

Evaluating synthetic population approaches for Italian municipal-level microdata

Silvia Scarpa ^{a,b} ^{*}, Massimo Baldini ^{a,b} , Andrea Barigazzi ^b , Giovanni Gallo ^{a,b} ,
Marcello Morciano ^{a,b} 

^a Department of Economics Marco Biagi, University of Modena and Reggio Emilia, Modena, Italy

^b CAPP - Centro di Analisi delle Politiche Pubbliche, Modena, Italy

ARTICLE INFO

Keywords:

Synthetic population
Spatial microsimulation
Reweighting algorithm
Policy evaluation

ABSTRACT

The growing demand for fine-scale microdata to support *ex-ante* and *ex-post* small area public policy evaluation is often hampered by a familiar trade-off: socio-economic surveys offer rich individual and household information but only at coarse geographic levels, while administrative sources provide detailed local counts with limited variable depth. This paper addresses this challenge by systematically implementing and evaluating deterministic (Iterative Proportional Fitting, Generalized Regression) and probabilistic (Simulated Annealing) reweighting algorithms to align survey weights with known population benchmarks. As a practical application, we construct a spatially disaggregated synthetic population representing the residents of the municipalities in the Emilia-Romagna region (NUTS 2 level) from the Italian EU-SILC data. Our comparative analysis highlights the trade-offs between the computational efficiency of deterministic methods and the precision of probabilistic approaches. By providing a robust foundation for spatial microsimulation models, this work offers practical guidance for researchers navigating the constraints of method fidelity, computational resources, and statistical accuracy.

1. Introduction

There is a growing demand for microdata at the territorial level, to evaluate public policies and monitor economic inequality and poverty [1,2]. However, researchers often face a significant lack of information at the desired level of aggregation. Subject to disclosure constraints, researchers may access individual census or administrative records, which provide fine-grained geographic detail but remain limited in the set of information available [3,4]. Conversely, survey data are rich in variables, e.g. income, consumption, wealth, and household characteristics, but typically lack detailed geographic identifiers due to confidentiality constraints. Moreover, survey data often suffer from issues such as top-coding [5], potential bottom-coding [6], and selective non-response and under-reporting [7–9], which introduce distortions into socio-economic estimates.

These issues are acknowledged across diverse research traditions; although they may lead to different interpretations, they nonetheless converge in identifying systematic biases. For example, with respect to the evolution of inequality, Pinkovskiy et al. [10] document declines in both between- and within-country inequality, while showing

that survey-based estimates are affected by misreporting, with under-reporting increasingly concentrated among lower-income groups. In contrast, Piketty et al. [11] emphasize that household surveys capture the upper tail of the distribution poorly and are not consistent with macroeconomic aggregates. Moreover, measurement and classification errors can distort the identification of the determinants of inequality, and the use of administrative data reveals patterns that differ from those suggested by survey data alone [12].

The integration of administrative sources can therefore improve the accuracy of estimates, but these sources generally lack many of the socioeconomic variables collected in surveys. This data gap gives rise to two main problems: policy evaluation below the regional level becomes unfeasible and the estimation of local socio-economic indicators suffers from inadequate statistical precision [13].

Statistical and economic research has addressed these issues through different strands of literature, including small area estimation (SAE) and spatial microsimulation (SM). SAE focuses on improving the precision of aggregated socio-economic indicators estimators at the small area level, which typically includes areas with few or no sampled

* Corresponding author at: Department of Economics Marco Biagi, University of Modena and Reggio Emilia, Modena, Italy.

E-mail addresses: silvia.scarpa@unimore.it (S. Scarpa), massimo.baldini@unimore.it (M. Baldini), andrea.barigazzi@unimore.it (A. Barigazzi), giovanni.gallo@unimore.it (G. Gallo), marcello.morciano@unimore.it (M. Morciano).

<https://doi.org/10.1016/j.seps.2026.102547>

Received 17 December 2025; Received in revised form 8 May 2026; Accepted 19 June 2026

Available online 23 June 2026

0038-0121/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

units. SAE models combine survey data with auxiliary information, sourced from censuses or administrative records. For detailed reviews, see Pfeffermann [14], Rao and Molina [15], Morales et al. [16]. While statistically rigorous, SAE models face two fundamental limitations. First, they require geographic identifiers in the survey data, which are often unavailable. Second, SAE is designed to produce estimates of specific indicators (e.g., mean income, poverty rates; see, for instance, Tonutti et al. [17]) rather than generating complete synthetic populations. These constraints limit their applicability: they enable monitoring of poverty and inequality through aggregate statistics but cannot support policy evaluation exercises that require individual-level data.

A complementary approach is survey-to-survey imputation, which provides an alternative strategy for recovering micro-level welfare information or territorial estimates when direct data are incomplete or unavailable. Although methodologically different from the synthetic reweighting framework, it shares similarities with small area estimation in that it uses models to produce disaggregated estimates, but relies only on another survey with common auxiliary information rather than census data [18–20].

Spatial Microsimulation (SM) offers a valuable alternative, though less rooted in the statistical literature, by focusing on the synthetic generation of microdata for small areas of interest, with the aim of reconstructing the target population while preserving its complexity and heterogeneity [21]. Microsimulation refers to the use of microdata to assess the effectiveness of a policy or the consequences of a general event by simulating its impact on individuals. SM inserts a spatial or geographic dimension into the simulation process, enabling evaluations at the small area level [22]. The applications of SM models spans a wide range of fields, including demography, tax-benefit systems, labour markets, urban planning, transport, and health care [23]. As noted by Alfons et al. [24] and Gussenbauer et al. [25], the goal is to construct a synthetic population that is statistically “close to reality” while maintaining the key characteristics of the original survey. To achieve this, a variety of algorithms have been developed — see Hermes and Poulsen [26], Tanton [27], Templ et al. [28] for a complete literature review. The resulting synthetic population at the desired small area level is crucial to microsimulate the implementation and impact of local policies and social transfers. Moreover, these synthetic populations can be used to estimate spatial indicators even in areas where the original survey data lack geographic identifiers, thus complementing and in some cases encompassing SAE methods.

This work presents a structured experimental exercise that rigorously compares and assesses different SM techniques. A central focus of this work is the systematic comparison of deterministic reweighting methods, namely Iterative Proportional Fitting (IPF) and the Generalized Regression Estimator (GREG), and a probabilistic approach based on combinatorial optimization, specifically Simulated Annealing (SA).

While the algorithms employed are established in the literature, this paper contributes to the spatial microsimulation in several ways. First, we provide a systematic and statistically grounded framework for the assessment of reweighting methods in a data-scarce context, including a reproducible benchmark-selection strategy. Second, we quantify the trade-offs between deterministic and probabilistic methods, specifically regarding computational efficiency and accuracy (i.e., the role of integration bias and its implications for small-area estimation). Third, we demonstrate the feasibility and reliability of the analysed methods on a complex real-world case: a territorial disaggregation at the municipal level in the complete absence of geographic identifiers in the survey data.

To achieve this, we construct a spatially disaggregated synthetic population representing the residents in the municipalities of the Italian region Emilia-Romagna, based on 2022 EU-Survey on Income and Living Conditions (SILC) data and a comprehensive set of municipality-level socio-demographic constraints derived from administrative data.

This territorial disaggregation, to our knowledge, has not been previously addressed with this degree of statistical rigour. By systematically documenting the trade-offs of each method across 330 municipalities with highly heterogeneous characteristics, ranging from rural areas with fewer than 100 inhabitants to the regional capital with nearly 400,000 residents, our analysis provides practical guidance for researchers and policymakers facing similar data constraints in other national or regional context, thus extending beyond the specific case study. The entire framework can be implemented with any survey microdata and publicly available administrative aggregates, making it transferable and replicable in other countries or subnational contexts.

The rest of the paper is organized as follows: Section 2 illustrates the reweighting methods that we employ. Section 3 describes the survey data and how we have derived the benchmark vector from available administrative data. Section 4 provides our main results, a detailed comparison of the trade-offs of the methods, and some distributional validations. A final Section is devoted to critically evaluate our main findings and extract lessons from the many trade-offs we faced to guide other researchers for building synthetic populations.

2. Methods for spatial microsimulation

Survey and population data may be used to generate a synthetic population through a spatial microsimulation (SM) model, whose quality depends on its statistical similarity to the target population [28]. Define $U = \{1, \dots, N\}$ as the finite population of interest, for example a country or a region, partitioned into D non-overlapping areas (e.g. single cities and towns or suburbs), each of size N_d , where $\sum_d N_d = N$. A sample $S \subset U$ of size $n < N$ is drawn from U according to a complex sampling design (e.g., by National Statistical Institutes).

