). *Transformed Variables in Survey Sampling*. S3RI Methodology Working Papers, M03/21, Southampton Statistical Sciences Research Institute, University of Southampton, UK.

Chambers, R.L. and Dunstan, R. (1986). Estimating distribution functions from survey data. *Biometrika*, 73, 597-604.

European Commission (2006), "EU-SILC user database description, Version 2004-1 from 24-05-06", EU-SILC/BB D(2005), European Commission - Eurostat.

Giusti, C., Pratesi, M., and Salvati, S. (2009). Small area methods in the estimation of poverty indicators: the case of Tuscany. [Paper submitted for publication. A copy is available upon request].

Grilli, L. and Pratesi, M. (2004) Weighted estimation in multi-level ordinary and binary models in the presence of informative sampling designs, *Survey Methodology*, 30, 93-1003.

Hall, P. and Maiti, T. (2006) On parametric bootstrap methods for small area prediction, *Journal of the Royal Statistical Society B*, 68, 221-238.

Korn, E. L. and Graubard, B. I.(2003) Estimating variance components by using survey data, *Journal of the Royal Statistical Society B*, 65,175-190.

Lessler, J. T. and Kalsbeek, W. D. (1992) *Nonsampling Error in Surveys*, John Wiley & Sons.

Lundström S. and Särndal C. E. Särndal (2005) *Estimation in Surveys with Nonresponse*, John Wiley & Sons.

Pfefferman, D., Skinner, C.J., Holmes, D.J., Goldstein, H., Rasbash, J. (1998) Weighting for unequal selection probabilities in multilevel models, *Journal of the Royal Statistical Society B*, 60, 23-40.

Pfefferman, D. and Sverchkov, M. (2007) Small-area estimation under informative probability sampling of areas and within the selected areas, *Journal of the American Statistical Association*, 102, 1427-1439.

Pothoff, R.F, Woodbury, M.A, and Manton, K.G. (1992) Equivalent sample size and equivalent degrees of freedom refinements for inference using survey weights under superpopulation models, *Journal of the American Statistical Association*, 87, 383-396.

Rao, J. N. K. (2003) *Small area estimation*, Wiley, New York.

Rabe-Hesketh, S. and Skrondal, A. (2006) Multilevel modelling of complex survey data, *Journal of the Royal Statistical Society A*, 169, 805-827.

Saei, A. and Chambers, R. (2003) *Small Area Estimation Under Linear and Generalized Linear Mixed Models With Time and Area Effects.* Southampton, UK, Southampton Statistical Sciences Research Institute, 31pp. (S3RI Methodology Working Papers, M03/15).

Skinner, C. J. (1989) Domain means, regression and multivariate analysis, In *analysis of complex surveys*, (eds C. J. Skinner, D.Holt and T. M. F. Smith) Chichester: Wiley, 59-87.

Stampleton L. (2002) The incorporation of sample weights into multilevel structural equation models, *Structural Equation modelling*, 9, 475-502.

You, Y. and Rao, J. N. K. (2002) A pseudo-empirical best linear unbiased prediction approach to small area estimation using survey weights, *The Canadian Journal of Statistics*, 30, 431-439.