Manchester Metropolitan University

Moretti, A and Shlomo, N and Sakshaug, JW (2019)Small Area Estimation of Latent Economic Well-being. Sociological Methods and Research. ISSN 0049-1241

Small Area Estimation of Latent Economic Wellbeing

Angelo Moretti[1], Natalie Shlomo[1] and Joseph Sakshaug[2]

[1]*Social Statistics Department, School of Social Sciences, University of Manchester, United Kingdom*
[2]*German Institute for Employment Research, Nuremberg, Germany.*

Summary. Small area estimation (SAE) plays a crucial role in the social sciences due to the growing need for reliable and accurate estimates for small domains. In the study of wellbeing, for example, policy-makers need detailed information about the geographical distribution of a range of social indicators. We investigate data dimensionality reduction using factor analysis models and implement SAE on the factor scores under the empirical best linear unbiased prediction approach. We contrast this approach with the standard approach of providing a dashboard of indicators, or a weighted average of indicators at the local level. We demonstrate the approach in a simulation study and a real data application based on the European Union Statistics for Income and Living Conditions (EU-SILC) for the municipalities of Tuscany.

Keywords: Composite estimation; Direct estimation; EBLUP; Factor analysis; Factor scores; Model-based estimation.

1. Introduction

Measuring poverty and wellbeing is a key issue for policy makers requiring a detailed understanding of the geographical distribution of social indicators. This understanding is essential for the formulation of targeted policies that address the needs of people in specific geographical locations. Most large-scale social surveys can only provide reliable estimates at a national level. A relevant survey for analyzing wellbeing in the European Union (EU) is the European Union Statistics for Income and Living Conditions (EU-SILC). However, these data can only be used to produce reliable direct estimates at the NUTS (Nomenclature of Territorial Units for Statistics) 2 level (Giusti, Masserini and Pratesi, 2015) which are generally large regions within a country. For example, in Italy one such NUTS 2 region is Tuscany. Hence, if the goal is to measure poverty and wellbeing indicators at a sub-regional level, such as NUTS 3 or LAU (Local Administrative Units) 2 which correspond to the Italian municipalities, the indicators may not be directly estimated from EU-SILC. In fact, the domains corresponding to the regions under NUTS 2 are so-called *unplanned domains* where domain membership is not incorporated in the sampling design, and therefore the sample size in each domain is random (and may be large or small) and in many cases zero. In this case, indirect model-based estimation methods, in particular small area estimation approaches, can be used to predict target parameters for the small domains.

Small area estimation (SAE) is defined as a set of statistical procedures with the goal of producing efficient and precise estimates for small areas, as well as for domains with zero sample size (Rao and Molina, 2015). An area is defined as *small*, if the area is an unplanned domain and the specific sample size may not be large enough to provide reliable direct design-based estimates. Small areas can also be defined by the cross-classification of geographical areas by social, economic or demographic characteristics.

SAE methods can be classified into two approaches: the unit-level and the area-level approach. The unit-level approach is used when covariates are available for each unit of the population, for example from census or administrative data, while the area-level approach is used when covariate information is known only at the area level. The use of the error-components model by Battese, Harter and Fuller (1988), also known as the Battese, Harter and Fuller (BHF) model, is commonly used for the unit-level SAE approach. In the SAE literature, estimation methods include empirical best linear unbiased prediction (EBLUP), empirical Bayes (EB), and hierarchical Bayes (HB). The EBLUP method can be used under linear mixed models, while the EB and HB methods can be used under generalized linear mixed models. For a review of these methodologies and their extensions we refer to Rao and Molina (2015).

A second important issue we consider in this paper is the multidimensionality of wellbeing indicators. Although it is generally agreed that wellbeing is a multidimensional phenomenon (OECD, 2013), there is continuing debate about the suitability of combining social indicators based on taking their average or using a dashboard of single indicators. On the one hand, Ravallion (2011) argues that a single multidimensional composite indicator in the context of poverty measurement leads to a loss of information, and on the other hand, Yalonetsky (2012) points out that composite indicators are necessary when the aim is to measure multiple deprivations within the same unit (individual or household). For a theoretical review of statistical properties of multidimensional indicators obtained by multivariate statistical techniques and related problems we refer to Krishnakumar and Nagar (2008) and Bartholomew et al. (2008).

Taking this latter view, an approach to reducing data dimensionality is to consider the multidimensional phenomena as a latent variable construct measurable by a set of observed variables and estimated using a factor analysis model. Factor scores are estimated from a factor analysis model and are defined as a *composite variable* computed from more than one response variable. Indeed, factor scores provide details on each unit's placement on the factor. When we have a substantive framework where a set of variables explains a latent construct, the confirmatory factor analysis modeling approach can be used. In the context of wellbeing measurement, a framework of indicators is generally provided a priori by official statistics or international organizations and thus are treated as fixed. The vector of unobserved variables represents a set of variables that jointly describe the underlying phenomenon. We note other work on the use of factor analysis modeling in latent wellbeing measurement to reduce data dimensionality in Ferro Luzzi, Fluckiger, and Weber (2008) and Gasparini et al. (2011). There are also other approaches which reduce dimensionality of measurement frameworks, such as the Fuzzy set approach in Lemmi and Betti (2006). Betti, Gagliardi, and Verma (2017) and Betti and Gagliardi (2017) discuss the variance estimation problem of multidimensional measures of poverty and deprivation obtained via the Fuzzy set approach using the jackknife method.

Once factor scores are estimated from the factor analysis model, they can be used to conduct further statistical analysis. For instance, they can be used as dependent or independent variables of a regression or predictive analyses to answer particular research questions. Kawashima and Shiomi (2007) use factor scores in order to conduct an ANOVA analysis on high school students' attitudes towards critical thinking and tested differences by grade level and gender. In addition, Bell, McCallum, and Cox (2003) investigated reading and writing skills where they extracted the factors and estimated factor scores before using them in a multiple regression analysis model. Skrondal and Laake (2001) note that using factor scores as dependent variables in regression modelling produces consistent estimates of model parameters since any measurement error from the factor analysis model is absorbed into the prediction error and coefficients are not attenuated (see also Fuller, 1987). Also, as highlighted in Kaplan (2009), we can assume that the specific variances from the factor analysis model are very small compared to the prediction error.

In the current literature on SAE of social indicators, there is a research gap on the estimation of multidimensional indicators. In particular, the use of factor scores and factor analysis in SAE models is an open area of research. This research area is important when we have to deal with data dimensionality in the estimation of social indicators at a local level. In this paper, we consider economic wellbeing as a latent variable construct with the aim of reducing the dimensionality of wellbeing indicators. We then implement the unit-level SAE approach on the factor scores in both a simulation study and on real data from EU-SILC for the region of Tuscany, Italy.

In particular, we address the problem of providing reliable small area estimation of multidimensional economic wellbeing phenomena starting from an established wellbeing measurement framework, such as the Italian Equitable and Sustainable Wellbeing framework (BES). As mentioned, these frameworks are already developed within countries and are commonly used for the measurement of the Sustainable Development Goals. Therefore, we follow a two-step procedure: first latent variables are estimated based on the measurement framework via a confirmatory factor analysis model, and second the small area estimates along with their measures of uncertainty are computed via the EBLUP approach.

This paper is organized as follows. In section 2, we describe the factor analysis model for reducing data dimensionality on a dashboard of economic wellbeing indicators. In section 3 we review the unit-level SAE approach according to the BHF model and present the point estimation of the EBLUP for small area means. In section 4, we show results of a simulation study considering factor scores for data dimensionality reduction and contrast our approach to the typical approach of averaging single univariate EBLUPs on the original variables. When averaging single univariate EBLUPS on the original variables, we consider both a simple average and a weighted average where the weights are defined by the factor loadings. Moreover, we develop a parametric bootstrap algorithm to estimate mean squared errors (MSE) of the EBLUP of factor score means and evaluate its properties. In section 5, we discuss multidimensional economic wellbeing in Italy considering indicators from the Italian framework BES (Equitable and Sustainable Wellbeing) 2015 (ISTAT 2015). Also, using real data from EU-SILC 2009 for the area of Tuscany, we apply the proposed method and compute small area EBLUPs for factor score means and their mean squared error (MSE) for each Tuscany municipality (LAU 2). Finally, in section 6, we conclude with some final remarks and a general discussion.

2. Using Factor Scores for Data Dimensionality Reduction

In this section, we provide a general discussion on the use of factor analysis models to reduce data dimensionality and focus on the estimation of factor scores. Since the focus of the application in Section 5 is on measuring economic wellbeing based on a given substantive framework and a small number of single indicators, we consider here a one-factor analysis model. We acknowledge that in the presence of more complex multidimensional phenomena, one factor may not explain the total variability. Moretti, Shlomo and Sakshaug (2017) investigate the issue of multiple latent factors under a multivariate SAE approach.

2.1 Issues in Composite Indicators

Multivariate statistical methods aim to reduce the dimensionality of a multivariate random variable $\boldsymbol{Y}$. Formally, consider a R^Kspace, where K denotes the number of observed variables where we want to represent the observations in a reduced space $R^M$with $M \ll K$. Bartholomew et al. (2008) suggests several multivariate statistical techniques in order to deal with data dimensionality reduction in the social sciences (e.g. principal component analysis, factor analysis models, multiple correspondence analysis, etc.). In this work, we consider the linear one-factor model, where the factor can be interpreted as a latent characteristic of the individuals revealed by the original variables. This model allows for making inference on the population, since the observable variables are linked to the unobservable factor by a probabilistic model to develop a composite indicator (Bartholomew et al., 2008).

There is an ongoing debate about how to construct indicators which are useful for decision makers to inform policies. Saisana and Tarantola (2002) and Nardo et al. (2005) emphasize that composite indicators are important when a summary of multidimensional phenomena is needed and propose factor analysis models. Nardo et al. (2005) highlight that factor analysis models reduce the data dimensionality of a set of sub-indicators whilst keeping the maximum proportion of the total variability of the observed data.

Given our focus on data dimensionality reduction from a well-established multidimensional wellbeing framework, the BES framework for Italy (ISTAT, 2015), the single indicators have already been grouped into wellbeing dimensions. One such dimension is the economic wellbeing dimension. Therefore, we use factor analysis models under a confirmatory approach.

Factor scores are estimated from a confirmatory factor analysis model. They are defined as composite estimates providing details on a unit's placement on the latent factor (DiStefano, Zhu and Mindrila, 2009). The factor scores, once estimated, are easy to interpret: they have the same economic interpretation of the observed responses as they are strongly linearly related to these via a linear model.

There have been some first attempts in SAE and data dimensionality reduction using factor analysis (e.g. Smith et al., 2015). Here, the construction of the composite indicators was on the small area EBLUPs of the single indicators. In our approach, we first construct the composite indicator from the factor analysis model and then obtain small area estimates of the average factor score. We also focus on mean squared error (MSE) estimation for the estimates.

2.2 The Linear One-Factor Analysis Model

Let us consider a $K \times 1$ vector of observed variables $\boldsymbol{Y}$ and we assume that they are linearly dependent on a factor $\boldsymbol{f}$. Thus, we can write the following linking model (Kaplan, 2009):

$$\boldsymbol{Y} = \boldsymbol{\Lambda f} + \boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{\epsilon}$ denotes a vector $K \times 1$ containing both measurement and specific error, and $\boldsymbol{\Lambda}$ is a $K \times 1$ vector of factor loadings.