Our goal is to artificially reconstruct the target population U within its D constituent areas. Each sampled individual j , $j \in S$, is subjected to a survey and associated with a sampling weight w_j , defined primarily as the inverse of the inclusion probability, and may be calibrated to known totals or adjusted for non-response. This weight represents the number of individuals in the target population represented by the sampled individual [29]. We assume that a matrix X of dimensions $D \times P$ containing area-specific benchmarks (known totals) is available from census or administrative sources. Let $X_d = (X_{d,1}, \dots, X_{d,P})^T$ denote the vector of P known totals for area d ($d = 1, \dots, D$). Correspondingly, let $x_j = (x_{j,1}, \dots, x_{j,P})^T$ be the vector of variables observed for individual j in the sample, matching the definitions of the benchmarks.

Crucially, the survey often lacks identifiers for the specific small area d to which individual j belongs. Consequently, the entire sample S (or a relevant subset) serves as a donor pool, and the reweighting process estimates area-specific weights $w_{j,d}^*$ such that the aggregated characteristics match the benchmarks X_d for each area.

As highlighted by Hermes and Poulsen [26], SM models require several a priori choices for the synthetic data generation process: the definition of the areas of interest; the choice of the model or algorithm to be applied; the identification of appropriate performance measures; the selection of the input data S and X .

2.1. On the choice of areas of interest

The definition of an “area” d ($d = 1, \dots, D$) is flexible and can vary depending on the application. It may refer to administrative units such as municipalities, provinces, or regions, but can also include custom aggregations based on policy relevance, functional geography, or socio-economic groups. Areas in this context are typically small in population size and are often poorly or not at all represented in the survey data. A strong requirement is the availability of a matrix X of known totals from administrative or census sources, which are assumed to be free of sampling error and cover the entire target population at the area level.

2.2. On the choice of the model

Spatial microsimulation literature comprises a wide range of methods, which can be probabilistic – such as synthetic reconstruction and combinatorial optimization – or deterministic methods, typically Iterative Proportional Fitting and the Generalized Regression Estimator. The latter two, together with combinatorial optimization, more generally belong to the class of reweighting approaches. All of them are reviewed in this section.

2.2.1. Synthetic reconstruction

When the input data in an SM model consist solely of published contingency tables, the *synthetic reconstruction* (SR) approach can be used to randomly generate a synthetic population from a series of observed conditional probabilities [30]. SR methods typically rely on random sampling or Monte Carlo techniques and often incorporate Iterative Proportional Fitting (IPF). IPF – also known as the raking method, as fully explained in Tillé [31] – allows the combination of partial marginal distributions to estimate a more complete joint distribution. A key advantage of this method is that it can be applied even when no survey microdata are available. However, there are also notable drawbacks. First, the process is entirely random, which means that each run may yield a different result. Second, the method can be designed to match the joint distribution either at the individual level or at the household level, but not both simultaneously [32]. Third, IPF matches the constraint variables sequentially rather than simultaneously, adjusting one at a time in an iterative loop until convergence is reached [27]. The SR method using IPF has been used in the SYNTHESIS model, as described by Birkin and Clarke [33].

It is important to note that while IPF is used within SR methods to estimate joint distributions from marginal tables, it can also be applied directly as a reweighting method when microdata are available, as discussed in the following section. Beyond IPF-based approaches, other SR methods include the use of Bayesian networks [34] and hidden Markov models [35].

2.2.2. Reweighting methods

Reweighting methods make use of all available data sources, as they aim to align the sample data S with a set of benchmarks (or constraints) X derived from census or administrative data at the small area level. A benchmark X_p may correspond to a univariate variable (e.g. age), or a multivariate combination (e.g. age by sex). Williamson et al. [36] highlighted the importance of selecting constraint tables that include equally both household-level and individual-level information. This selection should strike a balance: on the one hand, increasing the P number of constraints improves the similarity between the reweighted sample S^* and the target population U . On the other hand, including too many constraints can reduce the likelihood that the reweighting algorithm will converge, due to the difficulty of satisfying all the marginal totals simultaneously. Moreover, benchmarks should be chosen to ensure strong explanatory power for the target variables that will be analysed in subsequent microsimulation-based policy evaluations. For example, Chin et al. [22] performed a regression analysis to select the variables of interest.

Reweighting methods can be deterministic or probabilistic. Deterministic procedures iteratively adjust the initial sampling weights to match known population totals at the small area level. The calibration process continues until convergence is achieved, meaning the weighted survey totals closely match the administrative benchmarks. Once convergence is reached, the resulting calibrated weights must be rounded to integers to enable replication of survey individuals and generation of a synthetic population that preserves the full set of survey variables. Probabilistic methods, by contrast, produce integer weights by design. They randomly draw individuals from the sample to match known constraints at the area level, evaluating all constraints simultaneously rather than sequentially.

Iterative Proportional Fitting (IPF). Among the deterministic ones, Iterative Proportional Fitting (IPF) has the longest tradition [37]. IPF requires the initial set of weights for each individual $j \in S$ – often set equal to 1 – and adjusts them through a generalized raking procedure [38,39]. This procedure sequentially updates each weight by iteratively adjusting for each constraint $p = 1, \dots, P$, independently to each area d . Specifically, at step t for constraint p and area d , the updated weights are computed as

$$w_{j,d}^{(t,p)} = w_{j,d}^{(t,p-1)} \frac{X_{d,p}}{\widehat{X}_{d,p}^{(t,p-1)}}, \quad \forall j \in S,$$

where $\widehat{X}_{d,p}^{(t,p-1)} = \sum_{j \in S} w_{j,d}^{(t,p-1)} x_{j,p}$ is the estimated total for constraint p in area d . After cycling through all P constraints, the updated weights for the next iteration are set as $w_{j,d}^{(t+1)} = w_{j,d}^{(t,P)}$. This iterative process continues until convergence is achieved, meaning that the maximum relative change in weights across two successive iterations falls below a predefined threshold, or until a maximum number of iterations is reached. Key limitations of IPF include its sequential adjustment of constraints, which can lead to slower convergence when constraints are highly correlated or numerous. Additionally, IPF may fail to converge when constraints are incompatible or when the solution space is empty. IPF was, for example, used by Ballas et al. [40] in the development of the SimBritain spatial dynamic microsimulation model.

Generalized Regression Estimator (GREG). The Generalized Regression Estimator (GREG), originally introduced by Deville and Särndal [41], represents a robust deterministic reweighting method grounded in the regression relationship between the target variable and the auxiliary information. As detailed by Tanton et al. [42], GREG starts from the initial survey weights w_j , $j \in S$, and seeks a new set of area-specific weights $w_{j,d}^*$ that are as close as possible to the original ones, while ensuring that the calibration constraints are satisfied. The outcome of GREG depends on the choice of the distance function $D(w_j, w_{j,d}^*)$ used during the calibration process and solved by minimizing a Lagrangian function [31, Chap. 12]. Among the most common choices are the linear, truncated chi-squared, and raking ratio distance functions. Creedy [43] analysed the performance differences among these alternatives, while Pacifico [44] provided a broader review and comparison of calibration estimators. It is generally recommended to use distance functions that do not produce negative weights. In this work, we focus on the truncated chi-squared distance, which constrains the calibrated weights $w_{j,d}^*$ to lie within user-defined bounds L and H . This is defined as

$$D(w_j, w_{j,d}^*) = \begin{cases} \frac{(w_{j,d}^* - w_j)^2}{2w_j} & L < w_{j,d}^* < H \\ \infty & \text{otherwise.} \end{cases}$$

The minimization of the Lagrange function with respect to $w_{j,d}^*$ based on this distance function does not have a closed-form solution and must be solved iteratively, e.g. by Newton–Raphson method. In our implementation, for each area d , we set $L = 0$ to ensure non-negative weights and $H = N_d$ to allow sufficient flexibility in small areas, where a single sample unit may need to represent a large portion of the local population. While bounding the weights helps prevent extreme values, it also makes convergence more challenging. GREG has been widely adopted in practice. For example, the Australian microsimulation model developed at NATSEM uses the GREGWT routine implemented in SAS.

Probabilistic reweighting methods. Among probabilistic reweighting methods, combinatorial optimization (CO), originally proposed by Williamson et al. [36], is certainly the most well known. As described by Tanton et al. [45], CO consists of an iterative process that searches for the optimal combination of individuals or households by swapping individuals in and out from the survey microdata to represent each small area, based on benchmark totals. Conceptually, this is equivalent

to initializing all weights to zero and then populating the small area of size N by selecting N units, effectively assigning integer weights corresponding to the frequency of selection. To evaluate the quality of a specific combination, a goodness-of-fit metric, which we call for now generally *fit*, is required. We will give more details below.

