Book of the Short Papers

Editors: Francesco Maria Chelli, Mariateresa Ciommi, Salvatore Ingrassia, Francesca Mariani, Maria Cristina Recchioni

SIS Società Italiana di Statistica

UNIVERSITÀ POLITECNICA DELLE MARCHE

Dipartimento di Scienze Economiche e Sociali DISES

LIUC Università Cattaneo

LIUC Università Cattaneo | BUSINESS ANALYTICS AND DATA SCIENCE HUB

Pearson

CHAIRS

Salvatore Ingrassia (Chair of the Program Committee) - *Università degli Studi di Catania*
Maria Cristina Recchioni (Chair of the Local Organizing Committee) - *Università Politecnica delle Marche*

PROGRAM COMMITTEE

Salvatore Ingrassia (Chair), Elena Ambrosetti, Antonio Balzanella, Matilde Bini, Annalisa Busetta, Fabio Centofanti, Francesco M. Chelli, Simone Di Zio, Sabrina Giordano, Rosaria Ignaccolo, Filomena Maggino, Stefania Mignani, Lucia Paci, Monica Palma, Emilia Rocco.

LOCAL ORGANIZING COMMITTEE

Maria Cristina Recchioni (Chair), Chiara Capogrossi, Mariateresa Ciommi, Barbara Ermini, Chiara Gigliarano, Riccardo Lucchetti, Francesca Mariani, Gloria Polinesi, Giuseppe Ricciardo Lamonica, Barbara Zagaglia.

ORGANIZERS OF INVITED SESSIONS

Pierfrancesco Alaimo Di Loro, Laura Anderlucci, Luigi Augugliaro, Ilaria Benedetti, Rossella Berni, Mario Bolzan, Silvia Cagnone, Michela Cameletti, Federico Camerlenghi, Gabriella Campolo, Christian Capezza, Carlo Cavicchia, Mariateresa Ciommi, Guido Consonni, Giuseppe Ricciardo Lamonica, Regina Liu, Daniela Marella, Francesca Mariani, Matteo Mazziotta, Stefano Mazzuco, Raya Muttarak, Livia Elisa Ortensi, Edoardo Otranto, Ilaria Prosdocimi, Pasquale Sarnacchiaro, Manuela Stranges, Claudia Tarantola, Isabella Sulis, Roberta Varriale, Rosanna Verde.

FURTHER PEPOPLE OF LOCAL ORGANIZING COMMITTEE

Elisa D'Adamo, Christian Ferretti, Giada Gabbianelli, Elvina Merkaj, Luca Pedini, Alessandro Pionati, Marco Tedeschi, Francesco Valentini, Rostand Arland Yebetchou Tchounkeu

Technical support: Matteo Mercuri, Maila Ragni, Daniele Ripanti

Copyright © 2023
PUBLISHED BY PEARSON
WWW.PEARSON.COM
ISBN 9788891935618AAVV

Contents

Statistical Machine Learning for environmental applications

Statistical Process Monitoring for Complex Data in Industry 4.0

Advances in Data Science and Statistical Learning [IMS Invited Session]

ENBIS Session: System Maintenance, Boosting algorithms for regression, and Research Excellence

Educational Data mining: methods for complex data in students' assessment

Spatial and Spatio-Temporal Modeling: Theory and Applications

Statistical Framework for Measuring the Sustainability of Tourism

Statistical learning for well-being analysis

Statistics in Society 1

Assessment and Education

Bayesian methods and applications 1

Health and mortality

Mixture Models

Sampling methods and analysis of survey data

Space-time statistics

Clustering and classification 1

Dynamic models and time series

Environmental learning and indicators

Health statistics 2

Migrants in Italy and return migration

Sustainability assessment

Bayesian methods and applications 2

Clustering and classification 2

Economics and labour markets

Environmental modeling

Multivariate data analysis 2

Statistics in Society 2

Bayesian methods and applications 3

Functional Data Analysis

Machine Learning and text mining

Multivariate data analysis 3

Nonparametric statistical methods

Regression modeling

Preface

This book includes the contributions presented at the Intermediate Meeting of the Italian Statistical Society (SIS) "SIS 2023 - Statistical Learning, Sustanaibility and Impact Evolution" held in Ancona at the Università Politecnica delle Marche, from June 21th to 23th of 2023.

