Computational Statistics and Data Analysis 184 (2023) 107742

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Estimating heterogeneous causal effects in observational studies using small area predictors

Setareh Ranjbar [a,*], Nicola Salvati [b], Barbara Pacini [c]

[a] Department of Psychiatry, University of Lausanne, Switzerland
[b] Department of Economics and Management, University of Pisa, Italy
[c] Department of Political Science, University of Pisa, Italy

ARTICLE INFO

Article history:
Received 10 August 2021
Received in revised form 22 December 2022
Accepted 12 March 2023
Available online 17 March 2023

Keywords:
M-quantile regression
Linear mixed models
Potential outcome
Inverse propensity scores
Heterogeneity of effects

ABSTRACT

The official statistics produced by National Statistical Institutes are mainly used by policy makers to take decisions. In particular, when policy makers and decision takers would like to know the impact of a given policy, it is important to acknowledge the heterogeneity of the treatment effects for different domains. If the domain of interest is small with regard to its sample size, then the evaluator has entered the small area estimation (SAE) dilemma. Based on the modification of the Inverse Propensity Weighting estimator and the traditional small area predictors, new estimators of area specific average treatment effects are proposed for unplanned domains. A robustified version of the predictor against presence of the outliers is also developed. Analytical Mean Squared Error (MSE) estimators of the proposed predictors are derived. These methods provide a tool to map the policy impacts that can help to better target the treatment group(s). The properties of these small area estimators are illustrated by means of a design-based simulation using a real data set where the aim is to study the effects of permanent versus temporary contracts on the economic insecurity of households in different regions of Italy.

1. Introduction

In recent years, the thrust of planning process has shifted from the macro to the micro level. There is a demand from administrators and policy planners for reliable estimates of various parameters at the micro level (Chandra et al., 2011). In particular, policy makers and decision takers would like to know the impact of a given policy in certain unplanned geographic, socio-demographic, or socio-economic domains. Thus, they are faced with the problem of estimating heterogeneous causal effects. Unfortunately, very often it is not possible to design a randomized experiment, and observational data (from censuses, administrative archives and surveys that are not designed for the purpose) are used to evaluate the effects of the intervention. In some cases, large databases with baseline covariates are available and the assignment to treatment (the benefit received) is known, but there is not enough information on the outcome variables to be representative of the unplanned domains. In particular, direct estimates are not accurate because sample surveys are usually designed so that direct estimators for larger domains (states, regions - macro level) lead to reliable estimates. If the domain of interest for impact evaluation is small with regard to its sample size (or even zero in some domains), then the evaluator has entered the small

* Corresponding author at: CEPP, Rte de Cery 25, 1008 Prilly, Switzerland.
E-mail address: setareh.ranjbar@chuv.ch (S. Ranjbar).

https://doi.org/10.1016/j.csda.2023.107742

area estimation (SAE) dilemma. Small area techniques provide official statistics using the survey samples and other sources of available information from which the estimators can borrow strength.

It is still surprising that no link has been established between the SAE literature and causal analysis that would allow for evaluating the impact of such a policy or decisions at a finer population level. There are exceptions but with different intentions. Chan (2018) attempts to combine the strength of the two fields, causal inference and small area estimation, to provide more precise generalization of the randomized trials to the entire population. The paper uses model-based techniques borrowed from the SAE literature to get a better estimate of the average treatment effect in the sub-classification strata, which are defined by the propensity scores, that have a sparse sample from the randomized experiment. There has been some statistical research on how to assess the generalizability of randomized trials to the target population in which it may be implemented (external validity). Stuart et al. (2011) propose the use of propensity-score-based metrics to quantify the similarity of the participants in a randomized trial and a target population. Stuart et al. (2015) provide a case study using one particular method, which weights the subjects in a randomized trial to match the population on a set of observed characteristics. Methods for assessing and enhancing external validity are just beginning to be developed. These studies and SAE methods share the aim to generalize the sample treatment effect to the population. However, the heterogeneity of the effects in different sub-populations is usually out of the scope of external validity analysis.

In this paper, we propose new methods to estimate the area specific average treatment effects for small areas in observational studies. The main motivation behind this is that such methods allow for local rather than universal policy advice. Another advantage of our proposed method over existing ones is that the small area techniques can be used to predict the effects even if the sample size of the treated or control group is zero in the area of interest. In particular, if the size of the control group is zero in the small area of interest the proposed method still allows to predict the area-specific average treatment effect, whereas if the size of the treated group in the area of interest is zero the new approach is not able to estimate the area-specific average treatment effect, but it allows to predict an average treatment effect, that is not area-specific, by a synthetic estimation. We adopt the nested error unit-level regression models (Battese et al., 1988) and the M-quantile models (Chambers and Tzavidis, 2006) to estimate propensity scores and the unobserved outcomes for the population because they are methods vastly used in SAE; however in the general format of our proposal other prediction methods that are justified by the practitioner can be adopted. Then to estimate the area specific average treatment effects for unplanned domains we propose a modification of the Inverse Probability Weighting estimator based on the estimated propensity scores and predicted outcomes (Rosenbaum and Rubin, 1983; Hahn, 1998) and we prove that it is double robust.

Borrowing from some recent papers we report two examples of impact assessment based on observational data in which our methodological proposal could improve the accuracy of the results at a finer level (territorial classification or population subgroups).

Bachtrögler et al. (2020) analyze the impact of the European Union's Cohesion Policy (CP) on manufacturing firm growth. They aim to assess whether and to which extent the effects of the regional CP investments on supported manufacturing firms' performance vary across different territorial settings (European countries and NUTS-2 regions). The paper combines firm-level data with a set of territorial characteristics of NUTS-2 regions (data assembled from three databases). Firms for which no NUTS-2 information is available were dropped out from the study and the sample was further reduced due to poor availability of outcome variables (change in value added, employment growth, and growth in productivity) for some firms.

Starting from growing interest in studying the effectiveness of therapies in real-world conditions, Wendling et al. (2018) compare methods based on observational data from health care databases (e.g., commercial claims data, electronic health records, and national registries) to estimate treatment effects that are supposed to be heterogeneous, e.g. different in sub-populations which are excluded or underrepresented in Random Control Trials (RCTs).

Depending on the outcome variables of interest (which could also be rare), it may be difficult to have follow-up data for a representative sample of the subgroups of interest. In such a case, small area estimation techniques can help reconstruct the outcome variable where it is missing, using covariates available in the health care databases.

To show the potential of our proposal, in this paper we consider a design-based simulation experiment based on real data in an observational setting. Our experiment aims to approximate a real application in economic policy evaluation as closely as possible.

The paper is organized as follows. Section 2 is devoted to set out the theoretical background and the assumptions of the causal inference which are then used to extend the small area predictors. We propose extensions to the Empirical Best Linear Unbiased Predictor (EBLUP) and M-quantile-based predictors for causal inference in Section 3. Their corresponding MSE estimators are presented in Section 4. The performances of these newly proposed predictors are empirically assessed in Section 5 by a design-based simulation using EU-SILC data. Finally, in Section 6 we summarize our main findings, and provide directions for future research.

2. Notation and assumptions

To explain the methodology developed in this paper we need to link the notations and the terminologies conventionally used both in small area estimation and in causal inference. Specifically, in SAE, we usually use small letters to indicate the outcome variable because we analyze a finite population, while in the new framework we switch to capital letters to take into account the probabilistic assignment mechanism of treatment. In what follows we use the bold cases to indicate vectors

and matrices. The parameters of interest are shown using Greek letters, for example α, and their estimates are defined by adding a 'hat', for example $\hat{\alpha}$.

Consider a population $\mathcal{U}$ of size N that is partitioned into m mutually disjoint sub-populations/domains $\mathcal{U}_j$ of size N_j, $j = 1, \cdots, m$. We assume that values of a (continuous) outcome variable Y_{ij} are available from a random sample s, which includes units from all target domains. We assume that a set of auxiliary information, denoted as a vector of covariates and treatment status, $(\mathbf{x}_{ij}, w_{ij})$, is available for all the units in the population (from census or administrative data) and that provides predictive power for the unobserved part of the population. It is also assumed that the vector $\mathbf{x}_{ij}$ of dimension $p \times 1$ contains the set of all confounders and some additional covariates that are useful in predicting the outcome. More generally, the vector of covariates may include both individual and area-level covariates. It is a common practice for model-based techniques in SAE to assume a non-informative sampling scheme to allow valid inferences of non-sampled units based on models for sampled values. In what follows we implicitly assume a *Simple Random Sampling Without Replacement (SRSWOR)*.

We are interested in studying the impact of a binary treatment, W_{ij}, that takes the value 1 for treated and 0 for non-treated (control) units on the outcome Y_{ij} in the population. We focus on treatment assigned at the individual level and assume that the information on treatment status exists for all population units, for example from administrative sources. This is a plausible assumption in many applications, such as unemployment benefits, government subsidies and pensions.

We denote the sample size, the sampled part of the population and the non-sampled part of the population in each small area j by n_j, s_j and r_j respectively, with $\mathcal{U}_j = s_j \bigcup r_j$. The total sample size is given by $n = \sum_{j=1}^{m} n_j$.

To link the two methodologies on small area estimation and causal inference, we adopt the framework of Rubin Causal Model (RCM) (Rubin, 1974), and use the approach of potential outcomes to properly define the causal estimands of interest. In small area estimation setting the aim is to provide estimates of the average effects for each small sub-population or domain (i.e., these are the unplanned domains in the survey) rather than for the entire population. This is particularly relevant when heterogeneous effects are expected among different domains. In these cases our proposal can provide a map of policy impacts at a small area level, helping to better understand the impact of an intervention and to better target the treatment group(s).

The potential outcome approach is firstly developed under SUTVA (Stable Unit Value Assumption; Rubin, 1980), stating that the outcome of each unit is unaffected by the treatment assignment of any other unit and also that there are no different versions of each treatment level, which may lead to different potential outcomes. Within the simplest framework, each unit has only two potential outcomes, defined as Y_{ij}^0 and Y_{ij}^1 under control and under treatment, respectively. The former, Y_{ij}^0, denotes the outcome that would be realized by the unit if not treated and the latter, Y_{ij}^1, indicates the outcome that would be realized by the same unit if treated.

For the sampled units of area j (the set s_j) only one of the potential outcomes is observed for each individual; the other is necessarily missing and needs to be predicted, entering the so called fundamental problem of causal inference. We then observe the outcome variable Y_{ij} where $Y_{ij} = W_{ij}Y_{ij}^1 + (1 - W_{ij})Y_{ij}^0$, in this set. For the non-sampled units of area j (the set r_j), however, neither of the potential outcomes are available and both need to be predicted, implying that for the out of sample units Y_{ij}s are never observed. In this respect, our problem resembles that studied widely in the literature of imputation for missing data in the context of small area estimation. See Haziza and Rao (2010), Cantoni and de Luna (2018) and Chen and Haziza (2019) for a comprehensive review of this topic. The main difference of this line of literature with our work is twofold: (i) causal inference require additional assumptions and (ii) the percentage of missing values for which we need to predict the value is not negligible.

The individual treatment effect for the unit i in area j can be defined as a comparison of potential outcomes, such as the difference, conditioning on area j, and can be denoted as:

$$\tau_{ij} = Y_{ij}^1 - Y_{ij}^0 = \left(Y_i^1 - Y_i^0 | j\right).$$

For ease of notation, from now on we simply use the subscript j to indicate conditional operators or conditional expectations given j, avoiding the explicit conditioning in formulas.