Williamson et al. [36] reviewed three CO techniques: hill climbing, simulated annealing, and genetic algorithms. Among these, simulated annealing (SA) is the most extensively studied and applied. As explained by Harland et al. [46], SA is inspired by thermodynamics: it mimics the physical annealing process, in which a material is first heated to a high temperature and then slowly cooled to reach a stable structure. In optimization, this analogy translates into a Metropolis algorithm that allows both backward and forward steps to be taken when searching for an optimal solution. At high temperatures, the algorithm is more likely to accept poorer solutions, which helps it escape local optima. As the temperature decreases, it becomes more selective. This feature enables the algorithm to avoid getting trapped in local regions, thereby increasing the chance of converging to a near-global optimum. Formally, the SA algorithm implemented for each area d proceeds as follows:

1. Set the initial temperature T_0 , the cooling factor $\alpha \in (0, 1)$ and the maximum number of iterations Max_{iter} ;
2. Draw an initial sample $s^{(0)}$ of N_d individuals with replacement from the survey data S ;
3. Compute the initial goodness-of-fit, denoted fit_0 , between the marginal totals of the synthetic population and the known benchmarks.
4. Repeat the following steps until a stopping condition is met (i.e., maximum number of iterations reached, at iteration k $fit_k = 0$, or the temperature falls below a small threshold ϵ):
 - (a) Propose at step k a new sample $s^{(k)}$ by replacing one (or more) individual from the current sample with another drawn from the survey data;
 - (b) Compute fit_k based on the updated sample;
 - (c) Calculate the change in error $\Delta_k = fit_k - fit_{k-1}$. Accept the new sample if $\Delta_k < 0$. Otherwise, draw $u \sim \mathcal{U}(0, 1)$; if $u < P_{accept} = \exp\left(-\frac{\Delta_k}{T_k}\right)$, accept the new solution;
 - (d) Update the temperature using the cooling factor as $T_{k+1} = \alpha \cdot T_k$.

SA primary advantages lie in the simultaneous evaluation of constraints, which eliminates sensitivity to constraint ordering, and the generation of integer weights by design. By directly selecting discrete units, SA bypasses the integerisation bias inherent in deterministic methods, making it particularly suitable for microsimulation frameworks such as the Irish SMILE model [47].

2.2.3. Summary of methodological choices

Since this study relies on survey microdata (EU-SILC) alongside aggregate marginal totals from census and administrative sources, we adopt reweighting approaches rather than synthetic reconstruction (SR) methods. As noted by Tanton [27], SR is primarily reserved for scenarios where microdata is unavailable; conversely, reweighting allows us to preserve the correlation structure inherent in the sample.

Deterministic reweighting methods, such as IPF and GREG, offer significant computational efficiency and fully exploit the information contained in the original sampling weights. They can be effectively applied for the direct estimation of socio-economic indicators, such as mean of the target variable or inequality measures, as in Chin et al. [22]. However, a critical limitation arises from the nature of the output: these methods generate fractional (non-integer) weights. To generate a synthetic population at the area-level suitable for spatial

microsimulation modelling, these weights must be converted into integers to replicate survey observations. Defining the rounded calibrated weights for area d as $\lfloor w_{j,d}^* \rfloor$, the approximated benchmark total for variable p becomes $\lfloor \hat{X}_{d,p}^* \rfloor = \sum_{j \in S} \lfloor w_{j,d}^* \rfloor x_{j,p}$, contrasting with the exact benchmark total $\hat{X}_{d,p} = \sum_{j \in S} w_{j,d}^* x_{j,p}$. When the target population in small areas is smaller than the sample size of the survey data set used, this process can lead to a serious underestimation of population totals and size [23,36]. This limitation is relevant when the synthetic dataset needs to be used in microsimulation modules that assume integer replication of individuals (e.g., for behavioural or dynamic modelling). While specific strategies exist to optimize this integerisation process (e.g., the “truncate, replicate, sample” method by [48]), a detailed investigation of rounding techniques is beyond the scope of this work; therefore, standard rounding is applied.

Simulated annealing, on the other hand, circumvents the problem of non-integer weights by design, producing integer weights directly. Nevertheless, SA presents a notable drawback: it is extremely computationally intensive.

All methods in this study are applied at the individual level, a choice that reflects both data constraints and methodological considerations. It is worth noting that the literature provides established solutions for generating “two-layered” synthetic populations that simultaneously control for individual and household constraints [49], such as the Iterative Proportional Updating (IPF extension) proposed by Ye et al. [50], and the integrated weighting approach, as discussed in Lemaitre and Dufour [51], Tanton and Edwards [52, Chap. 9] and Haziza and Beaumont [53], to ensure consistency between individual and household weights. Similarly, SA can theoretically operate at the household level to preserve family structures. However, our choice to focus on individuals is driven by two decisive factors. First, the available administrative and census data offer a much greater availability of benchmarks for individuals compared to households. Second, all reweighting methods share a fundamental dependency on the representativeness of the source data. The “zero-cell problem” – where specific profiles absent from the source sample cannot be generated [49] – is severely exacerbated at the household level due to the high dimensionality of joint distributions. Our preliminary tests of SA at the household level confirmed this, yielding unsatisfactory convergence: given the sample size of the survey, the sparsity of specific household profiles prevented the algorithm from finding optimal matches. Consequently, we opted for an individual-based strategy; although this results in the loss of the original household structure, it allows for a significantly richer and more robust characterization of the synthetic units given the available data. For completeness, Appendix reports our household-level version of the SA algorithm.

2.3. On the choice of the performance measures

Another important issue in SM modelling is the definition of appropriate fitness measures for model evaluation. This work uses the Total Absolute Error (TAE), which is certainly the most widely used metric in literature. TAE measures the aggregate absolute difference between the known benchmark totals $X_{d,p}$ and the estimated marginal totals $\hat{X}_{d,p}$ for area d . It is defined as:

$$TAE_d = \sum_{p \in P} |\hat{X}_{d,p} - X_{d,p}|. \quad (1)$$

The Percentage Standardized Absolute Error (PSAE) values refines the TAE by scaling it to the population size of each area, thus facilitating comparisons across areas of different sizes. For the d th area, it is defined as

$$PSAE_d = \frac{TAE_d}{N_d} \times 100. \quad (2)$$

The ultimate goal of using these metrics in this work is to identify the most suitable algorithm for generating a synthetic population,

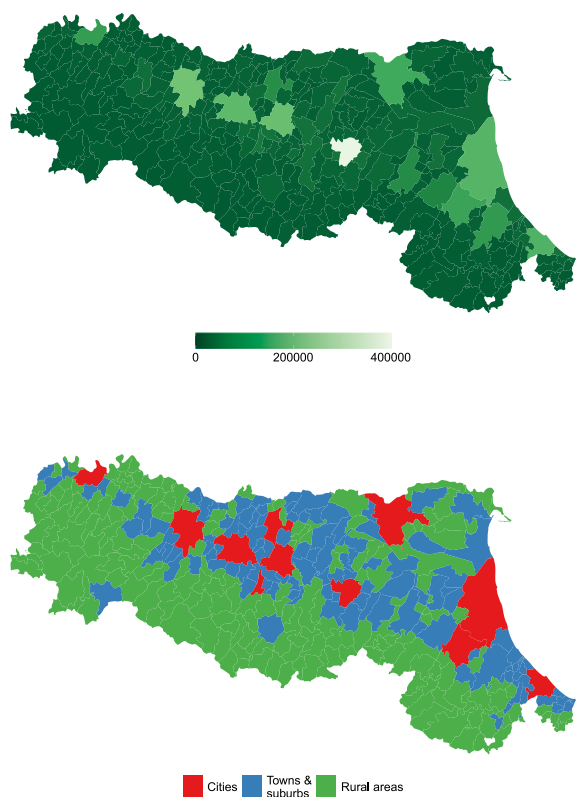


Fig. 1. Municipalities in Emilia-Romagna classified by total population size (left) and by DegUrba category (right).

conditionally on the P available constraints. It is important to note that this study applies the same vector of benchmarks across all small areas. Consequently, normalization by the dimension P is not required to ensure comparability across the spatial domain.

3. Survey data and benchmarks

To provide a rigorous empirical assessment of the reweighting methods discussed above, we apply them to the generation of a spatially disaggregated synthetic population for the 330 municipalities of Emilia-Romagna (E-R). The analysis relies on the 2937 observations available for this region in the 2022 EU-Statistics on Income and Living Conditions (EU-SILC) survey.

EU-SILC aims to produce harmonized statistics on inequality and poverty in Europe. The longitudinal data are collected using a rotational design consisting of four rotational panels, with each panel remaining in the survey for four consecutive years. Italy's EU-SILC (IT-SILC) is conducted annually by the Italian Institute of Statistics (ISTAT). It employs a two-stage stratified sampling design: in the first stage, municipalities are drawn with probability proportional to size; in the second stage, households are randomly selected from the municipal registries. The survey does not provide detailed location information and is statistically representative at the regional level.

This two-stage stratified sampling design ensures a high representation of cities and large towns, whereas the smallest municipalities are rarely included in the sample. This creates a representativeness challenge, as 40% of the municipalities in E-R have a population of less than 5000 inhabitants. The average and median municipality sizes are 13,515 and 6330, respectively; the largest municipality is Bologna (390,554 inhabitants) and the smallest is Zerba (71 inhabitants).