The new challenges of digitalization, innovation and sustainability are showing the crucial role of data-driven approaches in supporting decision-making processes. Methodologies resulting from the integration of different know-how seem to be a reliable way to deal with the increasing need to measure the impact of the policies and to forecast scenarios. This meeting welcomed any attempt to face new challenges.

The conference registered more than 250 presentations, including 3 keynote speakers in 3 plenary sessions and 72 presentations in 24 invited sessions, all dealing with specific themes in methodological and/or applied statistics and demography. Furthermore, more than 180 contributions, with one or more authors, have been spontaneously submitted to the Program Committee and arranged in 30 contributed sessions.

The numerous participation of researchers in the conference shows how the challenges of sustainability, in its broadest sense, are of interest to both methodological and applied statistics.

With the publication of this book, we wish to offer to all members of the Italian Statistical Society, all international academics, researchers, Ph.D. students, and all interested practitioners, a good snapshot of the on-going research in the statistical and demographic fields.
We aim to provide all members of the Italian Statistical Society - as well as international academics, researchers, Ph.D. students, and interested practitioners - with a comprehensive overview of the ongoing research in the fields of statistics and demography.

We extend our heartfelt gratitude to all the contributors for submitting their works to the conference and to the researchers for their outstanding job in serving as referees and discussants with precision and timeliness.

A special appreciation goes to the Scientific and Organizational Committees for their tremendous efforts in managing all the organizational aspects, as well as to the Università Politecnica delle Marche and the Department of Economic and Social Science for making this event possible.

Finally, we wish to express our gratitude to the publisher Pearson Italia for all the support received.

Assessing multidimensional poverty of the Italian provinces during Covid-19: a small area estimation approach

Mariateresa Ciommi[a], Chiara Gigliarano[b], Francesca Mariani[a], and Gloria Polinesi[a]

[a]Università Politecnica delle Marche; m.ciommi@staff.univpm.it, f.mariani@staff.univpm.it, g.polinesi@staff.univpm.it
[b]LIUC - Università Carlo Cattaneo; cgigliarano@liuc.it

Abstract

The aim of this paper is to analyse the effect of Covid-19 on multidimensional poverty in Italy and its provinces by measuring changes in individual poverty before and during the pandemic outbreak. To capture the multidimensional nature of poverty, we consider different dimensions: economic well-being, health condition, education, neighborhood quality, subjective well-being.

The empirical application is based on micro-data from the 'Aspects of daily life' (AVQ) survey by Istat for the years 2019 (pre-Covid period) and 2020 (Covid period).

Since survey direct estimates are reliable only at regional (NUTS 2) level, the introduction of small area estimation (SAE) techniques becomes of crucial importance to monitor and contrast the phenomenon at a finer geographical level.

Keywords: Multidimensional poverty index, composite indicators, small area estimation, EBLUP estimator.

1. Introduction

In the recent years there has been a considerable agreement that poverty is a multidimensional phenomenon, which cannot be adequately explained by considering only monetary variables. (1) confirms the existence of significant mismatches between monetary poverty and multiple deprivations and, therefore, the need to include many aspects of poverty, such as health, education, housing, satisfaction with life, security.

Few empirical studies have analyzed multidimensional poverty in Italy at sub-national level; see, among others, (5), (4) and (3). In this paper, we propose a deeper attempt in this direction by looking at the provincial level changes in the multidimensional poverty due to Covid-19 pandemic.

The empirical application is based on micro-data from the 'Aspects of daily life' (AVQ) Italian survey over the years 2019 (pre-Covid period) and 2020 (Covid period).

Since the AVQ survey is planned to obtain precise estimates at regional level, some alternative solutions should be evaluated to estimate poverty at provincial level. More specifically, two main possible strategies can be employed: i) increasing the sample size of AVQ for the specific province of interest (oversampling) so that direct estimates become reliable and ii) apply small area estimation (SAE) techniques; see, among others, (12) and (9).

We apply the Fay-Herriot (FH) small area estimation model proposed by (6) to estimate dashboard indicators at provincial level. Then, we propose an aggregation technique based on the Adjusted Mazziotta-Pareto Index (see (7) and (8)) for the FH estimates in order to obtain reliable estimates of the multidimensional poverty for the local areas of interest as in (11).