It is assumed that:

i) $E(\boldsymbol{\epsilon}) = \boldsymbol{0}$,

ii) $Var(\boldsymbol{\epsilon}) = \boldsymbol{\Theta}$,

iii) $\boldsymbol{\epsilon} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{\Theta})$,

iv) $\boldsymbol{\epsilon}$'s components are uncorrelated,

v) $E(\boldsymbol{f}) = 0$,

vi) $Cov(\boldsymbol{\epsilon}, \boldsymbol{f}) = \boldsymbol{0}$.

Therefore, the covariance matrix of the observed data is given by:

$$\boldsymbol{\Sigma} = Cov(\boldsymbol{YY}') = \boldsymbol{\Lambda}\Phi\boldsymbol{\Lambda}' + \boldsymbol{\Theta}, \tag{2}$$

where Φ denotes the factor variance, and $\boldsymbol{\Theta}$ is a $K \times K$ diagonal matrix of specific variance.

The maximum-likelihood (ML) approach is used to estimate the model parameters. ML equations under factor analysis models are complicated to solve, so iterative numerical algorithms are proposed in the literature (see e.g. Mardia, Kent and Bibby 1979). The log-likelihood function ℓ of the data $\boldsymbol{Y}$ can be written as follows (Hardle and Simar, 2012):

$$\ell(\boldsymbol{Y}; \boldsymbol{\Lambda}, \boldsymbol{\Theta}) = -\frac{nK}{2}\log(2\pi) - \frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{n-1}{2}tr(\boldsymbol{S\Sigma}^{-1}), \tag{3}$$

where $\boldsymbol{S}$ denotes the sample covariance matrix.

After the model parameters are estimated, the factor scores are also estimated. Factor scores are defined as estimates of the unobserved latent variables for each unit *i*. For a review of estimated factor scores we refer to Johnson and Wichern (1998). Using Bartlett's method, the individual factor scores estimate for $i = 1, \dots, n$ are given by (Bartholomew, Deary, and Lawn, 2009):

$$\hat{f}_i = \hat{\boldsymbol{\Gamma}}\hat{\boldsymbol{\Lambda}}'\hat{\boldsymbol{\Theta}}^{-1}\boldsymbol{y}_i\,. \tag{4}$$

Where $\hat{\boldsymbol{\Gamma}} = \hat{\boldsymbol{\Lambda}}'\hat{\boldsymbol{\Theta}}^{-1}\hat{\boldsymbol{\Lambda}}\ \hat{\boldsymbol{\Lambda}}$ and $\boldsymbol{y}_i$ denotes a K-dimensional vector of observations of K components of $\boldsymbol{Y}$ for $i = 1, \dots, n$.

Bartlett's method produces unbiased estimates of the true factor scores (Hershberger, 2005).

In the application presented in section 5, we also have binary dependent variables. According to Muthén and Muthén (2012) *logistic regression* is employed for binary dependent variables where the following transformation is applied in a single-factor model for each observed variable *k*:

$$logit\ [\pi_k(\boldsymbol{f})] = log\,\frac{\pi_k(\boldsymbol{f})}{1-\pi_k(\boldsymbol{f})} = \lambda_k \boldsymbol{f}\,, k = 1, \dots, K. \tag{5}$$

where $\pi_k(\boldsymbol{f})$ denotes the probability that the dependent variable is equal to one, and $\frac{\pi_k(\boldsymbol{f})}{1-\pi_k(\boldsymbol{f})}$ the odds. We can then write the following expression:

$$\pi_k(\boldsymbol{f}) = \frac{\exp(\lambda_k \boldsymbol{f}\,)}{1+\exp(\lambda_k \boldsymbol{f}\,)}, \tag{6}$$

which is monotonic in $\boldsymbol{f}$ and with domain in the interval [0,1].

In the presence of binary and continuous observed variables and under a maximum likelihood estimation approach, the factor scores may be estimated via the expected posterior method described in Muthén (2012) and applied in Mplus, Version 7.4.

3. Small Area Estimation using Empirical Best Linear Unbiased Prediction (EBLUP)

A class of models for SAE is the mixed effects models where we include random area-specific effects in the models and take into account the *between*-area variation.

3.1. Notation

Let $d = 1, \dots, D$ denote small areas for which we want to compute estimates of the target population parameter for each *d*, in our case the population mean $\bar{F}_d$ of the factor score. For a sample $s \subset \Omega$ of size n drawn from the target population of size N, the non-sampled units, $N - n$ are denoted by r.

Hence, $s_d = s \cap \Omega_d$ is the sub-sample from the small area d of size n_d, $n = \sum_{d=1}^{D} n_d$, and $s = \cup_d s_d$. r_d denotes the non-sampled units for the small area d of $N_d - n_d$ dimension.

3.2. Model based prediction using EBLUP

We consider the small area estimation problem for the mean under the EBLUP approach in the BHF model. Focusing on the population parameter of factor score means $\bar{F}_d,\ d = 1, \dots, D$, and as the population mean is a linear quantity, we can write the following decomposition:

$$\bar{F}_d = N_d^{-1}\left(\sum_{i \in s_d} f_{di} + \sum_{i \in r_d} f_{di}\right). \tag{7}$$

where f_{di} is the population factor score for unit *i* within small area *d* assuming that the factor model is implemented on the whole population.

When auxiliary variables are available at the unit level the BHF model can be used in order to predict the out-of-sample units. Considering the data for unit i in area d being $(f_{di}, \boldsymbol{x}_{di}^T)$ where $\boldsymbol{x}_{di}^T$ denotes a vector of p auxiliary variables, the nested error regression model is the following:

$$\begin{gathered} f_{di} = \boldsymbol{x}_{di}^T\boldsymbol{\beta} + u_d + e_{di}, i = 1, \dots, N_D, d = 1, \dots, D \\ u_d \sim \mathrm{iidN}(0, \sigma_u^2), e_{di} \sim \mathrm{iidN}(0, \sigma_e^2), \textit{independent}. \end{gathered} \tag{8}$$

In this model there are two error components, u_d and e_{di}, the random effect and the residual error term, respectively.

According to Royall (1970), we can write the best linear unbiased predictor (BLUP) for the mean as follows:

$$\tilde{\bar{F}}_d^{BLUP} = N_d^{-1}\left(\sum_{i \in s_d} f_{di} + \sum_{i \in r_d} \tilde{f}_{di}\right). \tag{9}$$

Where $\tilde{f}_{di} = \boldsymbol{x}_{di}^T\tilde{\boldsymbol{\beta}} + \tilde{u}_d$ is the BLUP of f_{di}, and $\tilde{u}_d = \gamma_d(\bar{f}_{ds} - \bar{\boldsymbol{x}}_{ds}^T\tilde{\boldsymbol{\beta}})$ the BLUP of u_d. Here, $\bar{f}_{ds} = n_d^{-1}\sum_{i \in s_d} f_{di}$, $\bar{x}_{ds} = n_d^{-1}\sum_{i \in s_d} x_{di}$, and $\gamma_d = \frac{\sigma_u^2}{\sigma_u^2 + \frac{\sigma_e^2}{n_d}} \in (0,1)$. γ_d is the *shrinkage estimator* measuring the unexplained between-area variability on the total variability.

Since in practice the variance components σ_e^2 and σ_u^2 are unknown, we replace these quantities by estimates, so we calculate the EBLUP of the mean:

$$\hat{\bar{F}}_d^{EBLUP} = N_d^{-1}\left(\sum_{i\in s_d} f_{di} + \sum_{i\in r_d} \hat{f}_{di}\right). \tag{10}$$

Where $\hat{f}_{di} = \boldsymbol{x}_{dj}^T\widehat{\boldsymbol{\beta}} + \hat{u}_d$ is the EBLUP of f_{di}. For details on $\widehat{\boldsymbol{\beta}}$ and $\hat{u}_d$ we refer to Rao and Molina (2015). As showed in Molina and Rao (2015), $\hat{\bar{F}}_d^{EBLUP}$ can be also written as follows:

$$\hat{\bar{F}}_d^{EBLUP} = \frac{n_d}{N_d}\bar{f}_{ds} + \left(\overline{\boldsymbol{X}}_{dp} - \frac{n_d}{N_d}\overline{\boldsymbol{x}}_{ds}\right)^T \widehat{\boldsymbol{\beta}} + \left(1 - \frac{n_d}{N_d}\right)\hat{u}_d. \tag{11}$$

$\overline{\boldsymbol{X}}_{dp}$ denotes the means of the auxiliary variable in the population for the d^{th} area.

If the sample size in a small area is zero, it holds that $\hat{\bar{F}}_d^{EBLUP} = \overline{\boldsymbol{X}}_{dp}\widehat{\boldsymbol{\beta}} = \hat{\bar{F}}_d^{Synthetic}$ where $\overline{\boldsymbol{X}}_{dp}$ denotes the means of the covariates in the population.

3.3. Mean Squared Error Estimation

The mean squared error (MSE) of (11) can be estimated via analytical approximations or resampling techniques. Prasad and Rao (1990) proposed an analytical approximation of MSE and González-Manteiga et al. (2008) proposed bootstrap techniques. Moreover, when large sample analytical approximations are available, the bootstrap might provide more accurate estimation alternatives to analytical approximations due to its second-order accuracy (González-Manteiga et al., 2008). Here, we suggest the use of a bootstrap method to estimate the MSE of (11). The bootstrap method proposed by González-Manteiga et al. (2008) has been adapted for the case of using factor score means as the dependent variable in the SAE models in order to take into account the variability arising from the factor analysis models. The steps are provided in appendix A and we evaluate our proposed algorithm via an extension to the simulation in Section 4.4. Analytical approximations of the MSE estimation of (11) under factor analysis models are a subject for future work.

4 Simulation Study

The simulation study was designed to assess the behavior of the EBLUP estimation of factor score means under a factor analysis model. We compare this approach with a weighted average of a dashboard of standardized univariate EBLUPs calculated from the original variables. We use a simple average and a weighted average where the weights are obtained by the factor loadings. We also assess the bootstrap MSE estimation for the EBLUP of factor score means which will be used in the application in Section 5.

The simulation is based on generating one population and drawing 500 simple random samples without replacement (SRSWOR) which is a mixture between a design- and model- based simulation approach where model assumptions are generally met and we mainly focus on sample variability. Drawing SRSWOR random samples from the population will result in the real setting of unplanned domains (zero sample sizes) within our small areas. Although EU-SILC may have complex survey designs, one important feature in the Italian EU-SILC for Tuscany is that every household (and hence adult in the household) has an equal inclusion probability (EPSEM) design and hence the simulation results based on an equal probability design are in line with the real data application. It is common to find in the literature other examples of simulation studies where simple random sampling is used to obtain unplanned domains, for example, Giusti, et al. (2013) used this approach when investigating a range of estimators also based on the EU-SILC. The subject of complex survey designs in SAE is a topic of ongoing research.

4.1 Generating the population

A single population is generated from a multivariate mixed-effects model, the *natural extension* of the BHF model (Fuller and Harter, 1987) with $N = 20{,}000$, $D = 80$, and $130 \leq N_d \leq 420$. N_d is generated from the discrete uniform distribution, $N_d \sim \mathcal{U}(a = 130, b = 420)$, with $\sum_{d=1}^{D} N_d = 20{,}000$ where the parameters are obtained from the Italian EU-SILC 2009 dataset used in the application in section 5. Here the multivariate model that we use to generate the population for the original variables (observed variables $\boldsymbol{Y}$) is:

$$\boldsymbol{y}_{di} = \boldsymbol{x}_{di}^{T}\boldsymbol{\beta} + \boldsymbol{u}_d + \boldsymbol{e}_{di}, i = 1, \dots, N_D, d = 1, \dots, D \qquad (12)$$
$$\boldsymbol{u}_d \sim \text{iid} MVN(\boldsymbol{0}, \boldsymbol{\Sigma_u}), \boldsymbol{e}_{di} \sim \text{iid} MVN(\boldsymbol{0}, \boldsymbol{\Sigma_e}), \textit{independent}.$$

$\boldsymbol{y}_{di}$ denotes a 3×1 vector of observed responses for unit i belonging to area *d*.