For the purpose of this analysis, E-R municipalities were classified according to the Eurostat Degree of Urbanisation (DegUrba), which

defines three categories, according to the population density and size: cities, towns and suburbs, and rural areas. The distribution of municipalities by population size and DegUrba category is displayed in Fig. 1. As illustrated, the vast majority of municipalities (approximately 60%) are classified as rural areas, while 36% are towns or suburbs, and only a small fraction are cities. However, the survey sample distribution is starkly different: only around 15% of observations come from rural areas, while 36% come from cities, with the largest share from towns and suburbs. This imbalance reflects the sampling design's bias towards larger urban centres, necessitating robust reweighting techniques for local estimation.

IT-SILC dataset provides cross-sectional sampling weights, which are derived via a multi-step procedure. This includes correction for non-response via logistic models, trimming of extreme weights to reduce variance, and two successive calibration steps: first on labour force survey totals (e.g., education level, work status by sex and age group), and subsequently on demographic totals from administrative sources (e.g., population by region, sex, age, and citizenship). These steps ensure representativeness at the regional level, though not necessarily at the municipal level.

To perform the spatial microsimulation, we utilized ISTAT census and administrative data to define the benchmarks (constraints) required for the reweighting process at the municipality level. Data preprocessing was performed on the IT-SILC survey data: variables were recoded to strictly match the benchmark definitions, and the micro-level dataset was transformed such that each categorical variable used for reweighting was converted into binary indicators.

To determine the optimal set of constraints from the available administrative data, we implemented a stepwise forward selection algorithm. The starting pool consisted of 147 potential attributes, ranging from basic demographics (age classes, gender, citizenship, education, employment status, household composition) to complex interactions. The selection algorithm was designed to navigate the critical trade-off between distributional accuracy and computational feasibility. On one hand, relying solely on univariate constraints (e.g., total males, total employed) ensures high convergence rates but fails to capture the correlation structure between variables (e.g., the specific relationship between age and education), leading to synthetic populations that match marginals but misrepresent joint distributions. On the other hand, introducing bivariate or multivariate constraints (interactions) significantly improves structural realism but exponentially increases the sparsity of the contingency tables. This increases the risk of the "zero-cell problem" [49]. To balance these conflicting objectives, the algorithm iteratively tests the inclusion of each candidate constraint. At each step, it selects the variable that minimizes the average TAE in (1) across all municipalities, conditional on the reweighting algorithm successfully converging for a predefined threshold of areas (e.g., allowing at most 1 failure). This procedure ensures a balance between maximizing the information content of the benchmarks and maintaining computational feasibility.

It is important to clarify that the primary objective of this study is to provide a rigorous methodological comparison of reweighting algorithms, rather than to tailor a specific policy model for each municipality. When the goal is instead to optimize the synthetic population for a specific outcome variable, we recommend performing correlation or regression-based benchmark selection, as in Chin et al. [22] and Tanton and Vidyattama [54], to ensure that the auxiliary information is relevant to the problem of interest.

To ensure strict comparability of the algorithms' performance across the spatial domains of this study, the identified vector of benchmarks is fixed and applied uniformly to all 330 municipalities, avoiding local-specific adaptations. Ultimately, we defined 6 constraints, some univariate and others based on interactions between variables, for a total of 48 attributes: age classes (6), gender (2), gender by age classes (8), education by gender and age class (15), work status by gender and age class (14), household-size group (3). Note that the latter refers to the size of the household to which the individual belongs, as household-level benchmarks are typically less available than individual-level data.

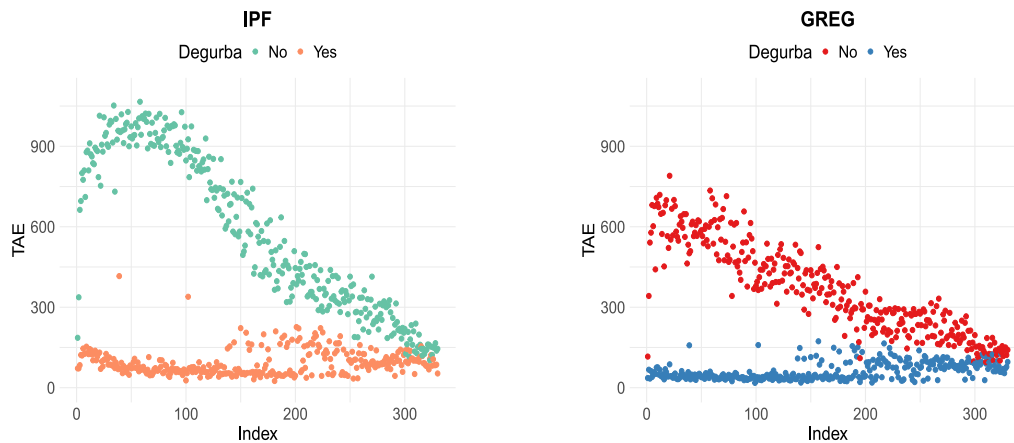


Fig. 2. Comparison between the inclusion and exclusion of the DegUrba classification in the IPF (left) and GREG (right) methods.

4. Comparing the performance of synthetic population approaches for emilia-romagna municipalities

This section presents a detailed comparison of the deterministic (IPF, GREG) and probabilistic (SA) reweighting algorithms on real data to identify the most effective strategy for population reconstruction, followed by a rigorous validation of the synthetic microdata against administrative benchmarks and external fiscal data.

4.1. Comparison of reweighting algorithms

4.1.1. Optimizing deterministic methods

The calibrated weights $w_{j,d}^*$ using IPF or GREG, computed for each area $d = 1, \dots, D$ are rounded and applied to replicate each individual survey observation according to its corresponding value, thereby generating a synthetic representation of the target population at the micro-level. We use the original design weights $w_j, \forall j \in S$, as starting values, as these yields significantly better convergence compared to initializing weights to unity.

In line with what reported by Harland et al. [46], restricting the survey data to observations from geographically or structurally similar areas can significantly improve performance. As shown in Fig. 2, the spatial filtering by the municipality Degree of Urbanisation classification (DegUrba, as explained in Section 3) leads to a substantial reduction in the TAE for both GREG and IPF. Given the limited size of our survey dataset, further spatial filtering – such as retaining only observations from municipalities of the same or bordering area – is not feasible, as it would severely reduce the donor pool, particularly for the smallest municipalities, despite the improvement in accuracy that such filtering can yield, as shown by Tanton and Vidyattama [54].

Fig. 3 compares the TAE values obtained with GREG and IPF across all municipalities. Most points lie above the bisector (red dashed line), indicating that GREG consistently outperforms IPF in terms of reconstruction accuracy. Consequently, GREG is selected as the benchmark deterministic method for the subsequent analysis and comparison.

4.1.2. The integerisation problem

While GREG ensures efficient calibration using floating-point weights, generating a synthetic population requires rounding these weights to integers, as $\lfloor w_{j,d}^* \rfloor$. This step inevitably breaks the perfect matching with benchmarks.

Fig. 4 (left panel) reports the distribution of absolute differences between the exact GREG estimates and their integerised counterparts, $\hat{X}_p - \lfloor \hat{X}_{d,p}^* \rfloor$, on average over all the D municipalities and P benchmark variables. While discrepancies are generally negligible for the region as a whole, with small discrepancies rarely exceeding ± 5 units, the right panel exposes a scale-dependency: when standardized by municipality

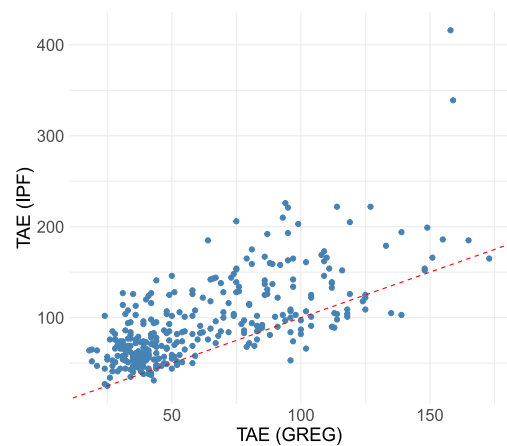


Fig. 3. TAE values obtained using GREG (x-axis) and IPF (y-axis) across municipalities. Red dashed line indicates the bisector.

size N_d , relative discrepancies become disproportionately large for the smallest municipalities. In these sparse domains, which constitute a minority of cases, calibration often yields small fractional weights, which, once rounded, result in significant distortions of the marginal totals.

These results suggest that while simple rounding is a viable and reliable strategy for the vast majority of the region, it introduces non-negligible bias in the smallest demographic domains. This limitation explicitly underscores the necessity of alternative approaches, such as Simulated Annealing, specifically for these sparse areas where preserving integer precision is paramount.

Probabilistic vs. Deterministic methods

To address the integerisation bias in small areas, we implemented the Simulated Annealing (SA) algorithm. The optimization process was initialized with a temperature $T_0 = 2000$ and a cooling factor $\alpha = 0.99$. The algorithm terminates when one of the following conditions is met: a perfect fit is achieved ($TAE = 0$), the TAE remains constant for 500 iterations, the temperature drops below a negligible threshold, or a maximum of 500,000 iterations is reached. This upper bound is deliberately set high to ensure the system has sufficient time to cool down and converge before the limit is exhausted.