This parameter τ_{ij} is not identifiable due to a lack of information for each unit, but several causal estimands can be defined as summaries of individual effects, which are identifiable and can be estimated out of the data under some additional assumptions. Here we distinguish between two sets of estimands that are essential for our analysis. The first includes the conditional average treatment effect ($SATE_j$) for the sample units. The second set of estimands includes the conditional ATE for the population, named $PATE_j$. Each of these estimands can be defined at the area (domain) level (i.e., conditioning on the specific area-membership, which acts as a special covariate) as follows:

$$\tau_{SATE_j} = \frac{1}{n_j} \sum_{i \in s_j} \left(Y_{ij}^1 - Y_{ij}^0\right), \tag{1}$$

$$\tau_{PATE_j} = \frac{1}{N_j} \sum_{i \in \mathcal{U}_j} \left(Y_{ij}^1 - Y_{ij}^0\right). \tag{2}$$

The aim of our proposal is to provide reliable estimates of τ_{PATE_j} for different areas/domains, borrowing strength from small area estimation techniques.

Causal effects from observational data can be identified under a set of assumptions, guaranteeing that the treatment is effectively randomized within cells defined by the values of a set of observed covariates. Slight modifications are needed in some cases for the identification of heterogeneous effects among different domains.

Here, we assume SUTVA, which is implied in the notation above, together with strong ignorability assumptions:

Assumption 1. Stable Unit Treatment Value

SUTVA states that there are no different forms or versions of each treatment level and the potential outcome for any unit does not vary with the treatments assigned to other units.

This assumption may become questionable in multilevel designs, where individual causal effects can vary depending on which cluster a unit is assigned to. In this case, interactions between units within a cluster are likely, mainly when treatment is administered at cluster level. The implications of a cluster structure, which may affect both the assignment to treatment and the potential outcomes, have not been intensively studied, with a few exceptions (Arpino and Mealli, 2011; Li et al., 2013; Kim et al., 2017; Cafri et al., 2019).

We explicitly included the cluster variable, which is here considered as a special confounder to condition on. In the definition of cluster-specific potential outcomes and conditional causal effects. We also assume that the treatment administered at the unit level will not affect other units within the same area and that there are no expected movements and interference across domains. Therefore, *a multilevel SUTVA* will be maintained throughout the paper.

Assumption 2. Unconfoundedness based on propensity scores

The assignment mechanism is unconfounded (with the potential outcomes, Rosenbaum and Rubin, 1983) if:

$$W_{ij} \perp (Y_{ij}^1, Y_{ij}^0) \mid \mathbf{x}_{ij}, \qquad \forall i \in \mathcal{U}_j,$$

or

$$W_{ij} \perp (Y_{ij}^1, Y_{ij}^0) \mid e(\mathbf{x}_{ij}), \qquad \forall i \in \mathcal{U}_j,$$

where $e(\mathbf{x}_{ij}) = Pr(W_{ij} = 1 | \mathbf{X}_{ij} = \mathbf{x}_{ij})$ to be decided is known as a propensity score.

We assume that, conditional on a set of pre-treatment covariates or conditional solely on the propensity scores, the assignment mechanism is independent from the cluster-specific potential outcomes.

Assumption 3. Common support (overlap)

We assume that the unconfounded assignment mechanism is probabilistic, that is all the unit-level probabilities for receiving treatment are strictly between zero and one:

$$0 < e(\mathbf{x}_{ij}) = Pr(W_{ij} = 1 | \mathbf{x}_{ij}) < 1 \qquad \forall i \in \mathcal{U}_j.$$

In other words, each unit in the defined population has a chance of being treated and a chance of not being treated (Rosenbaum and Rubin, 1983). We assume common support within area, based on the whole set of population auxiliary variables, using the propensity scores within each area at population level. Treatment assignment mechanisms satisfying both overlap and unconfoundedness are called strongly ignorable, so that we assume strong ignorability within each area/domain. Rubin et al. (2004) discuss the importance of using propensity scores to match the treatment and control units while using regression models in the complex survey settings. This approach can be also considered as a diagnostic tool to test the Assumption 3.

3. Small area estimators for causal inference

We propose a modification of the Augmented Inverse Probability Weighting estimator (Rosenbaum and Rubin, 1983), based on the estimated propensity scores and predicted outcomes.

We start from defining the Conditional Average Treatment Effect in area j, using the notation introduced in Section 2 where the subscript j denotes the conditional expectation given the area-membership j. Under unconfoundedness (Assumption 2), Imbens and Wooldridge (2009) show that:

$$E_j[Y_{ij}^1] = E_j\left[\frac{W_{ij}Y_{ij}}{e(\mathbf{x}_{ij})}\right],$$

and

$$E_j[Y_{ij}^0] = E_j\left[\frac{(1-W_{ij})Y_{ij}}{1-e(\mathbf{x}_{ij})}\right],$$

where $e(\cdot)$ is the propensity score (i.e., the probability for each unit to be treated) which is a function of confounding covariates, $\mathbf{x}_{ij}$. The natural sample estimator for this parameter is:

$$\tilde{\tau}_{SATE_j} = \frac{1}{n_j}\sum_{i\in s_j}\left[\frac{w_{ij}y_{ij}}{e(\mathbf{x}_{ij})} - \frac{(1-w_{ij})y_{ij}}{1-e(\mathbf{x}_{ij})}\right].$$

If the propensity scores are unknown, the $e(\mathbf{x}_{ij})$ values need to be replaced by their estimates. Using an estimated propensity score even when the scores are known is recommended for greater efficiency gain in estimating the average treatment effect (Hirano et al., 2003; Robins et al., 1995; Rosenbaum and Rubin, 1983). Therefore later on in our setting, even though the assignments to treatment are known for all units in the population, we use the estimated propensity scores:

$$\tau^{\star}_{SATE_j} = \frac{1}{n_j}\sum_{i\in s_j}\left[\frac{w_{ij}y_{ij}}{\hat{e}(\mathbf{x}_{ij})} - \frac{(1-w_{ij})y_{ij}}{1-\hat{e}(\mathbf{x}_{ij})}\right]. \tag{3}$$

Lunceford and Davidian (2004) and Imbens (2004) propose to improve the performance of estimator (3) by re-normalizing the weights so that they some up to one:

$$\hat{\tau}_{SATE_j} = \left(\sum_{i\in s_j}\left[\frac{w_{ij}y_{ij}}{\hat{e}(\mathbf{x}_{ij})}\right]\right)\left(\sum_{i=1}^{n_j}\frac{w_{ij}}{\hat{e}(\mathbf{x}_{ij})}\right)^{-1} - \left(\sum_{i\in s_j}\left[\frac{(1-w_{ij})y_{ij}}{1-\hat{e}(\mathbf{x}_{ij})}\right]\right)\left(\sum_{i=1}^{n_j}\frac{1-w_{ij}}{1-\hat{e}(\mathbf{x}_{ij})}\right)^{-1}. \tag{4}$$

In what follows we refer to (4) as the IPW-Direct estimator, which is the classical Inverse Propensity Weighting estimator proposed by Rosenbaum and Rubin (1983). Alternative direct estimators, that use the survey weights, have been proposed by Zanutto (2006) and by Miratrix et al. (2018).

In the context of small area estimation we assume the unit level auxiliary information and the treatment status are available for all units in the population. Under unconfoundedness and following Imbens and Wooldridge (2009) τ_{PATE_j} (2) can be written as:

$$\tau_{PATE_j} = E_j[Y_{ij}^1] - E_j[Y_{ij}^0] = E_j\left[\frac{W_{ij}Y_{ij}}{e(\mathbf{x}_{ij})}\right] - E_j\left[\frac{(1-W_{ij})Y_{ij}}{1-e(\mathbf{x}_{ij})}\right], \tag{5}$$

where the expectations are taken over all units in the population of area j. The population counterpart of these expectations can be re-expressed as:

$$\begin{aligned}\tau_{PATE_j} = &\left(\sum_{i\in s_j}\left[\frac{w_{ij}y_{ij}}{e(\mathbf{x}_{ij})}\right] + \sum_{i\in r_j}\left[\frac{w_{ij}y_{ij}}{e(\mathbf{x}_{ij})}\right]\right)\left(\sum_{i=1}^{N_j}\frac{w_{ij}}{e(\mathbf{x}_{ij})}\right)^{-1} - \\ &\left(\sum_{i\in s_j}\left[\frac{(1-w_{ij})y_{ij}}{1-e(\mathbf{x}_{ij})}\right] + \sum_{i\in r_j}\left[\frac{(1-w_{ij})y_{ij}}{1-e(\mathbf{x}_{ij})}\right]\right)\left(\sum_{i=1}^{N_j}\frac{1-w_{ij}}{1-e(\mathbf{x}_{ij})}\right)^{-1},\end{aligned} \tag{6}$$

where $1/N_j$ is replaced with $\left(\sum_{i=1}^{N_j}\frac{w_{ij}}{\hat{e}(\mathbf{x}_{ij})}\right)^{-1}$ and $\left(\sum_{i=1}^{N_j}\frac{1-w_{ij}}{1-\hat{e}(\mathbf{x}_{ij})}\right)^{-1}$ to normalize the sum.

This allows us to predict the outcome for the non-sampled part of the population in each small area and to fit a model on the whole population to estimate the propensity scores quite accurately (Hirano et al., 2003; Robins et al., 1995; Rosenbaum and Rubin, 1983). Then, using the predicted outcome $\hat{y}_{ij}$ for out of sample units and the estimated propensity scores $\hat{e}(\mathbf{x}_{ij})$, an estimator of (6) is:

$$\begin{aligned}\hat{\tau}_{PATE_j} = &\left(\sum_{i\in s_j}\left[\frac{w_{ij}y_{ij}}{\hat{e}(\mathbf{x}_{ij})}\right] + \sum_{i\in r_j}\left[\frac{w_{ij}\hat{y}_{ij}}{\hat{e}(\mathbf{x}_{ij})}\right]\right)\left(\sum_{i=1}^{N_j}\frac{w_{ij}}{\hat{e}(\mathbf{x}_{ij})}\right)^{-1} - \\ &\left(\sum_{i\in s_j}\left[\frac{(1-w_{ij})y_{ij}}{1-\hat{e}(\mathbf{x}_{ij})}\right] + \sum_{i\in r_j}\left[\frac{(1-w_{ij})\hat{y}_{ij}}{1-\hat{e}(\mathbf{x}_{ij})}\right]\right)\left(\sum_{i=1}^{N_j}\frac{1-w_{ij}}{1-\hat{e}(\mathbf{x}_{ij})}\right)^{-1}.\end{aligned} \tag{7}$$

This proposal is quite general in the sense that no restriction is imposed on the models used to predict the unobserved outcomes or to estimate the propensity scores for population units.

In Theorem 1 we show that the estimator (7) has consistent and double robust properties without any extra adjustment. On the contrary, if in equation (7) the weights are not re-normalized the estimator is no more double robust and an adjustment term is needed to obtain this desirable property, see the Section S.1 in Supplementary Material.