Two uncorrelated covariates are generated from the Normal distribution:

$$X_1 \sim N(9.93, 4.98^2), \qquad X_2 \sim N(57.13, 17.07^2).$$

These parameters reflect two real variables in the Italian EU-SILC 2009 dataset: the years of education and age (although we use here a normal (non-truncated) distribution). We selected *K*=3 response variables from the Italian EU-SILC 2009 data: the log of income, squared meters of the house, and the number of rooms, and fit regression models using the covariates X_1 and X_2.

From these models, we estimate the beta coefficient matrix and standard errors to build the simulation population by the model in (12). The $\boldsymbol{\beta}(3 \times 3)$ matrix of coefficients is given by:

$$\boldsymbol{\beta} = \begin{bmatrix} 3.983 & 0.018 & 0.001 \\ 1.263 & 0.007 & 0.005 \\ 0.404 & 0.006 & 0.002 \end{bmatrix}$$

The response vector was generated according to the following variance components, where the correlation was set at 0.5 as derived from the Italian EU-SILC 2009 data:

$$\boldsymbol{\Sigma}_e = \begin{bmatrix} 0.063 & 0.028 & 0.021 \\ 0.028 & 0.049 & 0.018 \\ 0.021 & 0.018 & 0.027 \end{bmatrix}.$$

We control the intra-class correlation ρ defined as $\rho_{y_k} = \sigma^2_{uy_k}/(\sigma^2_{uy_k} + \sigma^2_{ey_k})$, for the k^{th} component of $\boldsymbol{Y}$ and obtain the variance-covariance matrices of the correlated random effects. We chose three levels of intra-class correlations: 0.1, 0.3 and 0.8, and obtain the following matrices:

$$\boldsymbol{\Sigma}_u^{0.1} = \begin{bmatrix} 0.00693 & 0.00306 & 0.00227 \\ 0.00306 & 0.00539 & 0.00200 \\ 0.00227 & 0.00200 & 0.00297 \end{bmatrix},$$
$$\boldsymbol{\Sigma}_u^{0.3} = \begin{bmatrix} 0.02709 & 0.01195 & 0.00887 \\ 0.01195 & 0.02107 & 0.00782 \\ 0.00887 & 0.00782 & 0.01161 \end{bmatrix},$$
$$\boldsymbol{\Sigma}_u^{0.8} = \begin{bmatrix} 0.25500 & 0.11112 & 0.08249 \\ 0.11112 & 0.19600 & 0.07275 \\ 0.08249 & 0.07275 & 0.10800 \end{bmatrix}.$$

We first estimate the factor analysis model on the population to derive the population factor scores $f_i,\ i = 1, \dots, N$ according to (4). These will be treated as true values in our simulation study.
We note that although factor analysis models have been developed for multilevel structures within domains, it is not possible to use these models for *unplanned domains* given a random sample due to small and zero sample size domains. Thus, two-level factor analysis models in SAE is a subject for future work.

To derive the population factor scores, we first estimate an explanatory (unrestricted) factor analysis model (EFA) on the whole population, allowing for all possible factors. The EFA is estimated to check and identify the underlying relationships between observed variables (Norris and Lecavalier, 2009). The EFA results show that the first factor explains a large amount of the total variability. Table 1 shows the estimated eigenvalues under different scenarios and Figure 1 the scree plots. The eigenvalue represents the variance of factor m, and measures the variance in all the variables which is accounted for by that factor. With a large eigenvalue for the first factor, we then fit a one-factor confirmatory factor analysis model (CFA) on the population and estimate the population-based factor scores. The CFA one-factor model provides good fit statistics: $RMSEA = 0$ and $CFI = 1$, $TLI = 1$ (Hu and Bentler, 1999).

		Scenario		
		$\rho = 0.1$	$\rho = 0.3$	$\rho = 0.8$
Factors	***1***	2.060	2.055	2.139
	2	0.450	0.478	0.448
	3	0.440	0.450	0.402

Table 1 Eigenvalues from the EFA of the simulation population

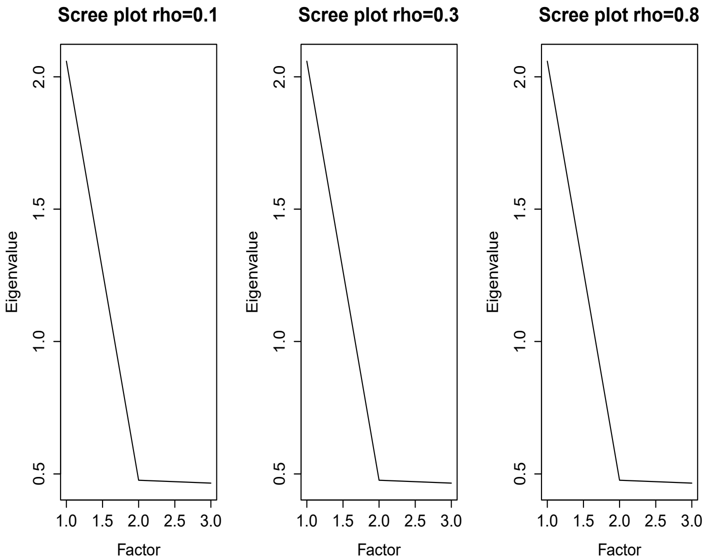

Figure 1 Scree plots from the EFA of the simulation population

We now define the following 'true' values for each of the small areas d from our simulated population for $i = 1, \dots, N$, area $d = 1, \dots, D$, and variables $k = 1, \dots, K$:

- the factor score means in area *d*: $\bar{F}_d = N_d^{-1} \sum_i f_{di}$,
- simple average of the observed variable standardized means in area *d*: $\bar{Y}_d^{S\_Averages} = \frac{\sum_{k=1}^{K} \bar{Y}_{dk}^{*}}{K}$;
- weighted average of the observed variables standardized means in area d using the factor

loadings: $\bar{Y}_d^{W\_Averages} = \frac{\sum_{k=1}^{K} \hat{\lambda}_k \bar{Y}_{dk}^*}{\sum_{k=1}^{K} \hat{\lambda}_k}$.

$\bar{Y}_{dk}^*$ denotes the standardized (mean zero and unit variance) true mean in area *d* and variable *k* where the standardization is obtained by subtracting the overall mean across all the areas and dividing by the standard deviation. $\hat{\lambda}_k$ denotes the estimated loading related to the k^{th} variable in the population obtained from the above CFA.

We highlight again that under factor analysis model assumptions the factor scores are strongly linearly related to the observed variables and have the same economic interpretation as the observed variables.

4.2 Simulation steps

The simulation study consists of the following steps:

1. Draw $S = 1, \ldots, 500$ samples using simple random sampling without replacement (note that this results in unplanned domains with small or zero sample size);
2. Fit the one-factor confirmatory factor analysis model on each sample and estimate the EBLUP of factor score means for each area *d* in each sample. We also calculate Horvitz-Thompson (HT) (Horvitz and Thompson, 1952) direct estimates of the factor score means for those areas with a non-zero sample size. In addition, the EBLUP for each of the original variables is also estimated in order to construct a simple average of the standardized small area EBLUPs and a weighted average using the factor loadings;
3. As the true values are known from the simulation population, we are able to calculate the root mean squared error (RMSE) and the relative bias (RBIAS) for each area *d* for the three types of estimates: EBLUPs of factor score means, and the simple and weighted average of EBLUPs. For example, for the EBLUPs of factor score means the RMSE is:

$$RMSE\left(\hat{\bar{F}}_d^{EBLUP}\right)_d = \sqrt{S^{-1} \sum_{s=1}^{S} \left(\hat{\bar{F}}_{ds}^{EBLUP} - \bar{F}_d\right)^2} \tag{13}$$

and the RBIAS is:

$$RBIAS\left(\hat{\bar{F}}_d^{EBLUP}\right)_d = S^{-1} \sum_{s=1}^{S} \frac{\left(\hat{\bar{F}}_{ds}^{EBLUP} - \bar{F}_d\right)}{\bar{F}_d}, \tag{14}$$

4. For the overall comparison across all areas, we rank the small areas according to the estimates averaged across the 500 samples and compare each to the ranking in the population. We also examine the average of the RMSE and RBIAS across all areas.

We estimate the EBLUP for each original variable separately on each of 500 samples, and then standardize them and construct weighted and simple averages. These are compared to the true values in the simulation population. The weighted mean in area *d* after standardizing the EBLUP estimates estimated on each sample *s* are given as follows:

$$\hat{\bar{Y}}_{ds}^{EBLUP\_W\_Averages} = \frac{\sum_{k=1}^{K}\left(\hat{\bar{Y}}_{dks}^{EBLUP*}\hat{\lambda}_{ks}\right)}{\sum_{k=1}^{K}\hat{\lambda}_{ks}}, d = 1, \dots, D, k = 1, \dots, K, \tag{15}$$

where k denotes the *k*th variable and $\hat{\lambda}_{ks}$ the factor loading estimated on the *s*th sample for the *k*th variable, and the standardized EBLUP of the mean is calculated as follows: $\hat{\bar{Y}}_{dks}^{EBLUP*} = (\hat{\bar{Y}}_{dks}^{EBLUP} - M_{ks}^{EBLUP})/SD_{ks}^{EBLUP}$ where $M_{ks}^{EBLUP} = D^{-1}\sum_d \hat{\bar{Y}}_{dks}^{EBLUP}$, and $SD_{ks}^{EBLUP} = \sqrt{(D-1)^{-1}\sum_d (\hat{\bar{Y}}_{dks}^{EBLUP} - M_{ks}^{EBLUP})^2}$.

In the following tables and figures we dropped the subscript *d* as we show the estimates averaged across all small areas.

4.3 Results: factor scores versus weighted and simple averages of standardized EBLUPs

In this section we show the main results of the simulation study. Table 2 contains the average eigenvalues across 500 samples under the EFA model and can be compared to Table 1 obtained from the simulation population. We can see that we are able to obtain good estimates for the eigenvalues across the samples. In parentheses we show the ratios between the sample and population eigenvalues. Table 3 presents the intra-class correlation coefficients estimated from the SAE model (averaged across 500 samples) showing that we approximate the known intra-class correlation coefficients as defined in the simulation population.

		Scenario		
		$\rho = 0.1$	$\rho = 0.3$	$\rho = 0.8$
Factors	***1***	2.058 (0.999)	2.050 (0.998)	2.135 (0.998)
	2	0.445 (0.989)	0.473 (0.990)	0.442 (0.987)
	3	0.442 (1.005)	0.455 (1.011)	0.405 (1.007)

Table 2 Average eigenvalues across 500 samples from EFA model.

Entries in parenthesis are ratios between the sample and population eigenvalues.

Scenario		
$\boldsymbol{\rho = 0.1}$	$\rho = 0.3$	$\rho = 0.8$
0.108	0.325	0.795

Table 3 Average intra-class correlation $\hat{\rho} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_e^2}$ *estimates across 500 samples*

For each of the three estimates in small area *d* averaged across the 500 samples, we compare the ranking of the small area domain estimates with the true ranking based on true area means according to our simulation population using a Spearman's correlation coefficient. These are shown in Table 4. The EBLUPs of the factor score means show an improvement and higher correlation to the true means in the population compared to the averages of EBLUPs, especially for the case of $\rho = 0.1$.