Fig. 5 presents the comparative accuracy (PSAE in (2)) of SA and GREG, revealing a clear crossover in performance driven by population size N_d . The SA approach proves particularly effective in rural areas, where it achieves significantly lower PSAE values compared to GREG.

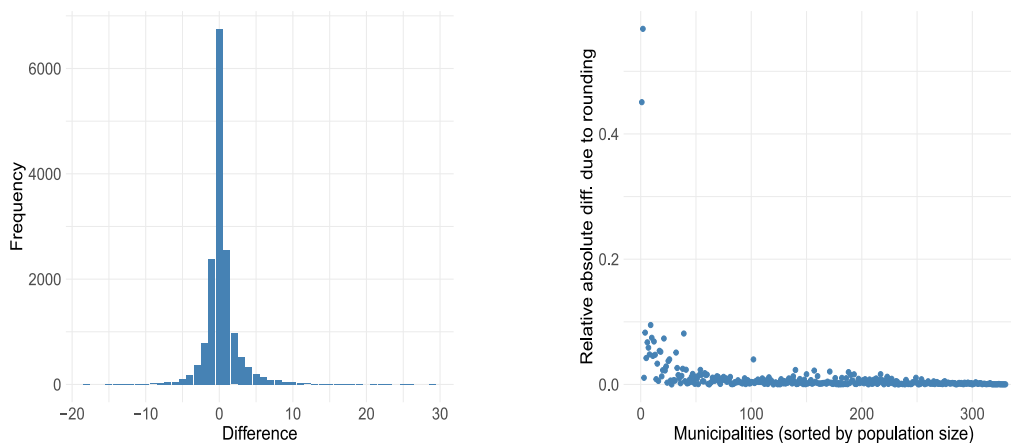


Fig. 4. Impact of rounding the calibrated weights: average differences across all P benchmark variables and D areas (left), and relative differences by municipality ordered by population size (right)

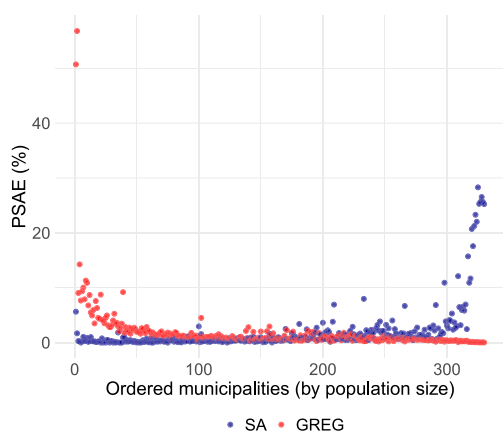


Fig. 5. Comparison of PSAE values between SA and GREG estimates across municipalities, ordered by population size.

The median PSAE in rural municipalities drops from 1.06 (GREG) to 0.22 (SA). Notably, SA achieves a perfect fit ($TAE = 0$) in 16 rural municipalities. This highlights SA's strength in handling small, sparsely populated areas, where GREG suffers from rounding issues and weight instability. Crucially, for these small municipalities, the SA algorithm is also computationally efficient. Since the search space is smaller, convergence is reached relatively quickly, making it a practical option for simulations at the local level.

However, this advantage reverses as population size increases. In cities, GREG outperforms SA substantially (Mean PSAE 0.06 vs 4.87). In these well-sampled urban contexts, the survey weights w_j used by GREG are robust and informative, whereas the stochastic search of SA struggles to converge to a global optimum within a reasonable time frame. As population size increases, the combinatorial search space expands exponentially. To address the high computational burden of SA, the custom R function has been parallelized to take advantage of all available CPU cores. Nevertheless, the entire simulation for the 330 municipalities in Emilia-Romagna required approximately 7 days of continuous computation, a burden driven primarily by the larger urban centres. Consequently, applying this individual-level swapping approach to even larger aggregations, such as regions (NUTS 2) or provinces (NUTS 3), would be prohibitively time-consuming.

Given this trade-off, a hybrid strategy that applies SA to small municipalities (in our studies, rural areas) and GREG to larger ones may offer an optimal balance between flexibility, precision, and computational feasibility.

4.1.3. Distributional validation

So far, performance has been evaluated using the TAE and PSAE metrics, see (1) and (2), which assess aggregate errors across target variables. However, since the ultimate goal is to reconstruct micro-level data for policy evaluation – beyond merely estimating totals or means – further validation is called for to ensure that the synthetic population accurately reflects the underlying variables distributions and structural relationships of the real data.

As an illustrative case, we consider the variable age, for which age classes were included among the benchmark constraints. Fig. 6 compares three density distributions for the municipalities of Bologna and Zerba (respectively, the largest and smallest cities in the region): the expanded observed ages (grey curve), derived directly from the survey data replicated according to the rounded GREG calibrated weights; the population register (light-grey curve); the simulated from age classes (red curve), obtained by randomly assigning an age within each age-class interval to individuals in the synthetic population. The age distribution resulting from the reweighted survey data exhibits significant irregularities for Bologna, with exaggerated peaks at specific age values. This is an artifact of the reweighting process: when a survey respondent is assigned a high calibrated weight, they are replicated multiple times in the synthetic population, creating a cluster of identical values. By contrast, the random draw strategy produces a smoother distribution that more closely resembles the true population structure. This phenomenon is particularly pronounced in such large urban centres, where the ratio between the population size and the sample size is high, resulting in larger weights and thus more massive replication of individual records. Conversely, in smaller municipalities, such as Zerba, calibrated weights tend to be lower and closer to unity. Consequently, the artificial peaks caused by replication are naturally less pronounced, and the direct synthetic distribution tends to align more organically with the observed one.

4.1.4. External validation: taxable income

Validation can also be performed using variables that were not included in the benchmarking constraints. In this section, we explore the results in terms of *individual overall taxable income*. Income is undoubtedly one of the most relevant variables for socio-economic policy evaluation, as well as for monitoring economic poverty and inequality. As a reference population source, we use data provided by the Italian Ministry of Economy and Finance (MEF), which include administrative records on annual taxable income for individuals under the Italian tax system. The MEF data, publicly available in aggregated form at the municipality level, include overall taxable income from entrepreneurial activity, employment, self-employment, shareholdings, pensions, and property.

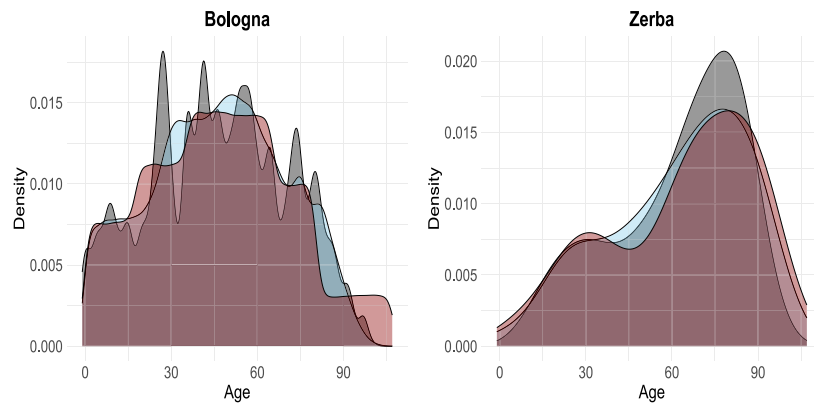


Fig. 6. Age distribution in Bologna (population $\approx 400,000$) and Zerba (population ≈ 100) comparing the population register (light-grey), the GREG-weighted survey (grey), and simulated ages drawn from class intervals (red)

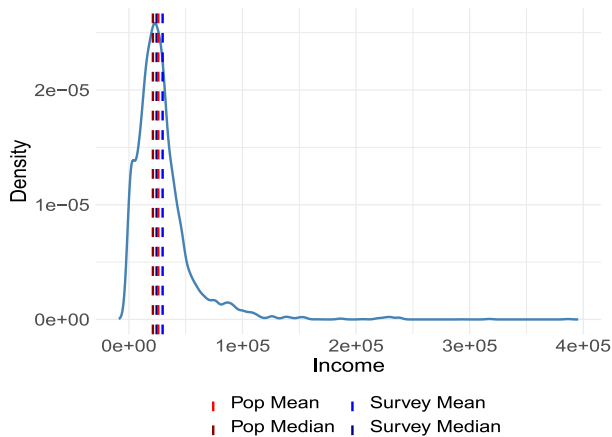


Fig. 7. Density plot of Emilia-Romagna income with vertical lines for population mean, population median, survey mean and survey median.

The overall taxable income variable derived from EU-SILC serves as a proxy for the administrative taxable income available in the MEF dataset. While conceptually similar, discrepancies are expected due to differences in definitions (e.g., tax evasion, non-taxable income components included in EU-SILC) and measurement methods (self-reported vs. administrative records). Consequently, a perfect alignment is not theoretically expected; however, the comparison remains valid for assessing the distributional consistency of the synthetic population. Fig. 7 shows the income distribution derived from EU-SILC data for Emilia-Romagna (using sample weights). Vertical lines in the figure indicate the weighted median and mean values estimated from the survey, as well as those computed on the MEF data for the same region. We observe quite good similarity overall.