We can show the consistency and double robust properties of the proposed estimator in equation (7) by expressing it as a weighted average of the outcomes y_{ij}s and their estimates $\hat{y}_{ij}$s in each small area:

$$\hat{\tau}_{PATE_j} = \left(\sum_{i\in s_j} a_{ij}y_{ij} + \sum_{i\in r_j} a_{ij}\hat{y}_{ij}\right)\left(\sum_{i=1}^{N_j} a_{ij}\right)^{-1} - \left(\sum_{i\in s_j} b_{ij}y_{ij} + \sum_{i\in r_j} b_{ij}\hat{y}_{ij}\right)\left(\sum_{i=1}^{N_j} b_{ij}\right)^{-1},$$

where $\{a_{ij} = \frac{w_{ij}}{\hat{e}(x_{ij})}\}$ and $\{b_{ij} = \frac{1-w_{ij}}{1-\hat{e}(x_{ij})}\}$ are the sequences of weights in area j. The $\hat{\tau}_{PATE_j}$ can be now expressed as a weighted average of i.i.d. random variables (rvs) conditioned on the small area j after few steps of mathematical developments, which are reported in Supplementary Material S.2, and we can show that the estimator is double robust and consistent as $N_j \to \infty$. Let A_{N_j} and B_{N_j} be $\sum_{i\in U_j} a_{ij}$ and $\sum_{i\in U_j} b_{ij}$, respectively. The theory will be developed under Assumptions 1, 2, 3. Further the following conditions have to be satisfied for the convergence of $\hat{\tau}_{PATE_j}$ to its true value:

(a) $A_{N_j} \to \infty$ and $a_{ij}/A_{N_j} \to 0$;
(b) $B_{N_j} \to \infty$ and $b_{ij}/B_{N_j} \to 0$;
(c) a_{ij}s and b_{ij}s are bounded;
(d) the $\text{var}(Y_{ij}) < \infty$ and $\text{var}(Y_{ij} - \hat{Y}_{ij}) < \infty$.

Assumption 2 and the fact that the propensity scores are estimated using the information on the whole population guarantee that weights are a deterministic sequence of values given the area population of size N_j. The conditions that $A_{N_j} \to \infty$ and $B_{N_j} \to \infty$ are linked with the assumptions that in each area there must be treated and non-treated units in the population. While it is possible to provide the estimates even if the entire sample units in some areas belong only to treated or control group, as pointed out in the Section 1, at the population level the presence of both groups is essential to provide area level estimates of the treatment. Further, according to Assumption 3 the propensity scores take values between 0 and 1 away from the boundaries. When $N_j \to \infty$, where the propensity scores are bounded away from 0, a_{ij}s are bounded and $a_{ij}/A_{N_j} \to 0$. Likewise, when the scores are bounded away from 1, b_{ij}s are bounded and $b_{ij}/B_{N_j} \to 0$.

Theorem 1. *Under Assumptions 1-3 and the conditions (a), (b), (c) and (d) the estimator (7) is double robust and consistent. That is:*

$$Pr\left(\lim_{N_j\to\infty} \hat{\tau}_{PATE_j} = E_j\left[Y_{ij}^1\right] - E_j\left[Y_{ij}^0\right]\right) = 1,$$

(i) *as long as the propensity score model is correct, even if the postulated prediction model is incorrect;*
(ii) *as long as the prediction model is correct, even if the postulated propensity model is incorrect.*

The proof of Theorem 1 is in Section S.2, Supplementary Material.

In estimator (7) different methods can be adopted to predict the unobserved y_{ij}s and to estimate the propensity scores. Here we propose two feasible strategies and discuss their implications on the estimation of (7). In the first proposal we predict the unobserved outcomes using EBLUP and a generalized linear mixed model to estimate the propensities. This estimator is referred to as IPW-EBLUP hereafter and can also be seen as a modification of the EBLUP estimator for the area level mean. In the second proposal we use a robust approach based on M-quantile models proposed by Chambers and Tzavidis (2006) for the continuous outcome and by Chambers et al. (2016) for the binary case to predict the unobserved outcomes and estimate the propensity scores. The resulting estimator is labelled IPW-MQ hereafter. We explain in more detail the models and the estimating strategies used for IPW-EBLUP and IPW-MQ in Section 3.1.

The properties of IPW-Direct estimators are widely studied in the literature; see for instance Hirano et al. (2003) and Wooldridge (2007) for more details. However, when the area/domain sample sizes are small these estimates are no longer reliable at this fine levels, that is, they could vary significantly. Our proposed estimators IPW-EBLUP and IPW-MQ overcome this problem by borrowing strength from additional sources of information rather than merely using the sample data. The second estimator can also deal with data that is contaminated by outlying values.

3.1. Data generating processes and estimation strategies

To explain the data generating process and justify our estimation strategies for predicting the unobserved population outcomes and estimating the population propensity scores once again we use the potential outcome framework. Consider the two potential outcomes for individual i in area j, Y_{ij}^1, and Y_{ij}^0, and τ_j be the area specific causal effect of a policy intervention. We are assuming a treatment effect specific to a subgroup of subjects, where the subgroup (that we call 'area') can be defined by subjects' attributes (e.g., gender), the context in which the intervention occurs (e.g., the subjects at a

specific site in a multi-site context), or a combination of both. We maintain a constant treatment effect assumption within subgroups ($\tau_{ij} = \tau_j$), bearing in mind that the subgroups can be very small in size.

To benefit from the hierarchical structure in the data, without loss of generality, we consider a nested error linear model (Battese et al., 1988) as the data generating process of the potential outcome in the absence of the treatment:

$$y_{ij}^0 = \mathbf{x}_{ij}^T \boldsymbol{\beta} + u_j + \epsilon_{ij},$$

where u_j is the area specific random effect and ϵ_{ij} is the individual error, the distributions of which are to be assumed (in general normal) if the model is fitted parametrically. This holds for the entire population as well as for the sample at hand in the absence of sample selection bias. Let w_{ij} be the individual treatment status, the outcome (observed in the sample and not observed for the population) is:

$$\begin{aligned} y_{ij} &= (w_{ij}) y_{ij}^1 + (1 - w_{ij}) y_{ij}^0 \\ &= \mathbf{x}_{ij}^T \boldsymbol{\beta} + w_{ij}\tau_j + u_j + \epsilon_{ij}. \end{aligned} \tag{8}$$

In the context of small area estimation we need to fit this model to the sample data and predict the outcome for the entire population by using the estimated parameters of the model and the auxiliary information that is available for the entire population. There are many different techniques that are developed in the SAE literature; two sets of parametric models are discussed in this paper, but, of course, others can also be adopted if appropriate. It is also worth noting that, if we have the area level variables in the model, then the interaction between these variables and the treatment variable must also be included in the random part of the model (Arpino and Mealli, 2011). It should be noted that since the proposed estimation methods in this section are parametric, they impose some restrictions on the conditional distribution of potential outcomes, and this directly translates to the shape of the treatment effect heterogeneity. In other words, we will explicitly assume that τ_js, are randomly distributed with $\tau_j \sim \mathcal{N}(\gamma_0, \sigma_\gamma^2)$. This assumption is relaxed when we introduce a robust estimator based on M-quantile models.

3.1.1. Out-of-sample estimation of outcome and propensities in the hierarchical structure

We start by assuming that the area specific causal effects, τ_js, are randomly distributed with $\tau_j \sim \mathcal{N}(\gamma_0, \sigma_\gamma^2)$. Then equation (8) can be rewritten as

$$y_{ij} = \tilde{\mathbf{x}}_{ij}^T \tilde{\boldsymbol{\beta}} + w_{ij}\gamma_j + u_j + \epsilon_{ij}, \tag{9}$$

where $\tilde{\mathbf{x}}_{ij} = (\mathbf{x}_{ij}^T, w_{ij})^T$ is of dimension $(p+1) \times 1$, $\tilde{\boldsymbol{\beta}} = (\boldsymbol{\beta}^T, \gamma_0)^T$ is the vector of fixed effects and we further assume that $u_j \sim \mathcal{N}(0, \sigma_u^2)$, and $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2)$. As a consequence of our assumption on the distribution of the area specific causal effects we have $\gamma_j \sim \mathcal{N}(0, \sigma_\gamma^2)$, that is the random slope associated with the treatment status. For obtaining the IPW-EBLUP, a mixed linear model (more specifically a random slope model) is fitted, using the maximum likelihood (ML) or restricted maximum likelihood (REML) method (McCulloch and Searle, 2001; Pinheiro and Bates, 2006). Then the estimated parameters are used to predict the outcome $\hat{y}_{ij}$ for $i \in r_j$ under model (9). The assumption of normality of the random components is mainly in place to specify the form of ML or REML used for estimating the unknown parameters of the model, including the unknown parameters of the variance-covariance matrix. However, this assumption can easily be relaxed using other existing methods for fitting random effect models, such as quasi-likelihood methods or Generalized Estimating Equation (Liang and Zeger, 1986) under some other mild conditions.

Proposition 1. *Under Assumption 2, unconfoundedness, the vector of random slopes* $\boldsymbol{\gamma}$ *and random intercepts* $\boldsymbol{u}$ *in equation* (9) *are independent, that is:*

$$\begin{bmatrix} \boldsymbol{\gamma} \\ \boldsymbol{u} \end{bmatrix} \overset{i.i.d.}{\sim} \left(\mathbf{0}, \quad \boldsymbol{\Sigma}_\omega \right),$$

where $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_m)^T$, $\boldsymbol{u} = (u_1, \cdots, u_m)$, *and* $\boldsymbol{\Sigma}_\omega = \begin{pmatrix} \boldsymbol{\Sigma}_\gamma & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_u \end{pmatrix}$.

Proof. Based on Assumption 2 the treatment assignment is independent from the potential outcomes conditional on the set of pre-treatment covariates (confounders). This assumption requires that conditional on observed covariates there are no unobserved factors that are associated both with the assignment mechanism and potential outcomes, that is, $E\left[\gamma_j \left(u_j + \epsilon_{ij}\right)\right] = 0$. Because $E\left[\gamma_j \epsilon_{ij}\right] = 0$ it goes that $E\left[\gamma_j u_j\right] = 0$. □

In equation (9) the average return to w_{ij} is captured by the fixed effect and the area specific heterogeneity of the return to w_{ij} is modelled through a random slope γ_j, that needs to be predicted (Li et al., 2013). However, our estimators of the total causal effect do not merely depend on the estimation/prediction of these two effects. In addition, we balance the characteristics of treated and control groups by weighting the outcomes based on the individual propensity scores.

Therefore, these estimators have doubly robust properties (Bang and Robins, 2005), that is, having misspecified only one of the models for the prediction of the outcomes or for the estimation of the propensity scores, we can still provide a consistent estimator for the causal effects of each area. Further, the hierarchical structure of the data as it is defined in equation (9) for the outcome model should also be considered in the estimation model of the propensity scores, see Arpino and Mealli (2011) and Arpino and Cannas (2016). Then, we consider the following model for the propensity scores:

$$\eta_{ij} = \Lambda(e(\mathbf{x}_{ij})) = \mathbf{x}_{ij}^T\boldsymbol{\alpha} + \nu_j, \tag{10}$$

where $\Lambda(.)$ is a logit link function. Substituting the estimated values $\hat{y}_{ij} = \tilde{\mathbf{x}}_{ij}^T\hat{\tilde{\boldsymbol{\beta}}} + w_{ij}\hat{\gamma}_j + \hat{u}_j$ and $\hat{e}(x_{ij}) = \Lambda^{-1}(\mathbf{x}_{ij}^T\hat{\boldsymbol{\alpha}} + \hat{\nu}_j)$ in equation (7) provides the estimates of IPW-EBLUP.

Remark 1. The $\hat{\tau}_{PATE_j}$ based on the IPW-EBLUP, obtained by fitting the random slope model to the sample data, is double robust and consistent estimator because this predictor satisfies the assumptions 1-3 and the conditions (a), (b), (c) and (d) of Theorem 1. Note that in this case the consistency of the EBLUP for the outcome variable is not needed, which can be obtained only if m and n_j tend to infinity as stated in Jiang (1999, 2010) and Lyu and Welsh (2021).

Remark 2. if the size of the control group is zero in the small area j the value of the response variable can be estimated by $\hat{y}_{ij} = \tilde{\mathbf{x}}_{ij}^T\hat{\tilde{\boldsymbol{\beta}}} + w_{ij}\hat{\gamma}_j + \hat{u}_j$, whereas if the size of the treated group in the area of interest is zero the response variable can be predicted by $\hat{y}_{ij} = \tilde{\mathbf{x}}_{ij}^T\hat{\tilde{\boldsymbol{\beta}}} + \hat{u}_j$.