	Scenario		
	$\rho = 0.1$	$\rho = 0.3$	$\rho = 0.8$
$\widehat{\boldsymbol{Y}}^{\boldsymbol{EBLUP\_S\_Averages}}$	0.780	0.996	0.999
$\widehat{\boldsymbol{Y}}^{\boldsymbol{EBLUP\_W\_Averages}}$	0.793	0.996	0.998
$\widehat{\boldsymbol{F}}^{\boldsymbol{EBLUP}}$	0.986	0.997	0.999

Table 4 Spearman's correlation estimates for the three approaches

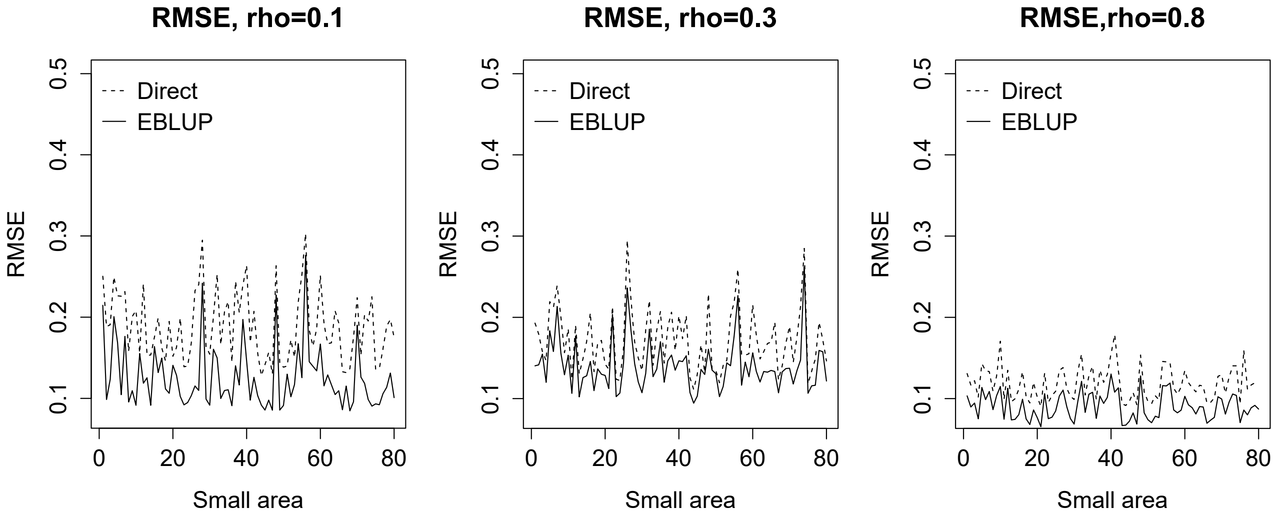

Figure 2 RMSE for Direct estimates and EBLUP of factor score means for small areas with $n_d > 0$.

Figure 2 shows the individual RMSE of the small areas for those areas with non-zero sample sizes. In line with the SAE literature the EBLUP approach produces estimates with lower variability than direct HT estimates. Table 5 shows the overall RMSE comparison defined in (10) across 500 samples for the EBLUPs of factor scores, and simple and weighted standardized EBLUPs. We do not show the overall relative bias *RBIAS* across the samples and areas since the estimates are all unbiased. In contrast to Figure 2, Table 5 presents the minimum, mean and maximum RMSE across all areas including those areas that had zero sample size and hence the synthetic estimator $\hat{F}_d^{Synthetic} = \overline{\boldsymbol{X}}_{dp}\hat{\boldsymbol{\beta}}$ where $\overline{\boldsymbol{X}}_{dp}$ denotes the means of the covariates in the population is used as the final estimator. The maximum values in Table 5 are generally obtained for those areas with zero or very small sample sizes. The larger the sample size, the smaller the RSME. The overall RMSEs for the EBLUP factor score means are lower than in the case of the simple and weighted averages of the dashboard of single EBLUPs for all levels of intra-class correlations, even after taking into account the extra modeling step of estimating factor scores. Hence, applying the EBLUP method on factor score means provides more precise estimates whilst reducing the data dimensionality of multiple observed variables.

Approach	**Statistics**	**Scenario**		
		$\rho = 0.1$	$\rho = 0.3$	$\rho = 0.8$
$\hat{\boldsymbol{Y}}^{\boldsymbol{EBLUP\_S\_Averages}}$	*Min*	0.590	0.247	0.083
	Mean	1.432	0.336	0.119
	Max	4.566	0.549	0.165
$\hat{\boldsymbol{Y}}^{\boldsymbol{EBLUP\_W\_Averages}}$	*Min*	0.610	0.247	0.083

	Mean	0.793	0.334	0.118
	Max	1.984	0.549	0.165
$\hat{\bar{F}}^{EBLUP}$	*Min*	0.085	0.094	0.065
	Mean	0.140	0.125	0.090
	Max	0.276	0.262	0.130

Table 5 RMSE estimates: comparison across 500 samples for the three approaches

4.4 Bootstrap MSE Estimation

In the application, we will use the algorithm defined in Appendix A to estimate the MSE of the EBLUP of the factor score means using a modified parametric bootstrap which take into account the variability arising from the factor analysis model. We extend here the simulation for the case of the intra-class correlation of 0.3 to assess the properties of our proposed bootstrap MSE estimation.

We compare the bootstrap RMSE according to the algorithm in Appendix A with the empirical RMSE (ERMSE) obtained across the 500 samples calculated as $ERMSE\left(\hat{\bar{F}}_d^{EBLUP}\right) = \sqrt{S^{-1}\sum_{s=1}^{S}(\hat{\bar{F}}_{ds}^{EBLUP} - \bar{F}_d)^2}$. We consider the ERMSE as the "true" MSE and assess whether our proposed modified parametric bootstrap MSE estimator is unbiased.

Figure 3 shows the ratio between the parametric bootstrap RMSE averaged across the 500 samples under two settings: (1) treating the factor scores as fixed, and (2) accounting for the variability of the factor analysis model, against the ERMSE. It can be seen that the RMSE estimated via parametric bootstrap without accounting for the factor model is underestimated with a relative bias of -34.6% across the small areas. However, the relative bias across the small areas when accounting for the variability in the factor analysis model is negligible at 4.0%.
To illustrate this point further, Figure 4 presents the coverage rate comparisons of the parametric bootstrap estimated MSE taking into account the factor analysis model variability and ignoring the factor analysis model variability. There are significantly smaller coverage rates if we ignore the factor analysis model variability. The coverage rate when taking the variability into account is relatively stable at 95%.

Therefore, we conclude from this extension to the simulation study that treating the factor scores as fixed in the standard parametric bootstrap approach leads to a severe underestimation in the RMSE and our modified parametric bootstrap in Appendix A performs well.

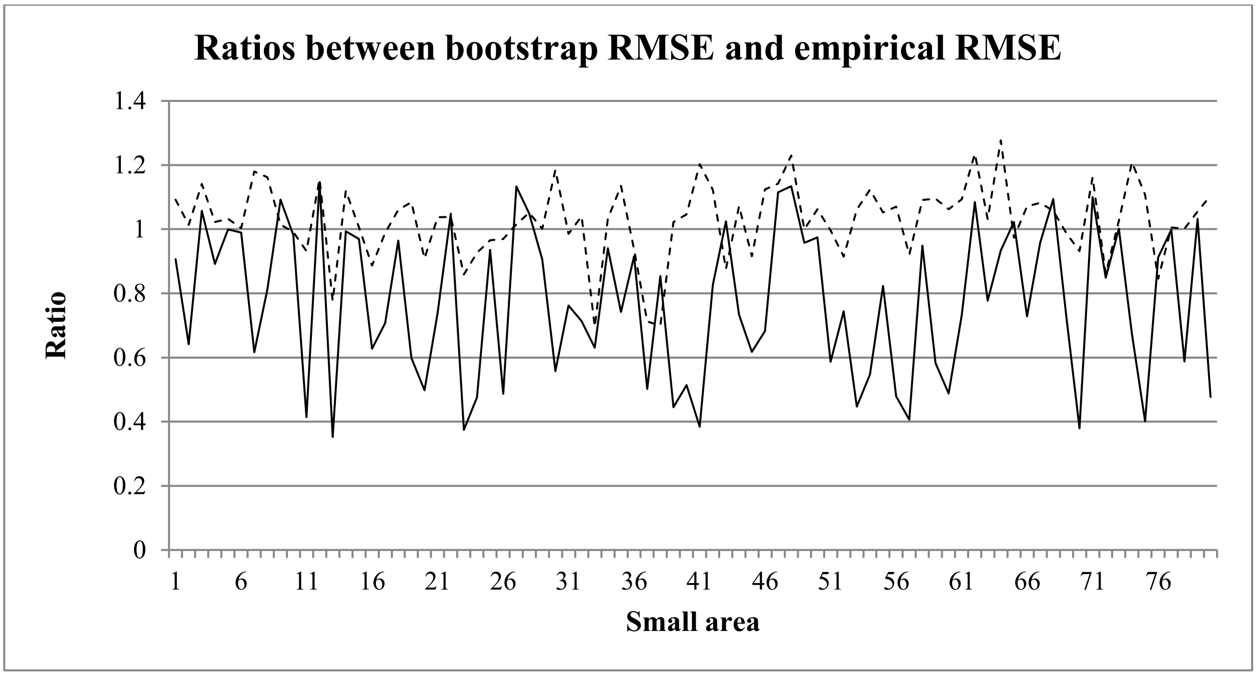

Figure 3 Ratios between bootstrap RMSE and empirical RMSE estimated via bootstrap taking into account the factor analysis model variability (---) and bootstrap ignoring the factor analysis model variability (__).

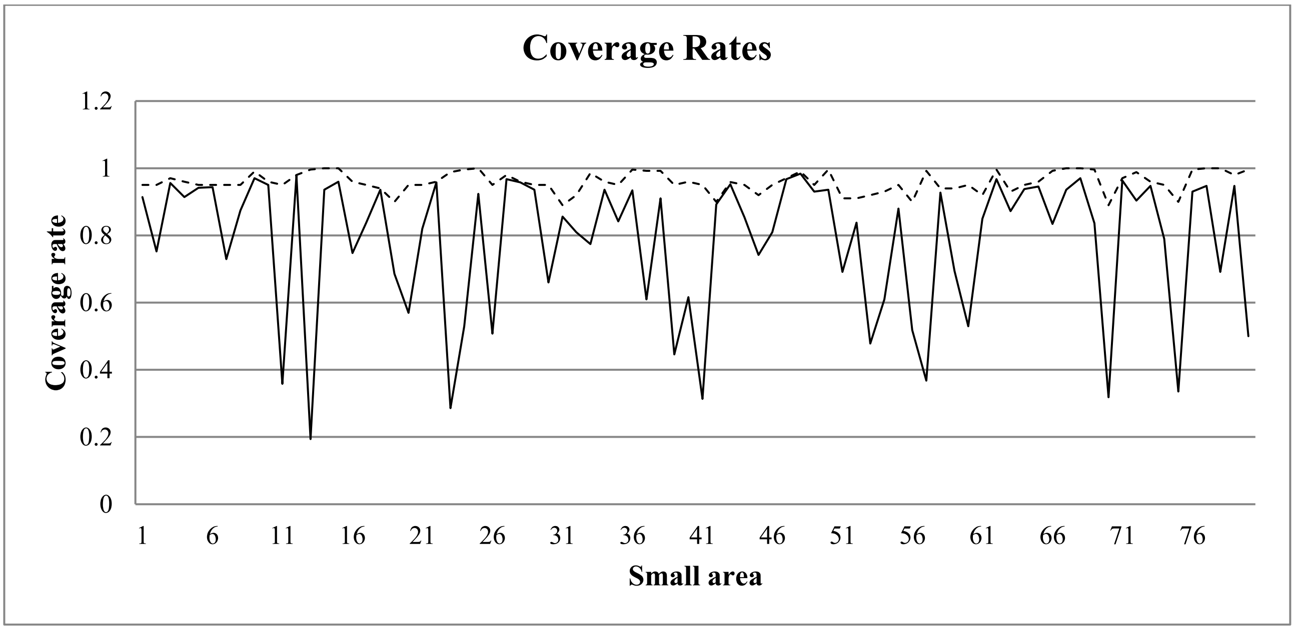

Figure 4 Coverage rates comparisons: bootstrap RMSE estimated taking into account the factor analysis model variability (---) and bootstrap ignoring the factor analysis model variability (__).