Therefore, the aim is to enhance the knowledge of the spatial distribution of the multidimensional poverty index at local level in Italy, focusing on Italian provinces, in order to help the policy maker to address resources towards the areas where the phenomenon is strongly present.

The empirical analysis reveals that the traditional Italian divide North vs. South becomes less clear when analyzing poverty at provincial level and for the different domains of poverty.

The remainder of the paper is as follows. Sections 2. and 3. describe data and methodologies considered in the analysis. Section 4. is devoted to results of the empirical application. Finally, Section 5. draws some conclusions.

2. Data

Our analysis is based on data from the 'Aspects of daily life' (AVQ) survey conducted by Istat, focusing on the years 2019 and 2020. The units of analysis are the Italian households and the original samples include 19536 and 18529 households, for the two years considered respectively.

To assess the multidimensional poverty over 105 Italian provinces[1], we consider five domains (health, education, economic well-being, neighborhood quality and subjective well-being), each composed of different dashboard indicators as described in Table 1.

In order to classify a household as deprived in each of the 13 dashboard indicators, we first identify the deprivation cut-offs following the approach proposed in (5), which we have slightly modified due to the data availability at provincial level. The deprivation cut-offs refers to the head of the household.

Table 1: Multidimensional poverty framework: domains, dashboard indicators, deprivation cut-offs.

Domains	**Deprivation Indicators**	**Deprivation cut-offs**
Health	Nutrition	A person is deprived if s/he consumes less than 3 portions of fruits or vegetables a day.
Education	Educational deprivation	A person is deprived if s/he has not completed higher-secondary school.
	Cultural deprivation	A person is deprived if in the 12 months before the interview s/he has joined less than 2 among the following activities: 1) at least once to cinema, theatre, exhibitions and museums, archaeological sites, monuments, concerts of classical music, opera, concerts of other kind of music; 2) read the newspaper at least once a week; 3) read at least a book.
Economic well-being	Material deprivation	A person is deprived if s/he possesses less than 4 out of 6 following items: washing machine, color tv, scooter/moto or car, phone, personal computer.
	Housing deprivation	A person is deprived if s/he experiences 3 or more among the following deprivations related to the house: overcrowding; distance from basic services (pharmacy, shops, school); overall poor condition of the floors and/or walls; expenses too high; house not owned).
	Gas	A person is deprived if the house is not served by methane gas.
	Water	A person is deprived if s/he declares to have irregularities in their water supply.
	Unemployment	A person is deprived if s/he is unemployed.
	Financial distress	A person is deprived if her/his economic sources are not sufficient to make ends meet.
Neighbourhood quality	Noise Crime Pollution	A person is deprived if the area in which s/he lives is declared to be very noisy; at risk of crime; polluted.
Subjective well-being	Life satisfaction and future expectations	A person is deprived if experiences 3 or more deprivations related to personal satisfaction (life, economic situation, health, familiar and friends relationship, leisure and future expectations).

[1]The provinces of Vibo-Valentia and Benevento have been excluded from the analysis due the very high percentage of missing data.

3. Methodology

Small area estimation (SAE) combines survey data with auxiliary variables of the population of interest to break down regional estimates into sub-regional ones. These variables are commonly obtained from population censuses or from administrative registers. If auxiliary information are available at unit-level (e.g., individual or household-level) one can consider unit-level SAE models[2]. When, on the contrary, the auxiliary data are available only at area-level (e.g., district, municipality or provincial level), one can use area-level SAE models.

We follow the latter approach and apply the Fay-Herriot area-level small area estimation model introduced by (6), which proposes the empirical best linear unbiased predictor (EBLUP) estimator to obtain estimates of the poverty dashboard indicators at provincial level[3].

Consider a finite population U, partitioned into $d = 1, \ldots, D$ mutually exclusive and exhaustive areas (in our case, provinces); the Fay-Herriot (FH) model is defined in two stages.

Let $\hat{\delta}_d^{DIR}$ be a direct estimator of δ_d, the parameter of inferential interest for area (province) d. In the first stage, we assume that $\hat{\delta}_d^{DIR}$ is an unbiased estimator of δ_d:

$$\hat{\delta}_d^{DIR} = \delta_d + e_d, \qquad e_d \overset{ind}{\sim} N(0, \psi_d), \tag{1}$$

where ψ_d is the sampling variance of the direct estimator $\hat{\delta}_d^{DIR}$ given δ_d, assumed to be known for all $d = 1, \ldots, D$.