3.1.2. Robust estimation for out-of-sample units

An alternative to mixed models and IPW-EBLUP is given by the M-quantile regression models for estimating the outcome variable and the propensity scores. If an outlying value can destabilize a population estimate based on a large survey sample, it can almost certainly destroy the validity of the corresponding direct estimate for the small area from which the outlier is sourced, since this estimate will be based on a much smaller sample size. This problem does not disappear when the small area estimator is a model based estimator such as EBLUP: large deviations from the expected response (outliers) are known to have a large influence on classical maximum likelihood inference based on generalized linear mixed models (GLMM). Chambers and Tzavidis (2006) and Sinha and Rao (2009) addressed the issue of outlier robustness in SAE proposing techniques that can be used to down-weight any outliers when fitting the underlying model. In particular, Chambers and Tzavidis (2006) proposed to apply the M-quantile regression models to SAE with the aim of obtaining reliable and outlier robust estimators without recourse to parametric assumptions for the residuals distribution using M-estimation theory. For details on M-quantile regression see Breckling and Chambers (1988).

When using the M-quantile method the unobserved outcomes are predicted as follows:

$$\hat{y}_{ij} = \mathbf{x}_{ij}^T\hat{\boldsymbol{\beta}}_{\hat{\bar{q}}_j} + w_{ij}\hat{\gamma}_{\hat{\bar{q}}_j}, \tag{11}$$

where $\hat{\boldsymbol{\beta}}_{\bar{q}_j}$ and $\hat{\gamma}_{\bar{q}_j}$ are the regression coefficients of the M-quantile model estimated at quantile $\hat{\bar{q}}_j$, that is, the average of the estimated quantiles for the sample units in area j. The Chambers and Tzavidis (2006) proposal is an alternative to the random effect models for characterizing the variability across the population not accounted for by the regressors based on the M-quantile coefficients of the population units. The authors observed that if a hierarchical structure does explain part of the variability in the population data, units within areas defined by this hierarchy are expected to have similar M-quantile coefficients. For details on the computation of M-quantile coefficients see Chambers and Tzavidis (2006).

For estimating the propensity scores the M-quantile for binary data proposed by Chambers et al. (2016) is adopted. Modelling the M-quantiles of a binary outcome presents more challenges than modelling the M-quantiles of a count outcome. A detailed account of these challenges is provided in Chambers et al. (2016). The authors proposed a new semiparametric M-quantile approach to small area prediction for binary data that extends the ideas of Cantoni and Ronchetti (2001) and Chambers and Tzavidis (2006). This predictor can be viewed as an outlier robust alternative to the more commonly used conditional expectation predictor (10) for binary data that is based on a logit GLMM with Gaussian random effects. With the proposed approach random effects are avoided and between-area variation in the response is characterized by variation in area-specific values of M-quantile indices. Furthermore, outlier robust inference is achieved in the presence of both misclassification and measurement error.

Under the M-quantile framework the propensity scores are estimated as:

$$\hat{\eta}_{ij} = \Lambda(\hat{e}(\mathbf{x}_{ij})) = \mathbf{x}_{ij}^T\hat{\boldsymbol{\alpha}}_{\hat{\bar{q}}_j}, \tag{12}$$

where the area level M-quantile coefficients are computed in a different way with respect to the continuous outcome. See Chambers et al. (2016) for details. Substituting the $\hat{y}_{ij}$ and $\hat{e}(x_{ij})$ in equation (7) provides the estimates of IPW-MQ. Note that this estimator is a special case of the equation (7) and so it is double robust and consistent.

Remark 3. if the size of the control group is zero in the small area j the value of the response variable can be estimated by $\hat{y}_{ij} = \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}}_{\hat{\tilde{q}}_j} + w_{ij}\hat{\gamma}_{\hat{\tilde{q}}_j}$, whereas if the size of the treated group in the area of interest is zero the response variable can be predicted by $\hat{y}_{ij} = \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}}_{\hat{\tilde{q}}_j}$.

4. MSE estimators in the finite population

In the context of randomized experiments Ding et al. (2019) proposed the decomposition of overall treatment effect variation into systematic and idiosyncratic components. In this paper we are in the framework of observational data and we are using the inverse propensity weighting; for this reason, we decompose the variation of the effect into the variation due to the estimation of the (i) outcome and the (ii) propensity scores. For the first component of variation we propose its estimation with an analytical derivation. In particular, for the IPW-EBLUP the proposal is based on the MSE estimation approach that is described in Prasad and Rao (1990) and represents an extension of the ideas in Opsomer et al. (2008). For IPW-MQ the MSE estimator is based on second order approximations to the variances of solutions of outlier robust estimating equations and represents an extension of the ideas in Chambers et al. (2014). The proposed analytical MSE estimators do not take into account the variability due to the estimation of the propensity scores. So to add this component of variability we suggest using re-sampling techniques. Miratrix et al. (2018) point out the importance of considering the extra variability that is introduced when estimating τ_{PATE} using weights, which is a similar problem to ours. In particular, for IPW-EBLUP we suggest using a parametric bootstrap technique, such as that proposed by Gonzalez-Manteiga et al. (2008) or a non-parametric bootstrap procedure as in Opsomer et al. (2008). For IPW-MQ, we suggest applying an outlier robust bootstrap estimator that is the modified version of the block-bootstrap approach of Chambers and Chandra (2013). These bootstrap methods are explained in detail in Supplementary Material, Section S.4.

We show in model-based simulation experiments (Section S.6) how these approaches can be useful for estimating the MSE of various small area predictors that are considered in this paper.

To develop the analytical MSE estimators for small area predictors based on EBLUP and MQ approaches, we rewrite the estimator in equation (7) as a linear combination of observed and unobserved outcomes:

$$\begin{aligned}\hat{\tau}_{PATE_j} &= K_j^{-1}\left(\sum_{i\in s_j}\left[\frac{w_{ij}y_{ij}}{\hat{e}(\mathbf{x}_{ij})}\right]+\sum_{i\in r_j}\left[\frac{w_{ij}\hat{y}_{ij}}{\hat{e}(\mathbf{x}_{ij})}\right]\right)- \\ &\quad T_j^{-1}\left(\sum_{i\in s_j}\left[\frac{(1-w_{ij})y_{ij}}{1-\hat{e}(\mathbf{x}_{ij})}\right]+\sum_{i\in r_j}\left[\frac{(1-w_{ij})\hat{y}_{ij}}{1-\hat{e}(\mathbf{x}_{ij})}\right]\right) \\ &= \sum_{i\in s_j} D_{ij}y_{ij}+\sum_{i\in r_j} D_{ij}\hat{y}_{ij}, \end{aligned} \tag{13}$$

where $K_j = \sum_{i=1}^{N_j} w_{ij}/\hat{e}(\mathbf{x}_{ij})$, $T_j = \sum_{i=1}^{N_j}(1-w_{ij})/(1-\hat{e}(\mathbf{x}_{ij}))$, and

$$D_{ij} = \left(\frac{K_j^{-1}w_{ij}}{\hat{e}(\mathbf{x}_{ij})} - \frac{T_j^{-1}(1-w_{ij})}{1-\hat{e}(\mathbf{x}_{ij})}\right).$$

4.1. MSE of the causal effect estimator IPW-EBLUP

We start from equation (13) to derive the analytic formula of the MSE for IPW-EBLUP. We consider that the D_{ij}s are known for the entire population, so we do not account for their variations originating from the estimation of the propensity scores. Therefore, if the proportion of observed outcomes, $f_j = \frac{n_j}{N_j}$, is small (negligible) we can write:

$$\hat{\tau}_{PATE_j} - \tau_{PATE_j} = \mathbf{D}_j^T\hat{\mathbf{y}}_j - \mathbf{D}_j^T\mathbf{y}_j = \mathbf{D}_j^T\left(\hat{\mathbf{y}}_j - \mathbf{y}_j\right),$$

where $\mathbf{D}_j$, $\hat{\mathbf{y}}_j$ and $\mathbf{y}_j$ are the vectors of D_{ij}s, the response variable and predicted outcomes, respectively, for the population in area j (Prasad and Rao, 1990). The prediction of outcome is obtained using the equation (9):

$$\hat{\mathbf{y}}_j = \tilde{\mathbf{X}}_j^T\hat{\tilde{\boldsymbol{\beta}}} + \tilde{\mathbf{W}}_j\hat{\boldsymbol{\gamma}} + \mathbf{Z}_j\hat{\mathbf{u}}, \tag{14}$$

where $\tilde{\mathbf{X}}_j$ is the matrix of auxiliary variables for area j of dimension $(p+1)\times N_j$, $\tilde{\mathbf{W}}_j$ is a sparse matrix with the jth column being replaced by the treatment status of individuals in area j, $\mathbf{Z}_j$ is a sparse matrix of area indicators with only the elements of column jth equal to one, so that $\text{var}(\mathbf{Y}) = \mathbf{V} = \tilde{\mathbf{W}}\boldsymbol{\Sigma}_\gamma\tilde{\mathbf{W}}^T + \mathbf{Z}\boldsymbol{\Sigma}_u\mathbf{Z}^T + \boldsymbol{\Sigma}_\epsilon$. Assuming that sampling is non-informative for the small area distribution of the response variable given the covariates, allows us to use population level models with the sample data. Where the variances of the random components are known, standard results from BLUP

theory (McCulloch and Searle, 2001, Chapter 9) guarantee that, given the model specifications (9) and Proposition 1, the generalized least squares estimator

$$\hat{\tilde{\boldsymbol{\beta}}} = \left(\tilde{\mathbf{X}}_s^T \mathbf{V}_s^{-1} \tilde{\mathbf{X}}_s\right)^{-1} \tilde{\mathbf{X}}_s^T \mathbf{V}_s^{-1} \mathbf{Y}_s$$

and the predictors

$$\hat{\boldsymbol{\gamma}} = \boldsymbol{\Sigma}_\gamma \tilde{\mathbf{W}}_s^T \mathbf{V}_s^{-1} \left(\mathbf{Y}_s - \tilde{\mathbf{X}}_s \hat{\tilde{\boldsymbol{\beta}}}\right)$$

$$\hat{\mathbf{u}} = \boldsymbol{\Sigma}_u \mathbf{Z}_s^T \mathbf{V}_s^{-1} \left(\mathbf{Y}_s - \tilde{\mathbf{X}}_s \hat{\tilde{\boldsymbol{\beta}}}\right)$$

are optimal among linear estimators and predictors, respectively. Here $\tilde{\mathbf{X}}_s$, $\mathbf{V}_s$, $\mathbf{Y}_s$, $\tilde{\mathbf{W}}_s$ and $\mathbf{Z}_s$ denotes the sample component of $\tilde{\mathbf{X}}$, $\mathbf{V}$, $\mathbf{Y}$, $\tilde{\mathbf{W}}$, $\mathbf{Z}$, respectively. Replacing $\hat{\mathbf{y}}_j$ with (14) we can write

$$\begin{aligned}\hat{\tau}_{PATE_j} - \tau_{PATE_j} &= \mathbf{D}_j^T \left(\hat{\mathbf{y}}_j - \mathbf{y}_j\right) \\ &= \mathbf{D}_j^T \mathbf{c}_j (\hat{\tilde{\boldsymbol{\beta}}} - \tilde{\boldsymbol{\beta}}) + \mathbf{D}_j^T \tilde{\mathbf{Z}}_j \left[\boldsymbol{\Sigma}_\omega \tilde{\mathbf{Z}}_s^T \mathbf{V}_s^{-1} \left(\mathbf{Y}_s - \tilde{\mathbf{X}}_s \tilde{\boldsymbol{\beta}}\right) - \omega\right],\end{aligned} \tag{15}$$