Next, we compare the estimated mean taxable income on the synthetic populations generated for each municipality with the corresponding values computed from the MEF data. Fig. 8 (left panel) displays the results when the GREG is performed with the selected benchmarks using the greedy algorithm explained in Section 3 (starting GREG). We observe that the estimates under GREG are substantially more concentrated than the MEF values, severely underestimating the actual variability across municipalities. This shrinkage effect is typical when auxiliary variables are not strongly correlated with the target variable.

To address this, we reinforce the role of income information within the benchmark variables. In the right plot of Fig. 8, we illustrate an example in which the same estimation procedure using GREG based on EU-SILC microdata is applied, but the calibration benchmarks are augmented to include the known frequencies across eight income classes

derived from MEF data (*augmented GREG*). In this case, the distributions of the estimated means are much closer to those based on the fiscal data, highlighting the importance of carefully selecting benchmarks when the research focus is on specific indicators.

Similarly to the procedure applied for age, one can calibrate the sampling weights to income-related benchmarks and then randomly assign continuous income values within each class interval. This helps mitigate discretization artefacts (“jumpy” distributions) and produces smoother, more realistic micro-level estimates, in line with the approach adopted by Alfons et al. [24], Templ et al. [28].

5. Discussion

The availability of a synthetic population for a target area, whether a region composed of multiple local areas or a city and its suburbs, provides a valuable foundation for *ex ante* and *ex post* small area policy evaluation through spatial microsimulation tools. Survey samples are inevitably constrained by budget limitations, which allow representativeness only at broad territorial levels (e.g. regions). By contrast, policymakers require disaggregated evaluations at the finest possible geographic scale to effectively monitor policy implementation and assess socio-economic phenomena.

This work demonstrates that spatial microsimulation, specifically through reweighting techniques, provides a robust solution to reconstructing these local populations even in a data-scarce context. These methods leverage survey microdata and known totals from census or administrative sources to generate synthetic populations at the desired level of aggregation, even in the absence of sample observations. An additional advantage is that, since synthetic microdata do not contain private information that leads to identification, they are free of confidentiality and disclosure concerns. Nevertheless, it is important to acknowledge that the synthetic population is based on specific assumptions and depends heavily on the chosen generation model.

Our analysis identifies the Generalized Regression Estimator (GREG) as the most balanced approach for real-data applications, striking an optimal trade-off between computational efficiency and distributional precision. GREG outperforms Iterative Proportional Fitting (IPF) in accuracy and ease of implementation. Furthermore, for the direct estimation of indicators (e.g., mean income) without integerisation, GREG with continuous weights remains the gold standard.

However, there is a crucial exception. In sparse domains, specifically very small municipalities, the integerisation bias of deterministic methods becomes significant. In this context, Simulated Annealing (SA) proves superior, achieving near-perfect reconstruction by design (e.g., $TAE \approx 0$). Despite its computational intensity and its stochastic nature – which introduces variability across runs [55] – SA is consistently shown to outperform deterministic reweighting methods [30,46]. However, its computational burden makes it impractical for

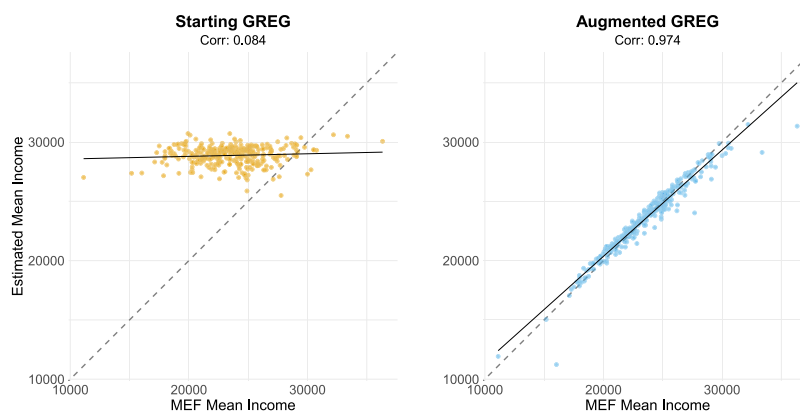


Fig. 8. Comparison of estimates of mean taxable income distributions across municipalities in Emilia-Romagna with starting and augmented GREG.

medium and large municipalities, where GREG is the best viable option. Consequently, our application to the Emilia-Romagna municipalities advocates for a hybrid strategy: deploying SA for small municipalities (rural areas, in our study) to maximize precision, and GREG for medium and large areas to ensure computational tractability.

Methodologically, the quality of the synthetic population is strictly conditional on the choice of benchmarks. Our stepwise forward selection algorithm – which at each step selects the benchmark that most reduces the average TAE across municipalities, subject to a convergence constraint – was designed to ensure global convergence across all the studied areas while maximizing the number of utilized benchmarks, rather than to optimize for a specific policy outcome. When the research focus is a specific indicator, benchmarks should instead be chosen based on their correlation with the target variable. As demonstrated with the income variable in the augmented GREG example, explicitly including relevant benchmarks is essential when the focus shifts to specific outcomes. Moreover, when the interest lies in specific areas, the trade-off between representativeness and convergence can be locally optimized. Although this work did not aim to isolate the specific contribution of each benchmark variable or to optimize performance for individual municipalities, the proposed method provides a general framework for constructing subnational synthetic populations. Researchers retain the flexibility to define benchmarks *ex ante* based on the specific indicators they intend to evaluate.

We highlight that people living in non-private dwellings that are excluded from the EU-SILC survey, such as individuals residing in collective housing or institutions, are not considered in our analysis. Nevertheless, we believe that synthetic reconstruction methods may assist in creating microdata for these populations, for which survey data are lacking, when auxiliary information is available; however, such data were not accessible to us.

Finally, the dependency on the input sample remains the ultimate constraint. The “zero-cell problem” highlights the importance of sample size and diversity. In this study, E-R EU-SILC data were filtered according to the Degree of Urbanisation (DegUrba) of each target municipality. Future research could mitigate this by expanding the donor pool (e.g. using the entire EU-SILC data or neighbouring or “similar” regions to Emilia-Romagna), employing alternative spatial filtering strategies, or integrating synthetic reconstruction techniques to impute missing profiles. These data constraints also limit the applicability of approaches that simultaneously account for both individual- and household-level information, such as integrated weighting. While not feasible in the present framework, such methods represent a valuable extension in contexts where richer household-level auxiliary data are available.

This work provides not just a synthetic population of the Emilia-Romagna region composed of its municipalities, but a scalable, validated framework for spatial microsimulation. By navigating the trade-offs between method accuracy and computational cost, the described

approach and its multiple trade-offs offer a reproducible instrument for generating synthetic data to support evidence-based local policy making.

CRediT authorship contribution statement

Silvia Scarpa: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Massimo Baldini:** Conceptualization, Funding acquisition, Project administration. **Andrea Barigazzi:** Writing – review & editing, Writing – original draft, Conceptualization. **Giovanni Gallo:** Conceptualization, Project administration. **Marcello Morciano:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Investigation, Conceptualization.

Declaration of generative AI use

During the preparation of this work the authors used AI, such as ChatGPT, in the writing process to improve the readability and language of the manuscript. After using this tool, the author reviewed and edited the content as needed and take full responsibility for the content of the published article.

Funding sources

Funder Silvia Scarpa: PNRR - Missione 4 Componente 2 Investimento 1.2 - D.D. n. 47/2025 PNRR - M4C2 INV.1.2, CUP: E93C25000610006;

Declaration of competing interest

The usual caveat applies: the views expressed here are those of the authors and do not reflect those of the institutions with which the authors are affiliated. The authors have no competing interests to declare that are relevant to the content of this article.

Acknowledgements

We are grateful to the wider team of the project “Amartya: Costruzione popolazione sintetica e implementazione modello di microsimulazione tax-benefit” for their suggestions throughout the project and to participants at the Bank of Italy conference for their valuable comments. We also thank the Office of Statistics of the Emilia-Romagna Region, member of the Italian National Statistical System (SISTAN), for granting access to the advanced version of the Italian EU-SILC database within the project. The usual caveats apply.

Appendix. Household-level Simulated Annealing

This appendix provides the household-level specification of the SA algorithm discussed in Section 2.

We did not adopt the household-level version of SA because the available data are not suitable for a robust implementation at that level. In our setting, the EU-SILC sample contains too few households observations and only a limited set of household-level benchmarks is available from administrative sources. These two factors exacerbate the “zero-cell problem” and prevent satisfactory convergence of the algorithm.