4.5 Final remarks of the simulation study

The use of factor scores provides better rankings to true values compared to weighted and simple averages of single variables, especially for the case of small intra-class correlations which are more common in real settings. Furthermore, it can be seen that factor scores provide estimates with lower variability (in terms of RMSE) than weighted and simple averages of single variables for estimating multidimensional phenomena at the small area level. We also conclude that it is crucial to consider the variability arising from the factor analysis model in the parametric bootstrap MSE estimation; otherwise, the true MSE will be underestimated.

Based on these results, we use the EBLUP of the factor score means approach to reduce the dimensionality of observed variables in a real application using the Italian 2009 EU-SILC data for the Tuscany region and the modified parametric bootstrap procedure for MSE calculations in Section 5.

5 Economic Wellbeing in Tuscany: a Multidimensional Approach

The aim of this section is to demonstrate how we can provide estimates of an economic wellbeing indicator following the BES guidelines for Tuscany municipalities. In our application, we use data from the EU-SILC 2009 and the 2001 General Census of Population and Housing. We note that the EU-SILC 2009 data were collected several years after the census and this is a limitation of the study since we assume stationarity of growth between the periods. Obviously the economic and financial crisis occurring in 2008 violates this assumption and further studies are needed with more current covariates. Nevertheless, the application is useful to demonstrate how small area estimates can be calculated for a multidimensional indicator. The specification of the main R functions used in this analysis are presented in Appendix C.

5.1 Data and variables

Income and economic resources can be seen as conditions by which an individual is able to have a sustainable standard of life. One of the dimensions in the Italian Equitable and Sustainable Wellbeing (BES) framework is dedicated to Economic Wellbeing (ISTAT 2015). It consists of ten single economic-related indicators (a dashboard of indicators). In this work, we focus on a subset of these highly correlated variables:

- Severe material deprivation according to Eurostat;
- Equivalized disposable income;
- Housing ownership;
- Housing density.

Appendix B in Figure B1 contains the variables nomenclature for the 2009 Tuscany EU-SILC dataset used in our study and descriptive statistics of these study variables which are explained in the next sections.

Material deprivation can be defined as the inability to afford some items considered to be desirable, or even necessary, to achieve an adequate standard of life. Indicators related to this are absolute measures useful to analyze and compare aspects of poverty in and across EU countries (Eurostat, 2012). According to Eurostat, material deprivation in the EU can be measured by the proportion of people whose living conditions are severely affected by a lack of basic resources. Technically, the severe material deprivation rate shows the proportion of people living in households that cannot afford *at least four* of the following nine items because of financial difficulty:

1. Mortgage or rent payments, utility bills, hire purchase installments or other loan payments;
2. One-week holiday away from home;
3. A meal with meat, chicken, fish or vegetarian equivalent every second day;
4. Unexpected financial expenses;
5. A telephone (including mobile telephone);
6. A color TV;
7. A washing machine;
8. A car;
9. Heating to keep the home sufficiently warm.

It can be argued that some of these indicators (e.g. 5 and 6) are nowadays less relevant than in the past. Nevertheless, these indicators are still used to describe the difficulties that households face in achieving a standard of life considered to be sufficient by society. This index is described in Table B3 in Appendix B. Disposable household income is the sum of gross personal income components plus gross income components at the household level minus employer's social insurance contributions, interest paid on mortgage, regular taxes on wealth, regular inter-household cash transfer paid and tax on income. In order to take into account differences in household size and composition, we consider disposable equivalized income I^{DE} defined as follows:

$$I_i^{DE} = \frac{I_i^D}{n_i^E}, i = 1, \dots, N, \tag{16}$$

where $i = 1, \dots, n$ denotes households, I_i^D is the disposable household income, n_i^E is the equivalized household size calculated in the following way (Haagenars et al., 1994):

$$n_i^E = 1 + 0.5 \cdot (HM_{14+} - 1) + 0.3 \cdot HM_{13-}, \tag{17}$$

where HM_{14+} and HM_{13+} are the numbers of household members aged 14 and over and 13 or younger at the end of the income reference period, respectively. This so-called 'OECD modified scaling' procedure is crucial to taking into account the economy of scales in the household. Due to the skewness of the variable, we use the log transformation in the factor model and SAE. The histograms are in Figure B2 and descriptive statistics in Table B1 of Appendix B. Housing ownership is measured by a dichotomous variable (0,1) where 0 denotes that the property where the household lies is not owned. According to the 2009 Tuscany EU-SILC data, 73.96% of households own the property where they live (see Table B3 in Appendix B. Overcrowding is one of the indicators that National Statistics Institutes include in their wellbeing measurement frameworks. A very simple indicator of housing density is given by the ratio between the number of rooms in the household (excluding kitchen, bathroom and rooms used for work purposes) and the household size:

$$\bar{r}_i = \frac{R_i}{M_i} \qquad (18)$$

where i is the household, M_i denotes the number of people in the i^{th} household, and R_i the number of rooms in the household. The histogram of this variable is in Figure B3 and descriptive statistics are in Table B2 of Appendix B.

EU-SILC is conducted yearly by ISTAT for Italy, and coordinated by EUROSTAT at the EU level. The survey is designed to produce accurate estimates at the national and regional levels (NUTS-2). Hence, for the Italian geography the survey is not representative of provinces, municipalities (NUTS-3 and LAU-2 levels, respectively), and lower geographical levels. The regional samples are based on a stratified two-stage sample design. The Primary Sampling Units (PSUs) are the municipalities within the provinces, and households are the Secondary Sampling Units (SSUs). The PSUs are stratified according to their population size and SSUs are selected by systematic sampling in each selected PSU. Each household has an equal probability of selection. The total number of households in the sample for Tuscany is 1,448.

The 14th Population and Housing Census 2001 surveyed 1,388,252 households of persons living in Tuscany permanently or temporarily, including the homeless population and persons without a dwelling.

5.2 The construction of the factor scores

The one-factor analysis model described in section 2 is fitted, and according to the goodness-of-fit statistics estimated on the one-factor model solution, the Root Mean Squared Error of Approximation (RMSEA=0.047) and the Comparative Fit Index criteria (CFI=0.966), the model provides good fit (Hu and Bentler 1999). This choice can be justified also substantively as our variables relate to economic wellbeing according to the BES framework, which is the phenomenon we want to measure.

The histogram, Q-Q plot, and box-plot of the factor scores are shown in Figure 5 as well as descriptive statistics in Table 6. We see evidence of a slight skewness in the factor scores likely due to discrete variables included in the factor analysis model. One interesting thing to note based on Table B4 in Appendix B is that the estimated intra-class correlation (ICC) for the factor scores is 0.1987 which is considerably higher than the estimated ICC's for the single study variables, thus as seen in the simulation study, we expect that the EBLUP of the factor scores will provide good rankings of the small areas compared to weighted and simple averages.

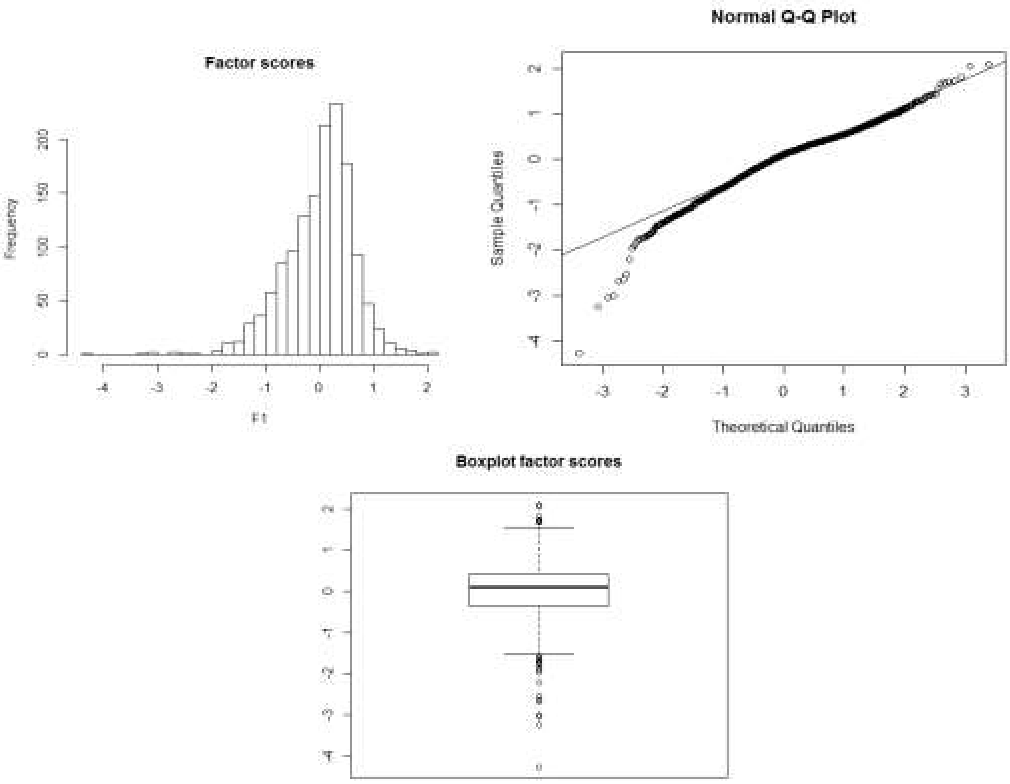

Figure 5 Factor scores distribution graphs.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max	S.d.	$\widehat{ICC}$
-4.2630	-0.3712	0.1050	0.0034	0.4120	2.0940	0.6436	0.1987

Table 6 Descriptive statistics of factor scores.

5.3 Small area estimates

In this application we treat municipalities as our small areas of interest. The municipalities within Tuscany are unplanned domains in EU-SILC and only 59 out of 287 were sampled. Sample sizes in municipalities range from 0 to 135 households.

First, we provide direct estimates for the small areas with $n_d > 0$. After this, we build a SAE model under the BHF approach where the response variable is the factor score interpreted as the latent economic wellbeing construct. The exploratory variables in the model relate to the head of the household and are those common to both the survey and Census data. In particular, after a preliminary analysis of the available data we chose gender, age, year of education, household size, size of the flat (in squared meters), and employment status as the explanatory variables.