where $\mathbf{c}_j = \tilde{\mathbf{X}}_j^T - \left(\tilde{\mathbf{Z}}_j \boldsymbol{\Sigma}_\omega \tilde{\mathbf{Z}}_s^T \mathbf{V}_s^{-1} \tilde{\mathbf{X}}_s\right)$, $\tilde{\mathbf{Z}}_j = (\tilde{\mathbf{W}}_j, \mathbf{Z}_j)$, $\tilde{\mathbf{Z}} = (\tilde{\mathbf{W}}, \mathbf{Z})$, $\tilde{\mathbf{Z}}_s = (\tilde{\mathbf{W}}_s, \mathbf{Z}_s)$ and $\boldsymbol{\omega} = \left(\boldsymbol{\gamma}^T, \mathbf{u}^T\right)^T$, $\boldsymbol{\Sigma}_\gamma = \sigma_\gamma \mathbf{I}_m$, $\boldsymbol{\Sigma}_u = \sigma_u \mathbf{I}_m$. If both the random slopes and the area specific intercepts are treated as true random effects in the underlying model (9), the mean prediction error is 0 and the covariance between the two terms in equation (15) is also 0, so that the MSE of the prediction errors is

$$E\left[(\hat{\tau}_{PATE_j}^{BLUP} - \tau_{PATE_j})^2\right] = \mathbf{D}_j^T \tilde{\mathbf{Z}}_j \boldsymbol{\Sigma}_\omega \left(\mathbf{I}_{2m} - \tilde{\mathbf{Z}}_s^T \mathbf{V}_s^{-1} \mathbf{Z}_s \boldsymbol{\Sigma}_\omega\right) \tilde{\mathbf{Z}}_j^T \mathbf{D}_j + \mathbf{D}_j^T \mathbf{c}_j (\tilde{\mathbf{X}}_s^T \mathbf{V}_s^{-1} \tilde{\mathbf{X}}_s) \mathbf{c}_j^T \mathbf{D}_j. \tag{16}$$

To extend these results for IPW-EBLUP, that is, where $\mathbf{V}$ is unknown, the variation that comes from the estimation of variance components has to be added. The resulting EBLUP version of equation (15) is

$$\mathbf{D}_j^T \hat{\mathbf{c}}_j (\hat{\tilde{\boldsymbol{\beta}}} - \tilde{\boldsymbol{\beta}}) + \mathbf{D}_j^T \tilde{\mathbf{Z}}_j \left[\widehat{\boldsymbol{\Sigma}}_\omega \tilde{\mathbf{Z}}_s^T \widehat{\mathbf{V}}_s^{-1} \left(\mathbf{Y}_s - \tilde{\mathbf{X}}_s \tilde{\boldsymbol{\beta}}\right) - \omega\right], \tag{17}$$

with $\hat{\mathbf{c}}_j = \tilde{\mathbf{X}}_j^T - \left(\tilde{\mathbf{Z}}_j \widehat{\boldsymbol{\Sigma}}_\omega \tilde{\mathbf{Z}}_s^T \widehat{\mathbf{V}}_s^{-1} \tilde{\mathbf{X}}_s\right)$ using restricted maximum likelihood estimators for the unknown variance components in $\mathbf{V}$ and $\boldsymbol{\Sigma}_\omega$. To derive a second-order approximation for the MSE as well as an estimator for the MSE that is correct up to the second order we follow the method proposed by Opsomer et al. (2008). The vector of unknown components of the variance-covariance matrix is $\boldsymbol{\theta} = (\sigma_\gamma^2, \sigma_u^2, \sigma_\epsilon^2)$ and we define

$$\mathcal{S}_t = \mathbf{D}_j^T \tilde{\mathbf{Z}}_j \left(\frac{\partial \boldsymbol{\Sigma}_\omega}{\partial (\boldsymbol{\theta})_t} \tilde{\mathbf{Z}}_s^T \mathbf{V}_s^{-1} + \boldsymbol{\Sigma}_\omega \tilde{\mathbf{Z}}_s^T \frac{\partial \mathbf{V}_s^{-1}}{\partial (\boldsymbol{\theta})_t}\right), \qquad t = 1, 2, 3.$$

Further, we define the 3×3 matrix $\mathcal{I}$, as the Fisher information matrix with respect to the variance components $\boldsymbol{\theta}$, then, the MSE of the IPW-EBLUP predictor is given by

$$MSE(\hat{\tau}_{PATE_j}^{EBLUP}) = E\left[(\hat{\tau}_{PATE_j}^{BLUP} - \tau_{PATE_j})^2\right] + tr\left(\mathcal{S} \mathbf{V}_s \mathcal{S}^T \mathcal{I}^{-1}\right) + o(m^{-1}), \tag{18}$$

and its estimator can be obtained as

$$\begin{aligned}mse(\hat{\tau}_{PATE_j}^{EBLUP}) =& \mathbf{D}_j^T \tilde{\mathbf{Z}}_j \widehat{\boldsymbol{\Sigma}}_\omega \left(\mathbf{I}_{2m} - \tilde{\mathbf{Z}}_s^T \widehat{\mathbf{V}}_s^{-1} \tilde{\mathbf{Z}}_s \widehat{\boldsymbol{\Sigma}}_\omega\right) \tilde{\mathbf{Z}}_j^T \mathbf{D}_j + \mathbf{D}_j^T \hat{\mathbf{c}}_j (\tilde{\mathbf{X}}_s^T \widehat{\mathbf{V}}_s^{-1} \tilde{\mathbf{X}}_s) \hat{\mathbf{c}}_j^T \mathbf{D}_j \\ &+ 2 \left(\mathbf{Y}_s - \tilde{\mathbf{X}}_s \hat{\tilde{\boldsymbol{\beta}}}\right)^T \widehat{\mathcal{S}}^T \hat{\mathcal{I}}^{-1} \widehat{\mathcal{S}} \left(\mathbf{Y}_s - \tilde{\mathbf{X}}_s \hat{\tilde{\boldsymbol{\beta}}}\right),\end{aligned} \tag{19}$$

substituting $\boldsymbol{\theta}$ by the restricted maximum likelihood estimates in $\mathcal{S}$ and $\mathcal{I}$. Using some results of this section, asymptotic properties of the IPW-EBLUP are obtained.

Proposition 2. *Under assumptions 1-3 and the conditions (a), (b), (c) and (d) of Theorem 1 and the normality assumption on the random effects and the error terms, the estimator IPW-EBLUP is double robust and asymptotically normally distributed:*

$$\sqrt{N_j m}(\hat{\tau}_{PATE_j}^{EBLUP} - \tau_{PATE_j}) \sim \mathcal{N}(0, \mathscr{V}_j(\theta)),$$

as $m \to \infty$.

The proof of Preposition 2 is provided in the Section S.3 of the Supplementary Material.

4.2. MSE of the robust causal effect estimator IPW-MQ

In this section we propose an analytical derivation of the MSE for the IPW-MQ type estimator. This is based on the linearization ideas that are set out in Booth and Hobert (1998) and that are used by Chambers et al. (2014) to propose a new estimator of the MSE of a small area estimator that is defined by the solution of a set of robust estimating equations. The MSE is the sum of prediction variance and squared bias term. The theoretical development, as in Chambers et al. (2014), is based on approximations that correspond to assuming that $max(n_j) = O(1)$, so that, as the number of small areas tends to infinity, the prediction variance and the squared bias are $O(1)$. We also make the standard assumption that a consistent estimator of the MSE of a linear approximation to the small area estimator of interest can be used as its MSE estimator. As noted by Harville and Jeske (1992), such an approach will not generally be consistent, and the resulting MSE estimator can be downward biased. However, in small sample problems, this is not generally an issue.

Note that we assume that the $\bar{q}_j$ values are known. The prediction error of the IPW-MQ estimator is then:

$$\hat{\tau}^{MQ}_{PATE_j} - \tau_{PATE_j} = \sum_{i\in r_j} D_{ij}\hat{y}_{ij} - \sum_{i\in r_j} D_{ij}y_{ij}, \tag{20}$$

where $\hat{y}_{ij} = \mathbf{x}_{ij}^T\hat{\boldsymbol{\beta}}_{\bar{q}_j} + w_{ij}\hat{\gamma}_{\bar{q}_j}$. Following Chambers et al. (2014) the prediction variance of IPW-MQ estimator is:

$$\mathrm{var}(\hat{\tau}^{MQ}_{PATE_j} - \tau_{PATE_j}|\bar{q}_j) = \sum_{i\in r_j}\left\{D_{ij}^2\begin{pmatrix}\mathbf{x}_{ij} & w_{ij}\end{pmatrix}^T \mathrm{var}\begin{pmatrix}\hat{\boldsymbol{\beta}}_{\bar{q}_j}\\ \hat{\gamma}_{\bar{q}_j}\end{pmatrix}\begin{pmatrix}\mathbf{x}_{ij} & w_{ij}\end{pmatrix}\right\} + \sum_{i\in r_j} D_{ij}^2\,\mathrm{var}(y_{ij}). \tag{21}$$

A first order approximation to $\mathrm{var}(\hat{\boldsymbol{\beta}}_{\bar{q}_j}, \hat{\gamma}_{\bar{q}_j})$ is obtained following Chambers et al. (2014) and Bianchi and Salvati (2015). These approximated expressions lead to the following sandwich estimator:

$$\widehat{\mathrm{var}}\begin{pmatrix}\hat{\boldsymbol{\beta}}_{\bar{q}_j}\\ \hat{\gamma}_{\bar{q}_j}\end{pmatrix} = \frac{n}{(n-p-1)}\frac{\sum_{j=1}^m\sum_{i\in s_j}\psi^2\left(\omega_{ij}^{-1}(y_{ij}-\mathbf{x}_{ij}^T\hat{\boldsymbol{\beta}}_{\bar{q}_j}-w_{ij}\hat{\gamma}_{\bar{q}_j})\right)}{\left\{\sum_{j=1}^m\sum_{i\in s_j}\psi'\left(\omega_{ij}^{-1}(y_{ij}-\mathbf{x}_{ij}^T\hat{\boldsymbol{\beta}}_{\bar{q}_j}-w_{ij}\hat{\gamma}_{\bar{q}_j})\right)\right\}^2}\left(\begin{pmatrix}\tilde{\mathbf{X}}_s & \tilde{\mathbf{W}}_s\end{pmatrix}^T\begin{pmatrix}\tilde{\mathbf{X}}_s & \tilde{\mathbf{W}}_s\end{pmatrix}\right)^{-1}, \tag{22}$$

where ω_{ij} is a robust estimator of the scale of the residual $y_{ij} - \mathbf{x}_{ij}^T\hat{\boldsymbol{\beta}}_{\bar{q}_j} - w_{ij}\hat{\gamma}_{\bar{q}_j}$ in area j. An estimator of the first-order approximation (21) is then

$$\widehat{\mathrm{var}}(\hat{\tau}^{MQ}_{PATE_j}|\bar{q}_j) = \sum_{i\in r_j}\left\{D_{ij}^2\begin{pmatrix}\mathbf{x}_{ij} & w_{ij}\end{pmatrix}^T\widehat{\mathrm{var}}\begin{pmatrix}\hat{\boldsymbol{\beta}}_{\bar{q}_j}\\ \hat{\gamma}_{\bar{q}_j}\end{pmatrix}\begin{pmatrix}\mathbf{x}_{ij} & w_{ij}\end{pmatrix}\right\} + \widehat{\mathrm{var}}(y_{ij})\sum_{i\in r_j}D_{ij}^2, \tag{23}$$

where $\widehat{\mathrm{var}}(y_{ij}) = (n-1)^{-1}\sum_{j=1}^m\sum_{i\in s_j}\left(y_{ij} - \mathbf{x}_{ij}^T\hat{\boldsymbol{\beta}}_{\bar{q}_j} - w_{ij}\hat{\gamma}_{\bar{q}_j}\right)^2$.