We outline below the proposed household-level SA version that we tested during the methodological assessment. The algorithm proceeds as follows:

1. Set the initial temperature T_0 , the cooling factor $\alpha \in (0, 1)$ and the maximum number of iterations Max_{iter} ;
2. Draw an initial sample $s^{(0)}$ of households from the survey data S , according to the known distribution of household sizes in the target area d . Specifically, for each household size category, select a number of households equal to the known total. Let m_h denote the number of individuals in household h ; the total population constraint is then enforced as $\sum_{h \in s} m_h = N_d$, where N_d is the known population size of area d ;
3. Compute the initial TAE, denoted TAE_0 , between the marginal totals of the synthetic population and the known benchmarks.
4. Repeat the following steps until a stopping condition is met (i.e., maximum number of iterations reached, $TAE_k = 0$, or the temperature falls below a small threshold ε):
 - (a) Propose at step k a new sample $s^{(k)}$ by replacing one (or more) household from the current sample with another drawn from the survey data. The replacement household(s) is required to have the same size (i.e., number of individuals) to preserve the population size constraint;
 - (b) Compute the new TAE, denoted TAE_k , based on the updated sample;
 - (c) Calculate $\Delta_k = TAE_k - TAE_{k-1}$. Accept the new sample if $\Delta_k < 0$. Otherwise, accept the sample with probability $P_{accept} = \exp\left(-\frac{\Delta_k}{T_k}\right)$. Draw $Z \sim \text{Bernoulli}(P_{accept})$, and accept the new solution if $Z = 1$;
 - (d) Update the temperature using the cooling factor as $T_{k+1} = \alpha \cdot T_k$.

Data availability

The municipal-level population totals are publicly available. The advanced version of the Italian component of the EU-SILC database used in this study are not publicly available. These data were provided to the authors within the project “Amartya: Costruzione popolazione sintetica e implementazione modello di microsimulazione tax-benefit” from the Office of Statistics of the Emilia-Romagna Region, as a member of the Italian National Statistic System (SISTAN). Access to the standard EU-SILC microdata can be requested from Eurostat.

References

- [1] Fotheringham AS, Sachdeva M. On the importance of thinking locally for statistics and society. *Spat Stat* 2022;50:100601. <http://dx.doi.org/10.1016/j.spasta.2022.100601>.
- [2] Benassi F, Naccarato A, Vellucci P. Exploring the territorial unevenness of one-person households and contextual factors of vulnerability: Evidence from the Italian context. *Socio-Economic Plan Sci* 2024;95:102014. <http://dx.doi.org/10.1016/j.seps.2024.102014>.

- [3] Connelly R, Playford CJ, Gayle V, Dibben C. The role of administrative data in the big data revolution in social science research. *Soc Sci Res* 2016;59:1–12. <http://dx.doi.org/10.1016/j.ssresearch.2016.04.015>.
- [4] Snoke J, Raab GM, Nowok B, Dibben C, Slavkovic A. General and specific utility measures for synthetic data. *J R Stat Soc Ser A: Stat Soc* 2018;181(3):663–88. <http://dx.doi.org/10.1111/rssa.12358>.
- [5] Jenkins SP, Burkhauser RV, Feng S, Larrimore J. Measuring inequality using censored data: a multiple-imputation approach to estimation and inference. *J R Stat Soc Ser A: Stat Soc* 2011;174(1):63–81. <http://dx.doi.org/10.1111/j.1467-985X.2010.00655.x>.
- [6] Tourangeau R. *Hard-to-survey populations*. Cambridge University Press; 2014. <http://dx.doi.org/10.1017/CBO9781139381635>.
- [7] D'Alessio G, Faiella I. *Non-response behaviour in the Bank of Italy's Survey of Household Income and Wealth*. Technical Report, Bank of Italy, Economic Research and International Relations Area; 2002.
- [8] Neri A, Ranalli MG. To misreport or not to report? The measurement of household financial wealth. 2012. <http://dx.doi.org/10.2139/ssrn.2154922>, The Measurement of Household Financial Wealth (July 26, 2012). Bank of Italy Temi di Discussione (Working Paper) No 870 (2012).
- [9] Cantarella M, Neri A, Ranalli MG. Estimating the distribution of household wealth: methods for adjusting survey data estimates using national accounts and rich list data. *Rev Income Wealth* 2024;70(3):551–80. <http://dx.doi.org/10.1111/roiw.12657>.
- [10] Pinkovskiy M, Sala-i Martin X, Chatterji-Len K, Nober WH. Inequality Within Countries is Falling: Underreporting-Robust Estimates of World Poverty, Inequality and the Global Distribution of Income. Cambridge, MA: National Bureau of Economic Research; 2024. <http://dx.doi.org/10.3386/w32203>, NBER Working Paper 32203.
- [11] Piketty T, Saez E, Zucman G. Distributional national accounts: Methods and estimates for the united states. *Q J Econ* 2018;133(2):553–609. <http://dx.doi.org/10.1093/qje/qjx043>.
- [12] Haltiwanger J, Hyatt HR, Spletzer JR. Increasing earnings inequality: Reconciling evidence from survey and administrative data. *J Labor Econ* 2023;41(S1):S61–109. <http://dx.doi.org/10.1086/726632>.
- [13] Grossman I, Wilson T, Temple J. Forecasting small area populations with long short-term memory networks. *Socio-Economic Plan Sci* 2023;88:101658. <http://dx.doi.org/10.1016/j.seps.2023.101658>.
- [14] Pfeiffermann D. New important developments in small area estimation. *Statist Sci* 2013;28:40–68. <http://dx.doi.org/10.1214/12-STS395>.
- [15] Rao JN, Molina I. *Small Area Estimation*. John Wiley & Sons; 2015. <http://dx.doi.org/10.1002/9781118735855>.
- [16] Morales D, Esteban MD, Pérez A, Hobza T. A course on small area estimation and mixed models. *Methods, Theory Appl R Stat Soc Behav Sci* 2021. <http://dx.doi.org/10.1007/978-3-030-63757-6>.
- [17] Tonutti G, Bertarelli G, Giusti C, Pratesi M. Disaggregation of poverty indicators by small area methods for assessing the targeting of the “reddito di cittadinanza” national policy in Italy. *Socio-Economic Plan Sci* 2022;82:101327. <http://dx.doi.org/10.1016/j.seps.2022.101327>.
- [18] Dang H-AH, Lanjouw PF. Regression-based imputation for poverty measurement in data-scarce settings. In: *Research handbook on measuring poverty and deprivation*. Edward Elgar Publishing; 2023, p. 141–50. <http://dx.doi.org/10.4337/9781800883451.00023>.
- [19] Dang H-AH, Kilic T, Abanokova K, Carletto C. Poverty imputation in contexts without consumption data: A revisit with further refinements. *Rev Income Wealth* 2025;71(1):e12714. <http://dx.doi.org/10.1111/roiw.12714>.
- [20] Betti G, Molini V, Mori L. An attempt to correct the underestimation of inequality measures in cross-survey imputation through generalized additive models for location, scale and shape. *Socio-Economic Plan Sci* 2024;91:101784. <http://dx.doi.org/10.1016/j.seps.2023.101784>.
- [21] Harding A, Lloyd R, King A. Assessing poverty and inequality at a detailed regional level: New advances in spatial microsimulation. In: McGillivray M, Clarke M, editors. *Understanding human well-being*. Helsinki: United Nations University Press; 2006, p. 239–61.
- [22] Chin S-F, Harding A, Lloyd R, McNamara J, Phillips B, Vu QN. Spatial microsimulation using synthetic small-area estimates of income, tax and social security benefits. *Australas J Reg Stud* the 2005;11(3):303–35, URL <https://search.informit.org/doi/10.3316/informit.058402399727354>.
- [23] O'Donoghue C, Morrissey K, Lennon J. Spatial microsimulation modelling: a review of applications and methodological choices. *Int J Microsimulation* 2014. <http://dx.doi.org/10.34196/ijm.00093>.
- [24] Alfons A, Kraft S, Templ M, Filzmoser P. Simulation of close-to-reality population data for household surveys with application to EU-SILC. *Stat Methods Appl* 2011;20(3):383–407. <http://dx.doi.org/10.1007/s10260-011-0163-2>.
- [25] Gussenbauer J, Templ M, Fritzmann S, Kowarik A. Simulation of calibrated complex synthetic population data with XGBoost. *Algorithms* 2024;17(6):249. <http://dx.doi.org/10.3390/a17060249>.
- [26] Hermes K, Poulsen M. A review of current methods to generate synthetic spatial microdata using reweighting and future directions. *Comput Environ Urban Syst* 2012;36(4):281–90. <http://dx.doi.org/10.1016/j.compenvurbsys.2012.03.005>.