In the second stage, we assume that the area parameters δ_d are linearly related with a p-vector x_d of area-level auxiliary variables as follows:

$$\delta_d = x_d^T \beta + u_d, \qquad u_d \overset{ind}{\sim} N(0, A). \tag{2}$$

Model (1) is known as *sampling model* because it represents the uncertainty due to the fact that δ_d is unobservable and the direct estimator is based on the sample data, $\hat{\delta}_d^{DIR}$. While, model (2) is called *linking model* because it relates all areas through the common regression coefficients β, allowing us to borrow strength from all areas.

Combining the two model components (1) and (2), we obtain the linear mixed model:

$$\hat{\delta}_d^{DIR} = x_d^T \beta + u_d + e_d, \tag{3}$$

where u_d is independent of e_d.

The EBLUP estimator is the combination of the direct and the regression-synthetic estimators:

$$\hat{\delta}_d^{EBLUP} = \hat{\gamma}_d \hat{\delta}_d^{DIR} + (1 - \hat{\gamma}_d) x_d^T \hat{\beta} \tag{4}$$

where $\hat{\gamma} = \frac{\hat{A}}{\hat{A} + \hat{\psi}_d}$ represents the shrinkage factor.

For each of the 13 dashboard indicators of deprivation j and for each province d we obtain an EBLUP estimate, denoted with $\hat{\delta}_{dj}^{EBLUP}$. For sake of simplicity we will denote estimator $\hat{\delta}_{dj}^{EBLUP}$ with $\hat{\delta}_{dj}$.

Then, in order to obtain a composite indicator of multidimensional poverty, we aggregate the EBLUP provincial estimates $\hat{\delta}_{dj}$, where d indicates the province and j the dashboard indicator, using the Adjusted Mazziotta-Pareto Index (AMPI); see (8).

The AMPI approach considers the normalized provincial estimates r_{dj} defined as:

$$r_{dj} = \frac{\hat{\delta}_{dj} - Min(\hat{\delta}_{dj})}{Max(\hat{\delta}_{dj}) - Min(\hat{\delta}_{dj})} 60 + 70. \tag{5}$$

Denoting with Ref_{δ_j} the reference value for the indicator j, the 'goalposts' $Max(\hat{\delta}_{dj})$ and $Min(\hat{\delta}_{dj})$ are defined as:

$$\begin{cases} Min(\hat{\delta}_{dj}) = Ref_{\delta_j} - \Delta \\ Max(\hat{\delta}_{dj}) = Ref_{\delta_j} + \Delta \end{cases} \tag{6}$$

[2]The Battese-Harter-Fuller (BHF) model is one of the most used unit-level small area estimation model (2).

[3]See (10) for notation details.

where $\Delta = \frac{Sup_{\delta_j} - Inf_{\delta_j}}{2}$. Inf_{δ_j} and Sup_{δ_j} are the overall minimum and maximum of the indicator $\hat{\delta}_{dj}$ across all provinces and the two years considered. To facilitate the interpretation of results, we choose as Ref_{δ_j} the mean of the year 2019. The normalized values r_{dj} will fall approximately in the range $(70; 130)$, where 100 represents the reference value.

Denoting with M_d and S_d , respectively, the mean and standard deviation of the normalized indicators r_{dj} of province d, the multidimensional poverty index of province d is computed as:

$$AMPI_d = M_d + S_d \times cv_d \tag{7}$$

where $cv_d = \frac{S_d}{M_d}$ is the coefficient of variation for the local unit d; see (8).

4. Empirical findings

Figures 1 and 2 illustrate a geographical representations of how the multidimensional poverty indicator as well as the domain-specific deprivation indices are distributed across the Italian provinces in the years 2019 and 2020, respectively. In each panel, value 100 corresponds to the Italian average in 2019, and darker colors indicate higher levels of deprivation.

Comparing Figures 1(d) and 2(d) it emerges how health deprivation strongly worsened during the Covid-19 pandemic especially in Sicily, Sardinia, Lombardy and Piedmont and for the provinces located along the Tyrrhenian coast. On the contrary, Figures 1(e) and 2(e) reveal that life satisfaction and future expectations have not been affected by Covid-19.