A corresponding estimator of the area-specific bias of the IPW-MQ estimator is

$$\hat{B}(\hat{\tau}^{MQ}_{PATE_j}|\bar{q}_j) = \sum_{k=1}^m\sum_{i\in s_k} c_{ij}(\mathbf{x}_{ik}^T\hat{\boldsymbol{\beta}}_{\bar{q}_k} + w_{ik}\hat{\gamma}_{\bar{q}_k}) - \sum_{i\in\mathcal{U}_j} D_{ij}\left(\mathbf{x}_{ij}^T\hat{\boldsymbol{\beta}}_{\bar{q}_j} + w_{ij}\hat{\gamma}_{\bar{q}_j}\right), \tag{24}$$

where $c_{ij} = b_{ij} + D_{ij}I(i\in j)$ and

$$\mathbf{b}_j = (b_i j) = \left(\sum_{i\in r_j} D_{ij}\begin{pmatrix}\mathbf{x}_{ij} & w_{ij}\end{pmatrix}\right)\mathbf{W}^{MQ}_{\bar{q}_j}\begin{pmatrix}\tilde{\mathbf{X}}_s & \tilde{\mathbf{W}}_s\end{pmatrix}\left(\begin{pmatrix}\tilde{\mathbf{X}}_s & \tilde{\mathbf{W}}_s\end{pmatrix}^T\mathbf{W}^{MQ}_{\bar{q}_j}\begin{pmatrix}\tilde{\mathbf{X}}_s & \tilde{\mathbf{W}}_s\end{pmatrix}\right)^{-1}.$$

The final expression for the estimator of the MSE of IPW-MQ is just the sum of equation (23) and the square of equation (24):

$$\widehat{MSE}(\hat{\tau}^{MQ}_{PATE_j}|\bar{q}_j) = \widehat{\mathrm{var}}(\hat{\tau}^{MQ}_{PATE_j}|\bar{q}_j) + \hat{B}^2(\hat{\tau}^{MQ}_{PATE_j}|\bar{q}_j). \tag{25}$$

Following the approach of Bianchi and Salvati (2015), a further adjustment to the approximation of the MSE is needed to account for the variation due to the estimation of the area M-quantile coefficient $\bar{q}_j$ in the equation (25). Therefore,

$$\mathrm{var}(\hat{\bar{q}}_j) = \begin{pmatrix}\tilde{\mathbf{X}}_j & \tilde{\mathbf{W}}_j\end{pmatrix}\mathbf{G}^T_{\bar{q}_j}\mathbf{G}_{\bar{q}_j}\begin{pmatrix}\tilde{\mathbf{X}}_j & \tilde{\mathbf{W}}_j\end{pmatrix}^T v^2_{\hat{\bar{q}}_j}, \tag{26}$$

where $\mathbf{G}_{\bar{q}_j} = n^{-1}\sum_{j=1}^m\left(\mathbf{H}^{-1}_{j\bar{q}_j}\left\{\partial_{\bar{q}_j}\mathbf{L}_{j\bar{q}_j} - \partial_{\bar{q}_j}\mathbf{H}_{j\bar{q}_j}\mathbf{H}^{-1}_{j\bar{q}_j}\mathbf{L}_{j\bar{q}_j}\right\}\right)$ with $\mathbf{H}_{j\bar{q}_j} = \tilde{\mathbf{X}}_j^T\mathbf{W}^{MQ}_{\bar{q}_j}\tilde{\mathbf{X}}_j$, $\mathbf{L}_{j\bar{q}_j} = \tilde{\mathbf{X}}_j^T\mathbf{W}^{MQ}_{\bar{q}_j}\tilde{\mathbf{y}}_j$, $\partial_{\bar{q}_j}\mathbf{H}_{j\bar{q}_j} = \tilde{\mathbf{X}}_j^T\partial_{\bar{q}_j}\mathbf{W}^{MQ}_{j\bar{q}_j}\tilde{\mathbf{X}}_j$, $\partial_{\bar{q}_j}\mathbf{L}_{j\bar{q}_j} = \tilde{\mathbf{X}}_j^T\partial_{\bar{q}_j}\mathbf{W}^{MQ}_{j\bar{q}_j}\tilde{\mathbf{y}}_j$, $\partial_{\bar{q}_j}\mathbf{W}^{MQ}_{j\bar{q}_j} = 2\boldsymbol{\Omega}_j\left|\psi\left\{\boldsymbol{\Omega}_j^{-1}(\tilde{\mathbf{y}}_j - \tilde{\mathbf{X}}_j^T\boldsymbol{\beta}_{\bar{q}_j}\right\}\right|\left\{\tilde{\mathbf{y}}_j - \tilde{\mathbf{X}}_j^T\boldsymbol{\beta}_{\bar{q}_j}\right\}^{-1}$, $\Omega_j = diag(\omega_{ij})$, $i\in s_j$ and $v^2_{\hat{\bar{q}}_j} = n_j^{-1}\sum_{i=1}^{n_j}(\hat{q}_{ij} - \hat{\bar{q}}_j)^2$ where $\hat{q}_{ij}$ are the M-quantile coefficients at unit level. This expression (26) can be estimated by

$$\widehat{\text{var}}(\hat{\bar{q}}_j) = \begin{pmatrix} \tilde{\mathbf{X}}_j & \tilde{\mathbf{W}}_j \end{pmatrix} \hat{\mathbf{G}}_{\bar{q}_j}^T \hat{\mathbf{G}}_{\bar{q}_j} \begin{pmatrix} \tilde{\mathbf{X}}_j & \tilde{\mathbf{W}}_j \end{pmatrix}^T \hat{v}_{\hat{\bar{q}}_j}^2. \tag{27}$$

The final form of the MSE estimator of $\hat{\tau}_{PATE_j}^{MQ}$ is then

$$mse(\hat{\tau}_{PATE_j}^{MQ}) = \widehat{\text{var}}(\hat{\tau}_{PATE_j}^{MQ}) + \hat{B}^2(\hat{\tau}_{PATE_j}^{MQ}) + \widehat{\text{var}}(\hat{\bar{q}}_j). \tag{28}$$

The validity of model-based inference depends on the validity of the model assumed. We empirically evaluate the properties of small area predictors and corresponding MSE estimators. In particular, we use the Monte Carlo simulation to evaluate the performance of the proposed small area estimators and their corresponding MSEs in comparison with the performance of the IPW-Direct estimator at small area level. Due to space constraints, the results and the discussion are not reported in the manuscript but they can be found in Supplementary Material, Section S.6. These results show that the proposed small area predictors, IPW-EBLUP and IPW-MQ, are much more efficient than the IPW-Direct and this suggests that it may be good to use these predictors to estimate the average treatment effect when the sample size in each area becomes small.

In addition, the benchmarking properties of the estimators are shown in Section S.5 of Supplementary Material.

5. A design-based simulation based on real data

In this section we perform a design based simulation study using the 2015 Italian module of the EU-SILC survey. The focus is on estimating the effect of permanent versus temporary contracts on the economic insecurity of households in different regions of Italy.

In the design based simulation we consider the following substantive policy issue. Suppose policy makers are interested in evaluating the impact of temporary employment contracts on the economic insecurity of households, measured by subjective poverty as defined in Kapteyn et al. (1988), with potential consequences on consumption behaviour, life satisfaction and well-being in general. The increase in non-standard forms of employment in many countries appears to have contributed to rising in-work poverty (Eurofound & the International Labour Office, 2017; Crettaz, 2013). The development of forms of flexible employment may have both positive and negative consequences. On the one hand it is expected to increase employment and reduce unemployment. On the other hand, this is often associated with greater economic insecurity and poorer working conditions. Relatively little research has been dedicated to the link between job instability and subjective poverty. According to some scholars, temporary as opposed to permanent employment contributes to lower general life satisfaction and well-being and a worse perceived household income situation. Scherer (2009) investigates the social consequences of insecure employment (fixed-term contracts), taking into account information on current family life, future family plans and general well-being. The analysis, for Western European countries, confirms that insecure employment is accompanied by more problematic social and family situations. These negative consequences are partly shaped by the specific institutional context (welfare state and labour market conditions). Filandri and Struffolino (2018), using the 2014 Italian wave module of the EU-SILC (EU Statistics on Income and Living Conditions) survey, find that subjective poverty is associated with instability of household members' job contracts, with effects on other life domains, such as well-being, adequate level of consumption, social integration.

Differently from the previous literature, in this paper we adopt a causal perspective and consider the effect of temporary employment on the feeling about the household economic status. As discussed above, an overall negative effect of temporary employment is expected compared to permanent employment. However, we expect the effect to be heterogeneous across Italian regions due to different quality and cost of living. The effects may be confounded by local institutional contexts (local welfare policies and labour market conditions), in addition to socio-demographics characteristics and information on the employment situation (work intensity and the skill level of the occupation). The presence of numerous regions with a very small sample size makes it difficult to obtain reliable direct estimates at the area level and motivates the use of SAE techniques. In this simulation each region of Italy is considered as a small area.

In this setting the units of the analysis are the Italian households and the treatment, Job stability, is a dichotomous variable that gets the value 1 if the head of the household (or the household respondent) has a temporary job and 0 if she/he has a permanent job at the time of the interview. We assume the existence of a causal path from this variable to the lowest monthly income to make ends meet, which is a subjective measure of the household economic status. This is one of the EU-SILC target variables in the domain of social exclusion/non-monetary household deprivation indicators. Respondents are asked to provide their own assessed indication of the very lowest net monthly income that the household would have to get in order to make ends meet, that is, to pay its usual necessary expenses. We use this continuous outcome as a proxy variable for subjective poverty in the following analysis. For the outcome model, as it is common for highly right skewed outcome distributions, we use the log transformation and then we consider the transformed values per individual in the household by dividing the total value by the equivalized household size. We consider two sets of plausible confounders and predictors at the individual and household level. The individual characteristics concern the head or the responsible person in the household. We assume unconfoundedness conditioning on the following set of covariates: `Age`, `Gender`, `Education`, `Marital status`, `Tenure`, `Family type`. In addition, the `Number of rooms` in the house, the `Dwelling type`, the existence of problems related to crime, violence and vandalism in the local area from the point of view of the respondents (`Crime`), and the `Household disposable income` are used as additional predictors of the outcome.

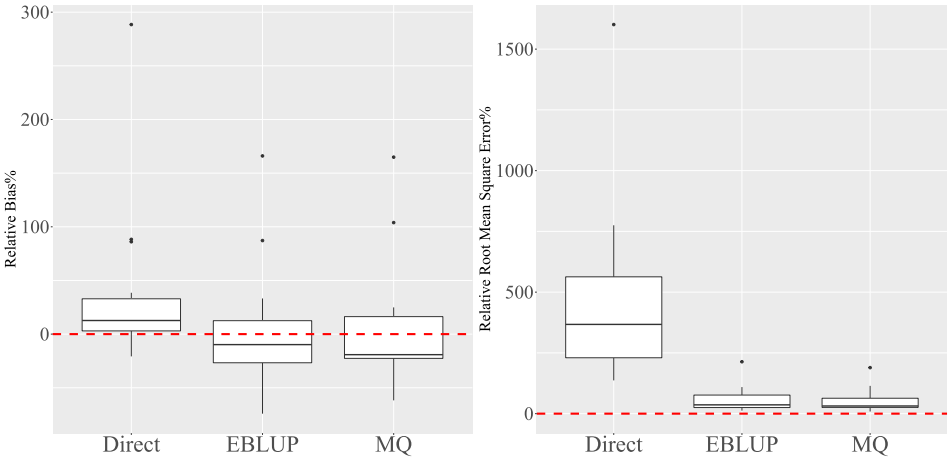

Fig. 1. Boxplots of the median values of area-specific relative bias and relative root mean square error computed over 1000 replications. Note that the Direct, EBLUP and MQ stand for IPW-Direct, IPW-EBLUP and IPW-MQ, respectively.