- [27] Tanton R. A review of spatial microsimulation methods. *Int J Microsimulation* 2014;7(1):4–25. <http://dx.doi.org/10.34196/ijm.00092>.
- [28] Tempel M, Meindl B, Kowarik A, Dupriez O. Simulation of synthetic complex data: The R package simpop. *J Stat Softw* 2017;79(10):1–38. <http://dx.doi.org/10.18637/jss.v079.i10>.
- [29] Pfeiffermann D. The use of sampling weights for survey data analysis. *Stat Methods Med Res* 1996;5(3):239–61. <http://dx.doi.org/10.1177/096228029600500303>.
- [30] Huang Z, Williamson P. A comparison of synthetic reconstruction and combinatorial optimisation approaches to the creation of small-area microdata. *Dep Geogr Univ Liverpool* 2001.
- [31] Tillé Y. Sampling and estimation from finite populations. John Wiley & Sons; 2020. <http://dx.doi.org/10.1002/9781119071259>.
- [32] Lenormand M, Deffuant G. Generating a synthetic population of individuals in households: Sample-free vs sample-based methods. 2012. <http://dx.doi.org/10.18564/jasss.2319>, arXiv preprint arXiv:1208.6403.
- [33] Birkin M, Clarke M. Synthesis — a synthetic spatial information system for urban and regional analysis: methods and examples. *Environ Plan A* 1988;20(12):1645–71. <http://dx.doi.org/10.1068/a201645>.
- [34] Sun L, Erath A. A Bayesian network approach for population synthesis. *Transp Res Part C: Emerg Technol* 2015;61:49–62. <http://dx.doi.org/10.1016/j.trc.2015.10.010>.
- [35] Saadi I, Mustafa A, Teller J, Farooq B, Cools M. Hidden Markov model-based population synthesis. *Transp Res Part B: Methodol* 2016;90:1–21. <http://dx.doi.org/10.1016/j.trb.2016.04.007>.
- [36] Williamson P, Birkin M, Rees PH. The estimation of population microdata by using data from small area statistics and samples of anonymised records. *Environ Plan A* 1998;30(5):785–816. <http://dx.doi.org/10.1068/a300785>.
- [37] Deming WE, Stephan FF. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann Math Stat* 1940;11(4):427–44. <http://dx.doi.org/10.1214/aoms/1177731829>.
- [38] Kolenikov S. Calibrating survey data using iterative proportional fitting (raking). *Stata J* 2014;14(1):22–59. <http://dx.doi.org/10.1177/1536867X1401400104>.
- [39] Hidalgo MEM, Vidyattama Y, Tanton R. A Comparison of the GREGWT and IPF Methods for the Re-weighting of Surveys. In: 5th world congress of the international microsimulation association. IMA, 2015.
- [40] Ballas D, Clarke G, Dorling D, Eyre H, Thomas B, Rossiter D. SimBritain: a spatial microsimulation approach to population dynamics. *Popul Space Place* 2005;11(1):13–34. <http://dx.doi.org/10.1002/psp.351>.
- [41] Deville J-C, Särndal C-E. Calibration estimators in survey sampling. *J Amer Statist Assoc* 1992;87(418):376–82. <http://dx.doi.org/10.1080/01621459.1992.10475217>.
- [42] Tanton R, Vidyattama Y, Nepal B, McNamara J. Small area estimation using a reweighting algorithm. *J R Stat Soc Ser A: Stat Soc* 2011;174(4):931–51. <http://dx.doi.org/10.1111/j.1467-985X.2011.00690.x>.
- [43] Creedy J. Survey reweighting for tax microsimulation modelling. In: *Studies on economic well-being: essays in the honor of John P. Formby*. Emerald Group Publishing Limited; 2004, p. 229–49. [http://dx.doi.org/10.1016/S1049-2585\(04\)12009-7](http://dx.doi.org/10.1016/S1049-2585(04)12009-7).
- [44] Pacifico D. Reweight: a Stata module to reweight survey data to external totals. Technical Report, Government of the Italian Republic (Italy), Ministry of Economy and Finance, Department of the Treasury Working Paper; 2014.
- [45] Tanton R, Williamson P, Harding A. Comparing two methods of reweighting a survey file to small area data. *Int J Microsimulation* 2014;7(1):76–99. <http://dx.doi.org/10.34196/ijm.00094>.
- [46] Harland K, Heppenstall A, Smith D, Birkin MH. Creating realistic synthetic populations at varying spatial scales: A comparative critique of population synthesis techniques. *J Artif Soc Soc Simul* 2012;15(1). <http://dx.doi.org/10.18564/jasss.1909>.
- [47] Morrissey K, O'Donoghue C, Clarke G, Li J. Using simulated data to examine the determinants of acute hospital demand at the small area level. *Geogr Anal* 2013;45(1):49–76. <http://dx.doi.org/10.1111/gean.12000>.
- [48] Lovelace R, Ballas D. 'Truncate, replicate, sample': A method for creating integer weights for spatial microsimulation. *Comput Environ Urban Syst* 2013;41:1–11. <http://dx.doi.org/10.1016/j.compenvurbsys.2013.03.004>.
- [49] Yaméogo BF, Gastineau P, Hankach P, Vandanjon P-O. Comparing methods for generating a two-layered synthetic population. *Transp Res Rec* 2021;2675(1):136–47. <http://dx.doi.org/10.1177/0361198120964734>.
- [50] Ye X, Konduri K, Pendyala RM, Sana B, Waddell P. A methodology to match distributions of both household and person attributes in the generation of synthetic populations. In: 88th annual meeting of the transportation research board, Washington, DC. vol. 36, 2009.
- [51] Lemaître G, Dufour J. An integrated method for weighting persons and families. *Surv Methodol* 1987;13(2):199–207.
- [52] Tanton R, Edwards K. In: *Spatial microsimulation: A reference guide for users*, vol. 6, Springer Science & Business Media; 2012. <http://dx.doi.org/10.1007/978-94-007-4623-7>.
- [53] Haziza D, Beaumont J-F. Construction of weights in surveys: A review. *Statist Sci* 2017;32. <http://dx.doi.org/10.1214/16-STS608>.
- [54] Tanton R, Vidyattama Y. Pushing it to the edge: Extending generalised regression as a spatial microsimulation method. *Int J Microsimulation* 2010;3(2):23–33. <http://dx.doi.org/10.34196/ijm.00036>.
- [55] Voas D, Williamson P. An evaluation of the combinatorial optimisation approach to the creation of synthetic microdata. *Int J Popul Geogr* 2000;6(5):349–66. [http://dx.doi.org/10.1002/1099-1220\(200009/10\)6:5<349::AID-IJPG196>3.0.CO;2-5](http://dx.doi.org/10.1002/1099-1220(200009/10)6:5<349::AID-IJPG196>3.0.CO;2-5).

Silvia Scarpa is a Post-Doc researcher at the Department of Economics Marco Biagi of the University of Modena and Reggio Emilia. She is a member of CAPP, the Center for the Analysis of Public Policies. She achieved a Ph.D. in Statistics at the Department of Statistical Sciences, University of Padua (Padua, Italy) and was a visiting scholar at Geneva University. Her research interests include applied statistics, small area estimation, socio-economic indicators and sampling theory.

Massimo Baldini is Full Professor of Economic Policy at the University of Modena and Reggio Emilia. He holds an M.Sc. in Economics from University College London and a Ph.D. in Economics from the University of Bologna. His research focuses on public policy evaluation, poverty, inequality, redistribution, and tax-benefit microsimulation. His research focuses on public policy evaluation, poverty, inequality, redistribution, and tax-benefit microsimulation. He has contributed to numerous national and international research projects and serves on several scientific committees. He is the author of books and many peer-reviewed articles and actively engages in public debate.

Andrea Barigazzi is a Data Officer at the Italian Ministry of Economy and Finance and a member of CAPP, the Center for the Analysis of Public Policies of the University of Modena and Reggio Emilia. He previously worked as a Post-Doctoral Research Fellow at the Department of Economics of the University of Modena and Reggio Emilia, where he also completed his Ph.D. His research interests include public policy evaluation, poverty and inequality, social mobility, and demographic dynamics. He was a visiting Ph.D. student at University College Dublin.

Giovanni Gallo is Associate Professor of Public Economics at the Department of Economics Marco Biagi of the University of Modena and Reggio Emilia. He is a Global Labour Organization fellow. He previously held research positions at Sapienza University of Rome and the National Institute for Public Policy Analysis (INAPP). He earned his Ph.D. in 2018 and was a visiting scholar at Tilburg University and the Luxembourg Institute of Socio-Economic Research. His research focuses on social policy evaluation, welfare systems, poverty, and inequality. He has edited four books and published in various national and international journals.

Marcello Morciano is Associate Professor of Public Finance at the Department of Economics Marco Biagi of the University of Modena and Reggio Emilia. He serves as Director of CAPP, the Center for the Analysis of Public Policies, and holds international appointments as Honorary Senior Lecturer in Health Policy and Economics at the University of Manchester (UK) and Visiting Research Associate at the Care Policy and Evaluation Centre of the London School of Economics (UK). His research focuses on the ex-ante and ex-post evaluation of public policies, with articles published in leading journals including *European Economic Review*, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *Economic Inquiry*, *Review of Income and Wealth*, *Fiscal Studies*, *Health Economics*, *Health Policy*, *Social Science & Medicine*, *European Journal of Health Economics*, and *Journal of Pension Economics and Finance*.