The single EBLUPs of the dashboard indicators have been estimated to construct the simple and weighted averages, as was done in the simulation study. In the case of binary variables the following linear logistic mixed effects model was fitted (MacGibbon and Tomber 1989):

$$logit(p_{di}) = log\left(\frac{p_{di}}{1 - p_{di}}\right) = \boldsymbol{x}_{di}^T\boldsymbol{\beta} + u_d, \tag{19}$$

where p_{di} is the probability that $y_{di} = 1$ and u_d ~iidN$(0, \sigma_u^2)$.

In Figure 6 we compare the relative root mean squared error (RRMSE) of the EBLUPs of factor score means with the coefficients of variation of the direct estimates for the sampled areas (to the right of the vertical line). We also include in Figure 6 the RRMSE for the non-sampled areas where $n_d = 0$ (to the left of the vertical line). Here, the estimates of the MSE for the predictions are obtained via the modified parametric bootstrap with $B = 500$ bootstrap samples as described in Appendix A. We can see the gain in efficiency (in terms of reduction in the RRMSE) obtained by the EBLUP compared to the direct estimates and in particular the RRMSE's are below 10%. In addition, even when the synthetic estimators are used in those areas with zero sample sizes, we still obtain an RRMSE that is below 20%. We note that an estimator with an RRMSE below 20% are considered reliable estimates (Australian Bureau of Statistics, 2015).

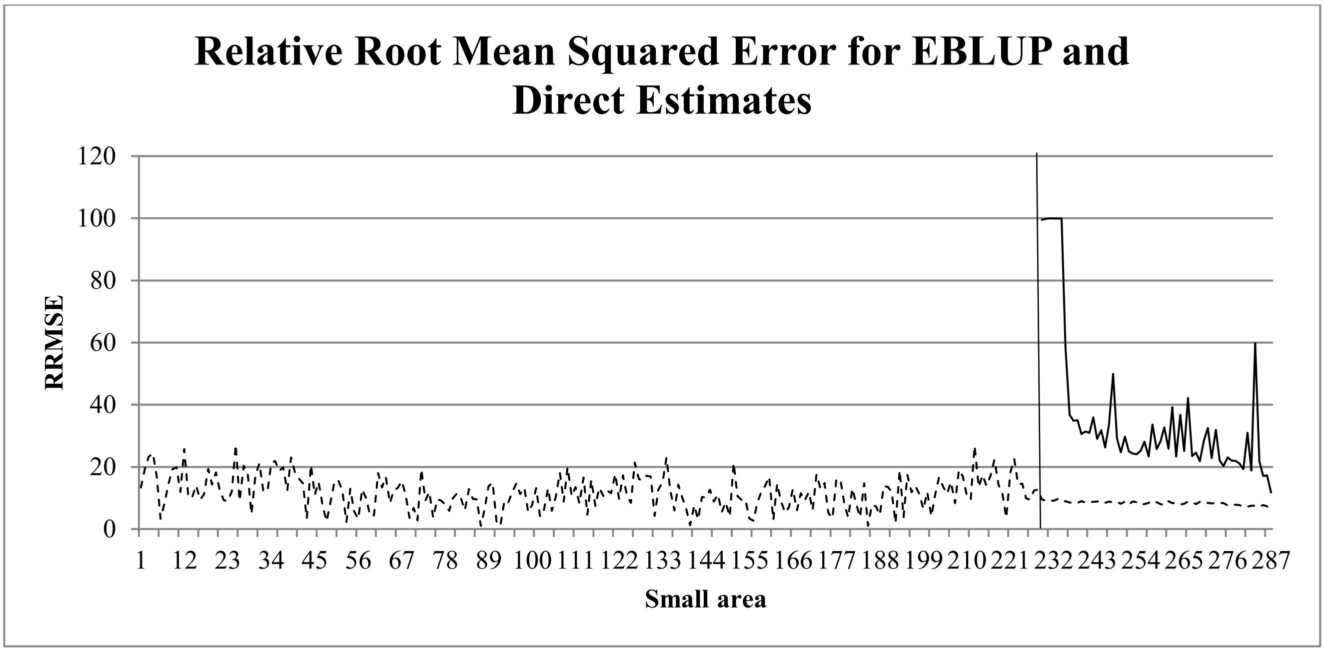

Figure 6 RRMSE direct estimates (__) and EBLUPs (---) ordered by growing sample size.

To facilitate the interpretation and provide a comparison between the different economic wellbeing indicators obtained from the EBLUP factor score means and the simple and weighted averages of the dashboard of EBLUPs, we have normalized the EBLUPs using the 'Min-Max' method (OECD-JRC, 2008), with range $[0,1]$. For the factor score EBLUPs, the normalization (denoted with a '*') is as follows:

$$\hat{\hat{F}}_d^{*EBLUP} = \frac{\hat{\hat{F}}_d^{EBLUP} - \min(\hat{\hat{F}}^{EBLUP})}{\max\left(\hat{\hat{F}}^{EBLUP}\right) - \min(\hat{\hat{F}}^{EBLUP})}, d = 1, \dots, D, \qquad (20)$$

where $\hat{\hat{F}}^{EBLUP} = col_{1 \le d \le D} \hat{\hat{F}}_d^{EBLUP}$. And similarly, for the simple and weighted averages of the dashboard of standardized EBLUPs.

Table 7 shows the percentiles for the latent economic wellbeing indicator based on the normalized EBLUP factor scores and the normalized averages of the dashboard of EBLUPs. Figure 7 and Figure 8 depict the maps of the quartiles of the EBLUPs under the different approaches for the Tuscany region.

Percentile	**0%**	**25%**	**50%**	**75%**	**100%**
EBLUP	0.0000	0.5110	0.5468	0.5819	1.0000

Simple	0.0000	0.4297	0.5297	0.6061	1.0000
Weighted	0.0000	0.4796	0.6006	0.7184	1.0000

Table 7 Percentiles for the transformed latent economic wellbeing indicator based on the EBLUP of factor score means and simple and weighted averages

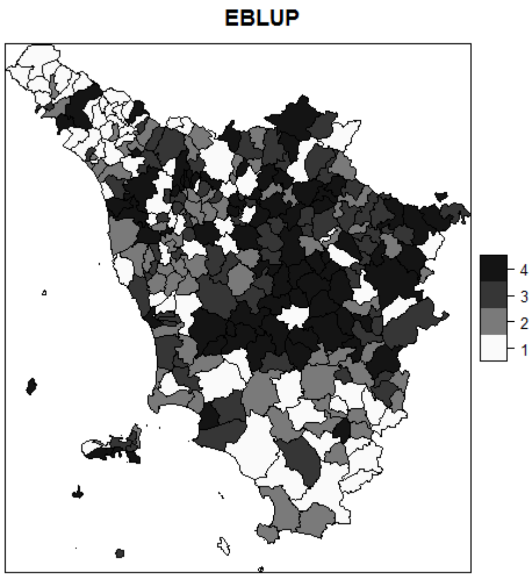

Figure 7 Latent economic wellbeing indicator based on transformed EBLUP of factor scores means {1=1st quartile; 2=2nd quartile; 3=3rd quartile; 4=4th quartile}

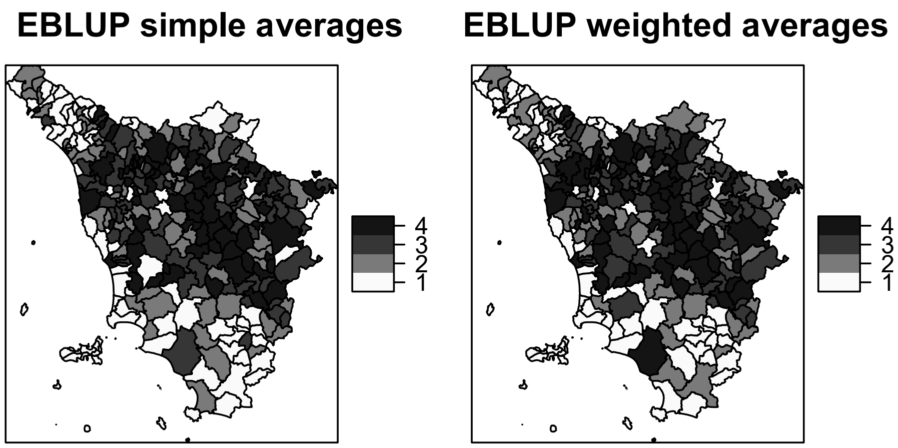

Figure 8 Latent economic wellbeing indicator based on simple and weighted averages of single EBLUPs {1=1st quartile; 2=2nd quartile; 3=3rd quartile; 4=4th quartile}.

In the maps of Figure 7 and Figure 8 a darker color denotes a better wellbeing phenomenon. Looking at these figures we can draw some interesting conclusions on economic wellbeing in the

Tuscany region.

The municipalities located in the Massa-Carrara province, which is based in the North of Tuscany (i.e. Pontremoli and Zeri municipalities), and municipalities based in Grosseto province (south of Tuscany), are the poorest ones. The small areas based in the Florence province are wealthy municipalities, as well as the ones located in the center of the region (Siena province). The lowest point estimates of the latent economic wellbeing indicator are estimated for Carrara and Seravezza municipalities, and the highest values for Firenze and Arezzo municipalities. Our results based on the EBLUPs of the factor scores in Figure 7 are more comparable with other SAE studies on welfare and poverty in Tuscany (Marchetti, Tzavidis, and Pratesi 2012; Giusti et al. 2015) compared to the averages of a dashboard of EBLUPs in Figure 8, though previous SAE studies consider only income variables rather than a composite indicator used here. This is not surprising given the low ICCs for each of the individual EBLUPs that form the dashboard which may result in more distortions on the rankings, particularly since some of the individual EBLUPs are based on discrete variables.

5.4 Model diagnostics

We assess the fit of the model by analyzing the *level-1* and *level-2* standardized residuals. In particular, the Q-Q plots of the residuals, shown in Figure 9 and Figure 10 show the leverage measures versus standardized scaled residuals from the linear model. Both figures show a presence of outliers in the left tail, although the factor scores distribution is approximately symmetric. Figure 10 also shows the contour of the Cook's distance which does not deviate much from zero and hence we can conclude that the outliers are not influential.

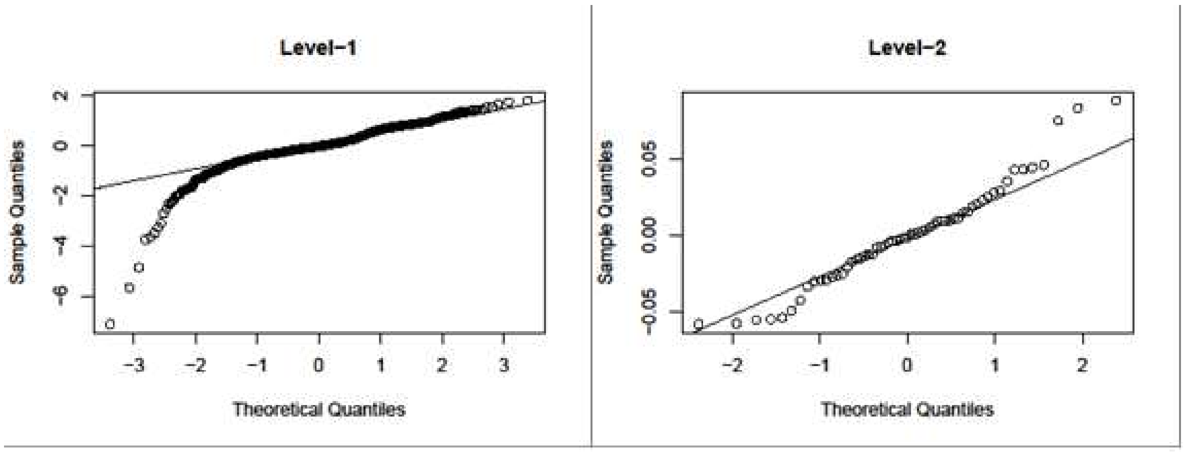

Figure 9 Q-Q plots for the level-1 and level-2 residuals of the BHF model fitting.