The aim of the design-based simulation is to compare the performance of different estimators for the impact in each domain under repeated sampling from a fixed population. For this reason we consider the sample of the workforce, aged between 25 and 80, in the 20 administrative regions of Italy, based on the 2015 Italian module of the EU-SILC survey as a pseudo-population (population hereafter). Due to sample size requirements Abruzzo and Molise are aggregated, leading to 19 areas. After accounting for common support within all areas 11011 units are left, from which 1254 units belong to the treated and the rest to the control group. The area population sizes range from 152 to 1329 with an average of 580. Figure S.10 in Section S.7 of the Supplementary Material shows the overall common support of the propensity scores among treated and control groups. In the same section we also report diagnostics on the balancing of covariates based on the estimated propensity score, both descriptive statistics (using the Standardized Mean Differences, SMD, and the maximum distance between empirical Cumulative Distribution Functions, eCDF) and inferential tests (running t-tests on the difference in the linearized estimated propensity score by treatment status within each area).

The original estimates of the impacts are considered as the true τ_j parameters at population level. This pseudo-population is then kept fixed over the Monte Carlo simulations. We draw $S = 1000$ independent random samples without replacement by randomly selecting individuals in the 19 regions with sample size of each area sets to 10% of its population size (resulting in a proportional stratified sampling). The samples from each region are drawn not considering the treatment status. This means that the sample of a specific region might include both treated and control units, or it might only contain the observations from one of the two groups. Three different estimators are evaluated in this simulation study: the IPW-Direct (4), the IPW-EBLUP (see Section 3.1.1) and the IPW-MQ (see Section 3.1.2). For the estimator based on M-quantile approach the influence function is the Huber-type function with tuning constant equal to 1.345 for the continuous response and 1.6 for the binary variable in the propensity scores estimation (Chambers and Tzavidis, 2006; Chambers et al., 2016). For each estimator and for each small area, we computed the Monte Carlo estimate of the percentage of relative bias and the percentage of relative root MSE and the corresponding efficiency.

Fig. 1 illustrates the box plots of the median values of area-specific relative bias and relative root MSE computed over replications, confirming the characteristics of the different estimators. We see that IPW-EBLUP and IPW-MQ work well in terms of both bias and relative root MSE compared with the IPW-Direct. This point is also highlighted in the series of model-based simulation studies in Section S.6 of the Supplementary Material. Fig. 2 shows that IPW-MQ, IPW-EBLUP and the IPW-Direct capture the heterogeneity of the average treatment effects over the areas. It illustrates that the distribution of the estimated effects by IPW-MQ, IPW-EBLUP (solid blue and red liens) and IPW-Direct (solid gray line) are close to the true distribution of the effects (dashed line). These results are confirmed in Table 1 where we report the distribution of the true and estimated average area treatment effects. However, the IPW-Direct shows a higher variability (almost 100 times) in the estimation of area specific treatment effects than the model-based predictors. The advantage of IPW-EBLUP and IPW-MQ in terms of variability is shown in the empirical confidence intervals presented in Fig. 3.

The relative efficiencies of the proposed estimators with respect to IPW-Direct are computed as the ratio of the average actual MSE for each area to the average actual MSE of the IPW-Direct. Table 2 presents the summary statistics over the 19 regions in the study. A value less than 100 for this ratio indicates that the MSE of the model-based estimate (i.e. IPW-EBLUP, IPW-MQ) is smaller than that of the direct estimate. The results reported in Table 2 indicate that the best method for this data appears to be the robust version, IPW-MQ. These results are consistent for all the areas in the study.

Fig. 3 illustrate the 95% Confidence Intervals that are obtained by sorting the 1000 estimates for each area and then taking the percentiles 2.5% and 97.5% as the lower and upper bound, respectively. In this illustration we can see that the length of the intervals for IPW-Direct estimator is much larger than our proposed IPW-EBLUP and IPW-MQ estimators due to the large variance of this estimator. This leads to the point that for all the 19 regions the CI of the direct estimator contain

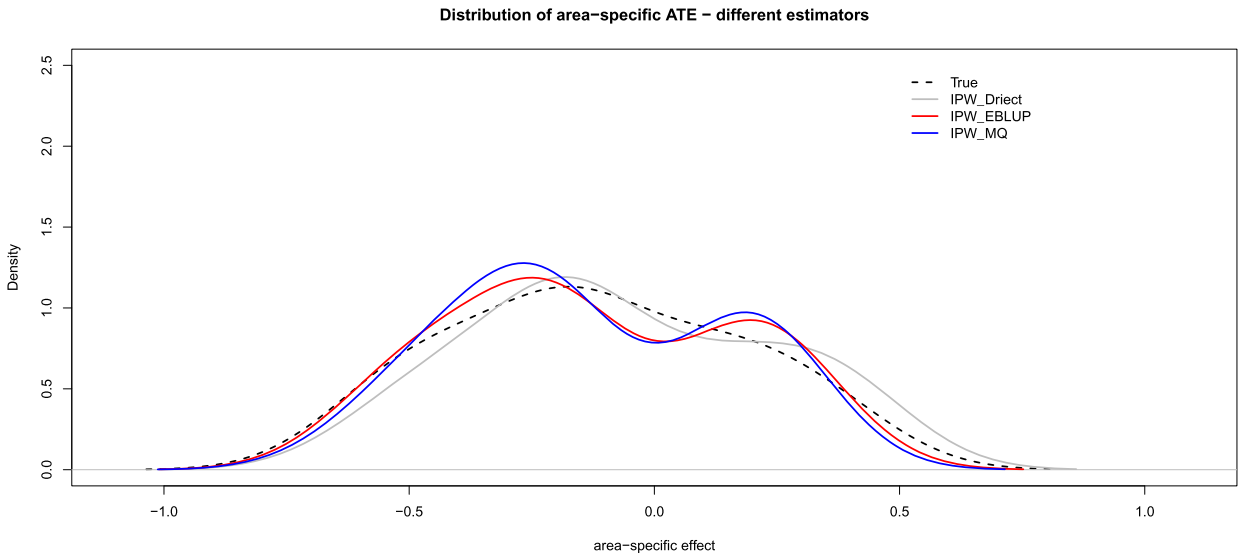

Fig. 2. The distribution of heterogeneous effects across areas. (For interpretation of the colours in the figure(s), the reader is referred to the web version of this article.)

Table 1
Summary statistics over 19 regions in the study of the true and the estimated average area treatment effects.

	min	Q1	Median	Mean	Q3	max
True	-0.5921	-0.3391	-0.1308	-0.1180	0.1023	0.3728
IPW-Direct	-0.5549	-0.2304	-0.1367	-0.0644	0.1825	0.4090
IPW-EBLUP	-0.5755	-0.3309	-0.1808	-0.1257	0.1744	0.3175
IPW-MQ	-0.5932	-0.3277	-0.2012	-0.1287	0.1639	0.2954

Table 2
The efficiency of each estimator compared to IPW-Direct. Summary statistics over 19 regions in the study.

Method	Min.	1st Qu.	Median	3rd Qu.	Max.
IPW-EBLUP	25.73	40.09	47.25	52.11	61.39
IPW-MQ	24.03	39.22	44.92	50.31	59.01

zero, implying that the direct method cannot not identify any significant effect and does not distinguish the heterogeneity of the effects among different areas. On the contrary, the CIs for IPW-EBLUP and IPW-MQ only contain zero in cases where the true area effects are very close to the zero line. Although the length of the intervals for our estimators are considerably lower than the direct estimator, they still mostly encompass the true values and manage to capture the heterogeneous effects among different regions.

6. Conclusion

Small area techniques provide official statistics for politicians and decision makers using sample surveys and other sources of information. However, to the extent of our knowledge there is no link between this literature and that on causal inference, even though sometimes the statements in the former literature are interpreted in a causal way.

In this paper we propose a methodological framework that links the two streams of literature and emphasize the relevance of such methods in many applications to real data. Our proposed methods take account of the heterogeneity of the effects across areas even at a very fine level (small area level). This allows policy makers and decision takers to know the impact of a given policy for a finer geographic, socio-demographic, or socio-economic grid and, consequently, to plan better local-targeted interventions.

Some of the usual assumptions for making causal inference with observational data are revisited and modified to be consistent with the context of small area estimations. The proposed methods IPW-EBLUP and IPW-MQ are mainly based on weighting with propensity scores. These estimators inherit the properties of doubly robust estimators, i.e., if one of the two

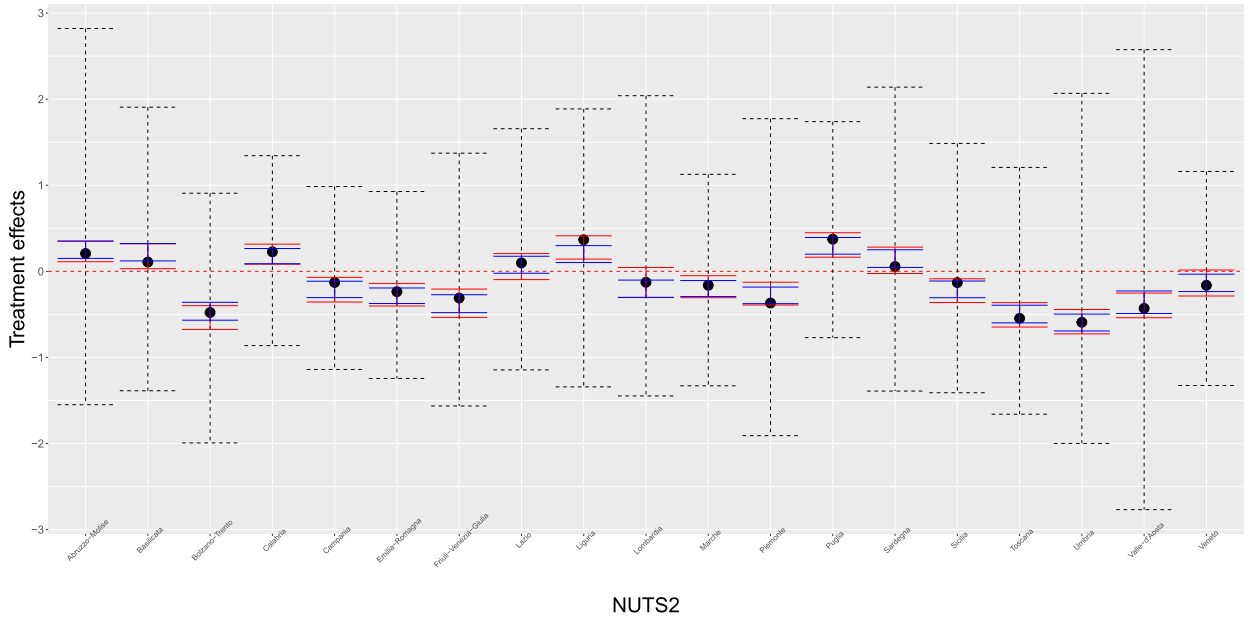

Fig. 3. 95% confidence intervals (CI) based on the quantiles of 1000 Monte Carlo replications. True values of the effect for the 19 NUTS2 area are depicted by black dots. CIs are shown for the Direct estimator in dashed black, for IPW-EBLUP in solid red and for IPW-MQ in solid blue line.

models (the one to estimate the propensity scores and the other to predict the outcome) is misspecified the estimator is still consistent.