Overall, the multidimensional poverty index (Figures 1(a) vs. 2(a)) does not seem to have deteriorated after the first year of the pandemic, although its levels remain higher in the South of Italy. We note that some domains of poverty (economic, education, health) have worsened in many provinces, but their effects have been compensated by the opposite trend registered for subjective well-being and quality of neighborhood. Anyway, the result refers only to the first year of the pandemic. In order to have a complete overview of the effect of Covid-19, one should analyze more recent data.

5. Conclusion

In this paper we have estimated the multidimensional poverty of the Italian households during the period of Covid-19 at provincial level. Since the AVQ survey is planned to obtain precise estimates at regional level, we have applied the FH small area estimation model in order to obtain reliable estimates of the dashboard indicators of deprivation at this finer geographical level. The overall composite indicator of multidimensional poverty has been then obtained using the AMPI approach.

Empirical findings suggest that, among the different domains of poverty considered, health deprivation has strongly worsened during the first year of the pandemic especially for provinces of the Islands (Sicily and Sardinia), Lombardy, Piedmont and the whole region located along the Tyrrhenian coast. Subjective deprivation, on the contrary, has reduced in the period of analysis. Overall, data reveals that the multidimensional poverty has not worsened relevantly, although the provinces of southern Italy are multidimensionally poorer than those of the north.

Further research could be devoted to the use of different techniques of aggregation of the dashboard indicators.

Acknowledgments

This research has been founded by Fondazione Cariplo, under the project *POST-COVID: POverty and vulnerability Scenarios in The era of Covid-19: how the pandemic is affecting the well-being of the Italians* - rif. 2020-4216.

The data used in this paper are from Istat's Aspects of Daily Life ('Aspetti di vita quotidiana') survey. The elaborations were carried out at Istat's ADELE ('Analisi dei Dati ELEmentari') laboratory

and in compliance with the legislation on protection of statistical confidentiality and personal data. The responsibility for results and opinions expressed in the paper rests solely with the authors. The analyses have been carried out without sampling weights.

References

[1] Atkinson, A. B.: Multidimensional deprivation: contrasting social welfare and counting approaches. The Journal of Economic Inequality **1 (1)**, 5–65 (2003)

[2] Battese, G. E., Harter, R. M., Fuller, W. A.: 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 (1988)

[3] Betti, G., Cheli, B., Lemmi, A., Verma, V.: The fuzzy set approach to multidimensional poverty: the case of Italy in the 1990s. In: Kakwani, N., Silber, J. (eds). Quantitative Approaches to Multidimensional Poverty Measurement. Palgrave Macmillan, London (2008).

[4] Coromaldi, M., Zoli, M.: Deriving multidimensional poverty indicators: Methodological issues and an empirical analysis for Italy. Social Indicators Research **107**, 37–54 (2012)

[5] De Rosa, D.: Are Italians getting multidimensionally poorer? Evidence on the lack of equitable and sustainable well-being. Italian Economic Journal **8 (1)**, 145–174 (2022)

[6] Fay, R. E., Herriot, R. A.: Estimation of income from small places: an application of James-Stein procedures to census data. Journal of the American Statistical Association **74 (366)**, 269–277 (1979)

[7] Mazziotta, M., Pareto, A.: On a generalized non-compensatory composite index for measuring socio-economic phenomena. Social Indicators Research **127 (3)**, 983–1003 (2016)

[8] Mazziotta, M., Pareto, A.: Measuring well-being over time: the adjusted Mazziotta-Pareto index versus other non-compensatory indices. Social Indicators Research **136 (3)**, 967–976 (2018)

[9] Molina, I., Rao, J. N. K.: Small area estimation of poverty indicators. The Canadian Journal of Statistics **38 (3)**, 369–385 (2010)

[10] Molina, I., Marhuenda, Y.: Sae: an R package for small area estimation. The R Journal **7 (1)**, 81 (2015)

[11] Pratesi, M., Quattrociocchi, L., Bertarelli, G., Gemignani, A., Giusti, C.: Spatial distribution of multidimensional educational poverty in Italy using small area estimation. Social Indicators Research **156**, 563–586 (2021)

[12] Rao, J. N. K.: Some new developments in small area estimation. JIRSS (2003)