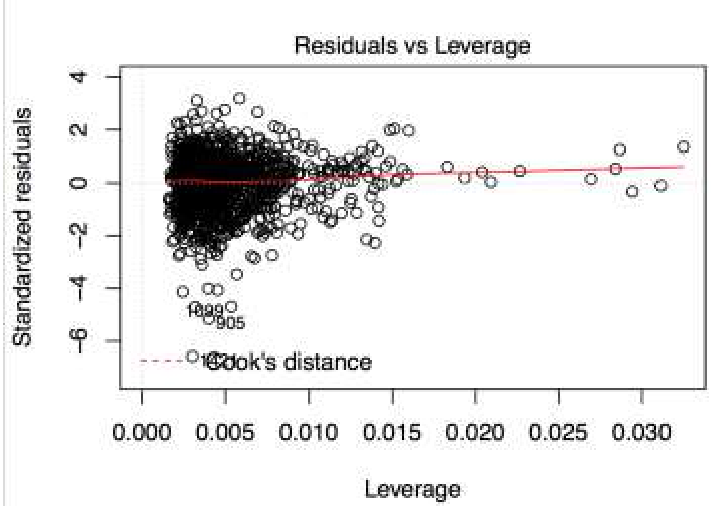

Figure 10 Standardized residuals versus leverage measure.

6 Conclusion and Discussion

In this paper we evaluated a method to estimate the mean of a latent economic wellbeing indicator at the local level for Tuscany using factor scores to reduce data dimensionality. We focused on the factor scores because they can be seen as a *latent* economic wellbeing composite variable. The simulation study demonstrated that factor score means provide a better ranking of the small areas compared to the true population means as measured by the Spearman's correlation coefficient, especially when intra-class correlations are small, which is common in real settings. The simple and weighted averages of univariate standardized EBLUPs also provide good rankings for the higher intra-class correlations that were examined. In addition, the use of factor scores provided more precise estimates in terms of the MSE for an estimate of a multidimensional phenomenon compared to the averages of the EBLUPs. The use of factor analysis models and factor scores has important advantages and implications in data dimensionality reduction: it avoids arbitrary weighting of single indicators and it generates continuous composite scores, which can be modeled using model-based SAE methods. Since the factor scores are strongly linearly related to the multidimensional observed variables, this leads to easier interpretation.

Another important point studied in this paper, is the MSE estimation of EBLUPs of factor score means. In this work, we proposed a modification to the González-Manteiga et al. (2008) parametric bootstrap algorithm to account for the additional variability added to the small area estimates by using factor scores obtained from a factor analysis model as the dependent variable. This has been tested via simulation and we showed that if the variability arising from the factor analysis model is ignored, the MSE is underestimated and therefore biased. For more theoretical details on the bootstrap, we refer to González-Manteiga et al. (2008). Analytical MSE approximations are left for future work.

There are several areas where this work could be extended. Future work might consider other geographical levels, such as SLL (Sistemi Locali del Lavoro – Labor Local System), by looking at the flow of daily travel home/work (commuting) detected during the General Census of Population and Housing. Further interesting applications would involve comparisons between Italian regions in the North, Central, and South.

Another worthwhile extension is accounting for more than one factor. When the goal is to reduce the dimensionality of the original data by identifying latent factors, one might face the issue of identifying multiple factors. Multiple latent factors can arise, particularly when we have many indicators referring to the same phenomenon which can be grouped substantively into subdomains. For example, if the goal is to study housing quality we may want to consider the following dimensions: type of dwelling and tenure status, housing affordability, and housing quality (e.g. overcrowding, housing deprivation, problems in the residential area). For multiple latent factors, we may have factor scores that are correlated, and hence future research should explore the use of the multivariate mixed effects model (Fuller and Harter, 1987). Datta, Day, and Basawa (1999) showed that the use of the multivariate mixed effects model might lead to gains in efficiency in terms of MSE for the EBLUP compared to the BHF model. Therefore, the multivariate small area estimation method might provide better dashboard estimates and averages if the correlation between the single variables is taken into account. These extensions are currently being carried out in Moretti, Shlomo and Sakshaug (2017).

Appendix A: Parametric bootstrap procedure for the EBLUPs of factor scores MSEs.

Here we show the bootstrap steps for the EBLUP's MSE. The bootstrap procedure is the one proposed by González-Manteiga et al. (2008) and we particularize the algorithm by taking into account the factor analysis model variability (in step 1).

1. Draw $b = 1, \dots, B$ simple random samples with replacement from the observed sample **S** and estimate factor analysis models to obtain factor scores. After this, the usual parametric bootstrap proposed by González-Manteiga et al. (2008) is run for the $b = 1 \dots, B$ bootstraps.
2. Fit the Battese, Harter and Fuller model to the sampled units $\boldsymbol{f}_b = (\boldsymbol{f}'_{1b}, \dots, \boldsymbol{f}'_{Db})'$, and estimate the model parameters $\widehat{\boldsymbol{\beta}}, \hat{\sigma}_u^2$ and $\hat{\sigma}_e^2$.
3. Generate $u_d^{*(b)} \sim \text{iidN}(0, \hat{\sigma}_u^2), d = 1, \dots, D$, which are the bootstrap area effects.
4. Generate the bootstrap errors for the sample units $e_{di}^{*(b)} \sim \text{iidN}(0, \hat{\sigma}_e^2)$, independently of the $u_d^{*(b)}$ and the error domain means $\bar{E}_d^{*(b)} \sim \text{iidN}\left(0, \frac{\hat{\sigma}_e^2}{N_d}\right), d = 1, \dots, D$.
5. Calculate the true means for each small area of the bootstrap population as follows:
$$\bar{F}_d^{*(b)} = \overline{\boldsymbol{X}}'_d \widehat{\boldsymbol{\beta}} + u_d^{*(b)} + \bar{E}_d^{*(b)}, d = 1, \dots, D,$$
where $\overline{\boldsymbol{X}}'_d$ denotes the means of the population (auxiliary variables).

6. Generate the responses for the sample units by using the sample covariates vectors $\boldsymbol{x}_{di}, i \in s_d$:
$$F_{di}^{*(b)} = \boldsymbol{x}'_{di} \hat{\beta} + u_d^{*(b)} + e_d^{*(b)},\ d = 1, \dots, D.$$
7. Fit the nested errors model to the bootstrap sample data $F_{di}^{*(b)}$ and obtain the bootstrap EBLUPs $\hat{\bar{F}}_d^{*(b)}, d = 1, \dots, D$.
8. Replicate steps from 1 to 7 for $b = 1, \dots, B$. The Monte Carlo approximation of the bootstrap estimator of the EBLUP is given by:
$$mse\left(\hat{\bar{F}}_d^{EBLUP}\right) = \frac{1}{B} \sum_{b=1}^{B} \left(\hat{\bar{F}}_d^{*(b)} - \bar{F}_d^{*(b)}\right)^2, d = 1, \dots, D.$$
$\bar{F}_d^{*(b)}$ denotes the true mean and $\hat{\bar{F}}_d^{*(b)}$ the EBLUP for the area d for replicate b.

We run the bootstrap procedure with B=500 both in the simulation and application.

Appendix B: EU-SILC data study variables

Here we describe the Italian EU-SILC 2009 data nomenclature and show some descriptive statistics on the study variables.

Variable name	*Description*
FCOM	Area code: comune (municipality)
HOUSEHOLD CROSS-SECTIONAL WEIGHT	Cross-sectional survey weight
TOTAL DISPOSABLE HOUSEHOLD INCOME	Total disposable household income
STANZE	Rooms in the flat (except: kitchen, toilet and bathroom, hallway, corridor, rooms used for work purposes).
GODAB_B	House ownership variable indicator
Material deprivation variables	
IMPREV	Ability to deal with unexpected expenses of €1000
FERIE	Affordability of one week per year away from home
PASTO	Affordability of a meat or chicken, or fish (or equivalent vegetarian) every two days
RISADE	Capacity of heating the house properly
LAVATR	Washing machine ownership
TV	TV ownership
AUTO	Car ownership
CELL	Telephone ownership
PAGAFF	Difficulties in paying the rent
PAGBOL	Difficulties in paying bills
PAGALDEB	Difficulties in paying loans or something similar
PAGMUT	Difficulties in paying the mortgage

Figure B1. Italian EU-SILC variables nomenclature

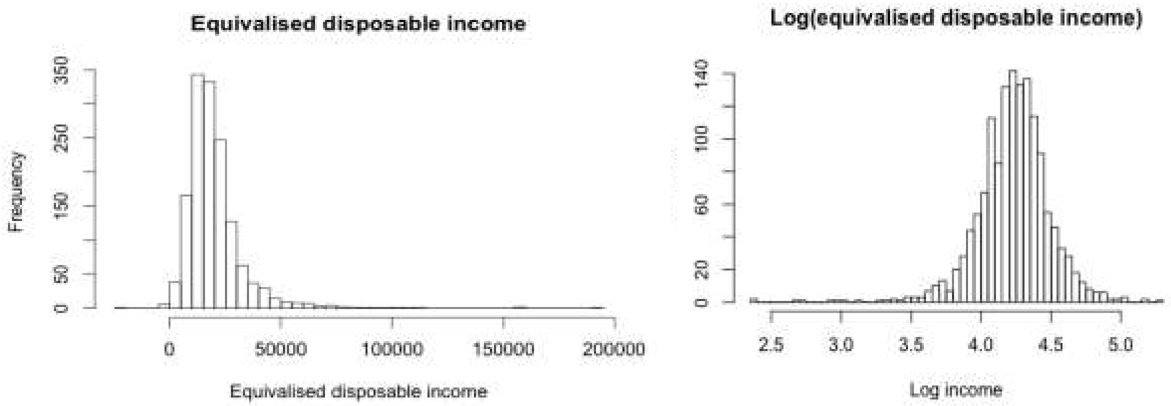

Figure B2. Disposable equivalized income histogram

Min.	1st Qu.	Median	Mean	3rd Qu.	Max	S.d.
-24,670	12,200	17,410	20,090	23,740	190,800	13,990.88
2.398	4.087	4.243	4.231	4.377	5.280	0.264

Table B1. Equivalized disposable income and log equivalized disposable income descriptive statistics

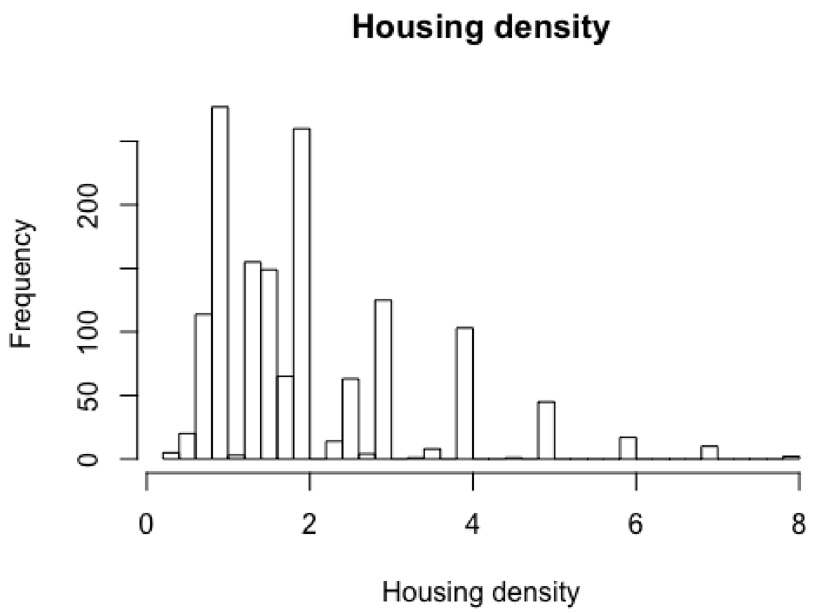

Figure B3. Housing density

Min.	1^{st} Qu.	Median	Mean	3^{rd} Qu.	Max	S.d.
0.250	1.000	1.600	1.989	2.500	8.000	1.239

Table B2. Descriptive statistics of the housing density

Variable	Frequency %
Material deprivation	3.94%
House ownership	73.96%

Table B3. Frequencies of the binary variables

Variable	Estimated ICC
Factor scores	0.1987
Disposable equivalized income	0.0019
Room average	0.0680
Material deprivation	0.0189
House ownership	0.0410

Table B4. Estimation of the ICCs of the study variables and factor scores

Factor	Eigenvalue
1	**1.791**
2	**1.000**
3	0.727
4	0.566

Table B5. Eigenvalues from exploratory factor analysis model on Tuscany EU-SILC 2009

Appendix C: Specification of the main R functions

Here we describe the main R packages we used for the small area estimates. All the other analyses were programmed manually.