For each of the proposed estimators, IPW-EBLUP and IPW-MQ, we developed an analytical MSE estimator under the assumption that the propensity score is known. We also suggest a correction for the bias in the analytical MSE, that can occur due to the estimation of the propensity score, by proposing two different bootstrap methods. They are defined as a parametric bootstrap and modified random effect block bootstrap, for IPW-EBLUP and IPW-MQ, respectively. The performance of the MSE estimators is studied via simulations.

Monte Carlo model based simulations are used to evaluate the performance of the proposed estimators in comparison with the performance of the IPW-Direct at small area levels. The results show that the proposed small area predictors, IPW-EBLUP and IPW-MQ, are much more efficient than the IPW-Direct and this suggests that it may be best to use these predictors to estimate the average treatment effect when the sample size in each area becomes small. However, as expected, these methods manifest higher bias than the direct estimator.

The application to real data, even if conducted as a design-based simulation analysis, has shown the potential of the proposed method in reconstructing the detail of the impact at the regional level, albeit with differences in the performance of the estimators. Job stability affects the perception of economic insecurity, but not in a homogeneous way in the different regions. The effect is negative in most cases with even significant differences, which we can attribute to the different levels of quality and cost of living, as well as to a different social context in general. Once again, this highlights the importance of adopting local policies to support families and combat poverty.

As future lines of research, we plan to extend our results for other robust estimators, such as REBLUP (Sinha and Rao, 2009). Moreover, due to the presence of bias in IPW-EBLUP and IPW-MQ, observed in our simulation experiment, we would like to investigate bias calibration methods to make the approach predictive rather than projective. We aim also to exploit the use of other matching techniques by properly defining distance measures and including the predicted random effects in the matching algorithm in small area estimation. Finally, in most small area estimation asymptotics, and also in this paper, m tends to infinity but n_js are bounded. As pointed out by Jiang and Lahiri (2006), the alternative asymptotic setting in which both number of areas m and area specific sample sizes n_j tend to ∞ (possibly at differential rates) is indeed an important problem. This alternative asymptotic framework has received relatively less attention in the SAE research. Lyu and Welsh (2021) recently put forward an asymptotic approach where both m and n_j are allowed to tend to infinity. The extension of the asymptotics of our proposal using this framework is an avenue of future research.

Acknowledgement

The authors gratefully acknowledge that the work of Nicola Salvati and Setareh Ranjbar has been partially carried out with the support of project InGRID 2 (grant agreement 730998, EU), and the work of Barbara Pacini and Nicola Salvati has been partially carried out with the support of project PRA2018-9 (‘From survey-based to register-based statistics: a paradigm shift using latent variable models’). The work of Setareh Ranjbar has also benefited from the financial supports of

the SNSF project (reference: 100018-178964). The authors are also grateful to Ray Chambers and Katarzyna Reluga for their insight and constructive comments in writing the current version of the paper.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.csda.2023.107742.

References

Arpino, B., Cannas, M., 2016. Propensity score matching with clustered data. An application to the estimation of the impact of caesarean section on the Apgar score. Stat. Med. 35, 2074–2091.
Arpino, B., Mealli, F., 2011. The specification of the propensity score in multilevel observational studies. Comput. Stat. Data Anal. 55, 1770–1780.
Bachtrögler, J., Fratesi, U., Perucca, G., 2020. The influence of the local context on the implementation and impact of EU cohesion policy. Reg. Stud. 54, 21–34.
Bang, H., Robins, J.M., 2005. Doubly robust estimation in missing data and causal inference models. Biometrics 61, 962–973.
Battese, G., Harter, R., Fuller, W., 1988. An error-components model for prediction of county crop areas using survey and satellite data. J. Am. Stat. Assoc. 83, 28–36.
Bianchi, A., Salvati, N., 2015. Asymptotic properties and variance estimators of the m-quantile regression coefficients estimators. Commun. Stat., Theory Methods 44, 2416–2429.
Booth, J.G., Hobert, J.P., 1998. Standard errors of prediction in generalized linear mixed models. J. Am. Stat. Assoc. 93, 262–272.
Breckling, J., Chambers, R., 1988. M-quantiles. Biometrika, 761–771.
Cafri, G., Wang, W., Chan, P.H., Austin, P.C., 2019. A review and empirical comparison of causal inference methods for clustered observational data with application to the evaluation of the effectiveness of medical devices. Stat. Methods Med. Res. 28, 3142–3162.
Cantoni, E., de Luna, X., 2018. Robust semiparametric inference with missing data. arXiv preprint arXiv:1803.08764.
Cantoni, E., Ronchetti, E., 2001. Robust inference for generalized linear models. J. Am. Stat. Assoc. 96, 1022–1030.
Chambers, R., Chandra, H., 2013. A random effect block bootstrap for clustered data. J. Comput. Graph. Stat. 22, 452–470.
Chambers, R., Chandra, H., Salvati, N., Tzavidis, N., 2014. Outlier robust small area estimation. J. R. Stat. Soc., Ser. B Stat. Methodol. 76, 47–69.
Chambers, R., Salvati, N., Tzavidis, N., 2016. Semiparametric small area estimation for binary outcomes with application to unemployment estimation for local authorities in the UK. J. R. Stat. Soc., Ser. A, Stat. Soc. 179, 453–479.
Chambers, R., Tzavidis, N., 2006. M-quantile models for small area estimation. Biometrika 93, 255–268.
Chan, W., 2018. Applications of small area estimation to generalization with subclassification by propensity scores. J. Educ. Behav. Stat. 43, 182–224.
Chandra, H., Salvati, N., Sud, U., 2011. Disaggregate-level estimates of indebtedness in the state of Uttar Pradesh in India: an application of small-area estimation technique. J. Appl. Stat., 2413–2432.
Chen, S., Haziza, D., 2019. Recent developments in dealing with item non-response in surveys: a critical review. Int. Stat. Rev. 87, S192–S218.
Crettaz, E., 2013. A state-of-the-art review of working poverty in advanced economies: theoretical models, measurement issues and risk groups. J. Eur. Soc. Policy 23, 347–362.
Ding, P., Feller, A., Miratrix, L., 2019. Decomposing treatment effect variation. J. Am. Stat. Assoc. 114, 304–317.
Eurofound, the International Labour Office, 2017. Working Anytime, Anywhere: The Effects on the World of Work.
Filandri, M., Struffolino, E., 2018. Individual and household in-work poverty in Europe: understanding the role of labor market characteristics. Eur. Soc. 21, 1–28.
Gonzalez-Manteiga, W., Lombardia, M., Molina, I., Morales, D., Santamaría, L., 2008. Analytic and bootstrap approximations of prediction errors under a multivariate Fay–Herriot model. Comput. Stat. Data Anal. 52, 5242–5252.
Hahn, J., 1998. On the role of the propensity score in efficient semiparametric estimation of average treatment effects. Econometrica, 315–331.
Harville, D.A., Jeske, D.R., 1992. Mean squared error of estimation or prediction under a general linear model. J. Am. Stat. Assoc. 87, 724–731.
Haziza, D., Rao, J., 2010. Variance estimation in two-stage cluster sampling under imputation for missing data. J. Stat. Theory Pract. 4, 827–844.
Hirano, K., Imbens, G.W., Ridder, G., 2003. Efficient estimation of average treatment effects using the estimated propensity score. Econometrica 71, 1161–1189.
Imbens, G.W., 2004. Nonparametric estimation of average treatment effects under exogeneity: a review. Rev. Econ. Stat. 86, 4–29.
Imbens, G.W., Wooldridge, J.M., 2009. Recent developments in the econometrics of program evaluation. J. Econ. Lit. 47, 5–86.
Jiang, J., 1999. On unbiasedness of the empirical blue and blup. Stat. Probab. Lett. 41, 19–24.
Jiang, J., 2010. Large Sample Techniques for Statistics. Springer Science & Business Media.
Jiang, J., Lahiri, P., 2006. Mixed model prediction and small area estimation (with discussions). Test 15, 1–96.
Kapteyn, A., Kooreman, P., Willemse, R., 1988. Some methodological issues in the implementation of subjective poverty definitions. J. Hum. Resour., 222–242.
Kim, G., Paik, M., Kim, H., 2017. Causal inference with observational data under cluster-specific non-ignorable assignment mechanism. Comput. Stat. Data Anal. 113, 88–99.
Li, F., Zaslavsky, A.M., Landrum, M.B., 2013. Propensity score weighting with multilevel data. Stat. Med. 32, 3373–3387.
Liang, K.-Y., Zeger, S.L., 1986. Longitudinal data analysis using generalized linear models. Biometrika 73, 13–22.
Lunceford, J.K., Davidian, M., 2004. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. Stat. Med. 23, 2937–2960.
Lyu, Z., Welsh, A.H., 2021. Asymptotics for eblups: nested error regression models. J. Am. Stat. Assoc. https://doi.org/10.1080/01621459.2021.1895178.
McCulloch, C.E., Searle, S.R., 2001. Generalized, Linear, and Mixed Models.
Miratrix, L.W., Sekhon, J.S., Theodoridis, A.G., Campos, L.F., 2018. Worth weighting? How to think about and use weights in survey experiments. Polit. Anal. 26, 275–291.
Opsomer, J., Claeskens, G., Ranalli, M., Kauermann, G., Breidt, J., 2008. Non-parametric small area estimation using penalized spline regression. J. R. Stat. Soc., Ser. B 70, 265–286.
Pinheiro, J., Bates, D., 2006. Mixed-Effects Models in S and S-PLUS. Springer Science & Business Media.
Prasad, N., Rao, J., 1990. The estimation of the mean squared error of small-area estimators. J. Am. Stat. Assoc. 85, 163–171.
Robins, J.M., Rotnitzky, A., Zhao, L.P., 1995. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. J. Am. Stat. Assoc. 90, 106–121.
Rosenbaum, P.R., Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70, 41–55.
Rubin, D.B., 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. J. Educ. Psychol. 66, 688.
Rubin, D.B., 1980. Randomization analysis of experimental data: the Fisher randomization test comment. J. Am. Stat. Assoc. 75, 591–593.
Rubin, D.B., Stuart, E.A., Zanutto, E.L., 2004. A potential outcomes view of value-added assessment in education. J. Educ. Behav. Stat. 29, 103–116.

Scherer, S., 2009. The social consequences of insecure jobs. Soc. Indic. Res. 93, 527–547.
Sinha, S., Rao, J., 2009. Robust small area estimation. Can. J. Stat. 37, 381–399.
Stuart, E.A., Cole, S.R., Bradshaw, C.P., Leaf, P.J., 2011. The use of propensity scores to assess the generalizability of results from randomized trials. J. R. Stat. Soc., Ser. A, Stat. Soc. 174, 369–386.
Stuart, E.A., Cole, S.R., Leaf, P.J., 2015. Assessing the generalizability of randomized trial results to target populations. Prev. Sci. 16, 475–485.
Wendling, T., Jung, K., Callahan, A., Schuler, A., Shah, N.H., Gallego, B., 2018. Comparing methods for estimation of heterogeneous treatment effects using observational data from health care databases. Stat. Med. 37, 3309–3324.
Wooldridge, J.M., 2007. Inverse probability weighted estimation for general missing data problems. J. Econom. 141, 1281–1301.
Zanutto, E.L., 2006. A comparison of propensity score and linear regression analysis of complex survey data. J. Data Sci. 4, 67–91.