C.1 Estimation of small area means and MSE under EBLUP approach with the "sae" package (Molina and Marhuenda 2015)

- Required packages: nlme, MASS
- Functions: eblupBHF( ) and pbmseBHF( ).

C.2 Running Mplus models in the R environment via MplusAutomation (Muthén and Muthén, (2012), Hallquist and Wiley (2014))

- Functions: mplusObject( ), mplusModeler( ).

C.3 Mapping using spdep, maptools, sp, Hmisc

- Functions: readShapePoly( ), spplot( )

References

Australian Bureau of Statistics. 2015. Survey of disability, ageing and carers, 2012. Modelled estimates for small areas, projected 2015. Available at: https://www.health.gov.au/internet/main/publishing.nsf/Content/98DCE47FC10BDD51CA257F15000413F5/$File/SDAC%202012%20Modelled%20Estimates%20for%20Small%20Areas%20projected%202015_Explanatory%20Notes%20-%20Release%201.pdf (accessed 30 April 2017).

Bartholomew, David J., Ian J. Deary, and Martin Lawn. 2009. "The origin of factor scores: Spearman, Thomson and Bartlett". British Journal of Mathematical and Statistical Psychology 62: 569–582.

Bartholomew, David J., Fiona Steele, Irini Moustaki, and Jane I. Galbraith, I. 2008. *Analysis of multivariate social science data*. New York: Chapman & Hall/CRC statistics in the social and behavioural sciences series.

Battese, George. E., Rachel M. Harter, and Wayne A. Fuller. 1988. "An Error-Components model for Prediction of County crop areas using Survey and Satellite data." Journal of the American Statistical Association 83 (401): 28-36.

Bell, Sherry M., Steve R. McCallum, Elizabeth Cox. 2003. "Toward a research-based assessment of dyslexia: Using cognitive measures to identify reading disabilities." Journal of Learning Disabilities 36(6): 505-516.

Betti Gianni, and Francesca Gagliardi. 2017. "Extension of JRR Method for Variance Estimation of Net Changes in Inequality Measures." Social Indicators Research 137(1): 45-60.

Betti Gianni, Francesca Gagliardi and Vijay Verma. 2017. "Simplified Jackknife Variance Estimates for Fuzzy Measures of Multidimensional Poverty." International Statistical Review 86(1): 68-86.

Datta, Gauri. S., Bannmo Day, and Ishwar Basawa. 1999. "Empirical best linear unbiased and empirical Bayes prediction in multivariate small area estimation." Journal of Statistical Planning and Inference 75: 269-279.

DiStefano, Christine., Min Zhu, Diana Mindrila. 2009. "Understanding and using factor scores: considerations for the applied researcher." Practical assessment, Research and Evaluation, 14(20).

EURAREA. 2004. Enhancing Small Area Estimation Techniques to meet European Needs. Deliverable D7.1.4

Eurostat. 2012. "Measuring material deprivation in the EU". Methodologies and Working papers, 2012. Retrieved February 13, 2017 (http://ec.europa.eu/eurostat/documents/3888793/5853037/KS-RA-12-018-EN.PDF/390c5677-90a6-4dc9-b972-82d589df77c2).

Ferro Luzzi, Giovanni, Yves Fluckiger, and Sylvain Weber. 2008. A cluster analysis of multidimensional poverty in Switzerland in Kakwani, Nanak. and Jacques Silber. 2008. Quantitative Approaches to Multidimensional Poverty Measurement. Palgrave Macmillan, Basingstoke.

Fuller, Wayne A., and Rachel M. Harter. 1987. The Multivariate Components of Variance Model for Small Area Estimation, in Richard, Platek, J. N. K. Rao, Carl-Erik. Sarndal, and M. P. Singh (Eds.), Small Area Statistics, New York: Wiley, 103-123.

Gasparini, Leonardo, Walter Sosa Escudero, Mariana Marchionni, and Sergio Olivieri. 2011 "Multidimensional social poverty in Latin America and the Caribbean: new evidence from the Gallup World Poll." Journal of Economic Inequality 11(2): 1-20.

Giusti, Caterina, Lucio Masserini, and Monica Pratesi. 2015. "Local Comparisons of Small Area Estimates of Poverty: An Application within the Tuscany Region of Italy." Social Indicators Research 1-20.

Giusti, Caterina, Nikos Tzavidis, Monica Pratesi, and Nicola Salvati. 2013. "Resistance to Outliers of M-Quantile and Robust Random Effects Small Area Models" Communications in Statistics – Simulations and Computations 43(3): 549-568.

González-Manteiga, Wenceslao., Maria J. Lombardía, Isabel Molina, Domingo Morales and Laureano Santamaría. 2008. "Bootstrap mean squared error of a small-area EBLUP." Journal of Statistical Computation and Simulation 78: 443–462.

Haagenars, Aldi J. M., Klaas de Vos, K. and Ashar Zaidi. 1994. Poverty Statistics in the Late 1980s: Research Based on Micro-data. Office for Official Publications of the European Communities. Luxembourg.

Hallquist, Michael and Joshua Wiley. 2014. MplusAutomation: Automating Mplus Model Estimation and Interpretation. R package version 0.6-3. http://CRAN.R-project.org/package=MplusAutomation.

Hershberger, Scott. L. 2005. Factor scores. In B. S. Everitt and D. C. Howell (Eds.) Encyclopedia of Statistics in Behavioral Science. (636-644). New York: John Wiley.

Horvitz Daniel G. and Donovan J. Thompson. 1952. "A Generalization of Sampling without Replacement from Finite Universe". Journal of the American Statistical Association 47: 663-685.

Hu, Litze., and Peter M. Bentler. 1999. "Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives." Structural Equation Modeling. 6(1): 1-55.

ISTAT. 2015. BES 2015. Il benessere equo e sostenibile in Italia. *Report* by ISTAT and CNEL in 2015. Available from http://www.istat.it/it/files/2015/12/Rapporto_BES_2015.pdf In Italian.

Johnson, Richard. A. and Dean W. Wichern. 1998. Applied Multivariate Analysis, 4th edition, Prentice.

Kawashima, Noriaki and Kunio Shiomi. 2007. "Factors of the thinking disposition of Japanese high school students." Social Behaviour and Personality, 35(2): 187-194.

Kaplan, David. 2009. Structural Equation Modeling. Foundations and Extension. Second Edition. Sage.

Krishnakumar, Jaya and Anirudh Lal Nagar. 2008. "On Exact Statistical Properties of Multidimensional Indices Based on Principal Components, Factor Analysis, MIMIC and Structural Equation Models." Social Indicators Research 86: 481-496.

Lemmi, Achille and Bruno Cheli. 2006. Fuzzy Set Approach to Multidimensional Poverty Measurement. Springer.

MacGibbon, Brenda and Tomberlin, Thomas. J. 1989. "Small Area Estimation of Proportions Via Empirical Bayes Techniques." Survey Methodology 15: 237-252.

Marchetti, Stefano, Nikos Tzavidis, and Monica Pratesi. 2012. "Non-parametric bootstrap mean squared error estimation for M-quantile estimators of small area averages, quantiles and poverty indicators." Computational Statistics and Data Analysis*,* 56: 2889–2902.

Mardia, Kanti V., John Kent, John, T. and J. M. Bibby. 1979. *Multivariate Analysis*. London: Academic Press.

Molina, Isabel, and Yolanda Marhuenda. 2015. sae: "An R Package for Small Area Estimation." The R Journal, **7**(1): 81-98.

Moretti, Angelo, Natalie Shlomo and Joseph Saukshaug. 2017. "Multivariate Small Area Estimation of Multidimensional Latent Economic Wellbeing Indicators". The Cathie Marsh Institute for Social Research working papers series.

Muthén, Bengt O. 2012. Mplus Technical Appendices. Los Angeles, CA: Muthén and Muthén.

Muthén, Linda .K. and Bengt O. Muthén. 2012. Mplus User's Guide. Seventh Edition. Los Angeles, CA: Muthén and Muthén.

Nardo, Michela, Michaela Saisana, Andrea Saltelli, Stefano Tarantola. (2005). Tools for Composite Indicators Building. European Commission, report EUR 21682 EN (Joint Research Centre, Ispra, Italy).

Norris, M. and L. Lecavalier, 2009. "Evaluating the Use of Exploratory Factor Analysis in Developmental Disability Psychological Research". Journal of Autism and Developmental Disorders. 40 (1): 8–20.

OECD-JRC. 2008. Handbook on constructing composite indicators: methodology and user guide.

OECD Statistics report.
OECD. 2013. OECD framework for statistics on the distribution of household income, consumption and wealth OECD Statistics report.
Prasad, N. G. Narasimha and J. N. K, Rao. 1990. "The Estimation of Mean Squared Error of Small-Area Estimators". Journal of the American Statistical Association 85. 163-171.
Rao, J. N. K. and Isabel Molina. 2015. *Small area estimation*. New York: Wiley.
Ravallion, Martin. 2011. "On multidimensional indices of poverty." Journal of Economic Inequality 9(2): 235-248.
Royall, Richard M. 1970. "On finite population sampling theory under certain linear regression." Biometrika 57: 377-387.
Saisana, Michaela, and Stefano Tarantola. 2002. State-of-the-art report on current methodologies and practices for composite indicator development, EUR 20408 EN.
Skrondal, Anders, and Petter Laake. 2001. "Regression among factor scores." Psychometrika 66(4): 563-575.
Smith, Tom, Michael Noble, Stefan Noble, Gemma Wright, David McLennan, and Emma Plunkett. 2015. The English Indices of Deprivation 2015. Technical Report. (Available from https://www.gov.uk/government/publications/english-indices-of-deprivation-2015-technical-report).
Yalonetzky, Gaston. 2012. "Conditions for the most robust multidimensional poverty comparisons using counting measures and ordinal variables." ECINEQ working paper, 2012-2257.