

Integration of socio-economic data for the estimation of indicators at municipal level

G. Bertarelli¹ S. Casacci² M. D'Alò² D. Ercolani²
A. Guandalini² A. Fasulo² M. G. Ranalli³ F. Solari²

¹University of Pisa – ²ISTAT – ³Univeristy of Perugia

SIS 2018, Palermo, June 22nd, 2018

The presentation at a Glance

LabInn

Introduction

Data

Area-level models

Conclusions and Future Work

Why are we together?



Introduction

- ★ Provide a statistical tool that can drive local policies on the basis of urban specificities.
- ★ Very **detailed** and **updated** statistical information at finer geographic level is necessary.
- ★ **Census** ensures accurate information but its temporal discrepancy has been a big limit.
- ★ **Sample surveys** ensure information on annual base but have limitations at finer geographical resolution.
- ★ Need to provide solutions that exploit the availability of new sources of information.

Administrative data

- ★ The integration of this information with survey data can overcome the lack of information at a more detailed territorial level: timely and accurate estimates.
- ★ ISTAT produces indicators using **ARCHIMEDE** (*Integrated ARCHives of socio-Economic and DEmographic Microdata*) at municipal level.
- ★ **Issue**: due to a different taxonomy they do not match the estimates of the indicators obtained using sample surveys.

Aim of the presentation

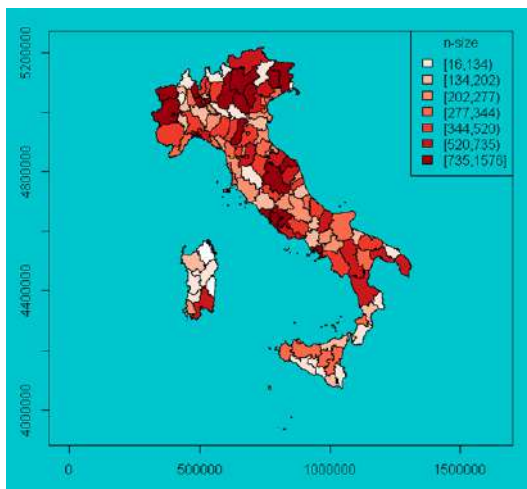
Provide first results of the following analyses

- ★ Comparison the indicators computed by the two sources of information for
 - all the metropolitan cities (14)
 - and for provincial areas (excluding metropolitan cities) (110)
- ★ Introduction of administrative data as proxy information in area-level SAE model.

Data: ARCHIMEDE & EU-SILC

- ★ ARCHIMEDE 2015 (Garofalo, 2014; Wallgren and Wallgren, 2007): Integrated archives of socio-economic and demographic microdata
 - The administrative sources used for constructing the collection of microdata are:
 1. Municipal Population Registers
 2. Tax Returns Register
 3. Central Register of Pensioners
 4. Social Security and Fiscal sources (workers)
 5. Social Security Benefits registers
 6. Population Census
 - ≈ 60 million individuals (≈ 24 million households)
- ★ EU-SILC 2016 (EUropean Survey on Income and Living Conditions)(Ceccarelli et al., 2008)

EU-SILC sample sizes



Response Variables

- ★ **At risk of poverty rate**: share of people with an equivalised disposable income below the at the risk of poverty threshold, which is set at 60% of the national median equivalised disposable income;
- ★ **Severe material deprivation rate**: share of population living in households lacking at least 4 items out of the following 9:
 - to pay rent or utility bills,
 - keep home adequately warm,
 - face unexpected expenses,
 - eat meat or fish every second day,
 - a week holidays away from home,
 - a car,
 - a washing machine,
 - a colour TV,
 - a telephone.

Response Variables

- ★ **At risk of poverty rate**: share of people with an equivalised disposable income below the at the risk of poverty threshold, which is set at 60% of the national median equivalised disposable income;
- ★ **Severe material deprivation rate**: share of population living in households lacking at least 4 items out of the following 9:
 - to pay rent or utility bills,
 - keep home adequately warm,
 - face unexpected expenses,
 - eat meat or fish every second day,
 - a week holidays away from home,
 - a car,
 - a washing machine,
 - a colour TV,
 - a telephone.

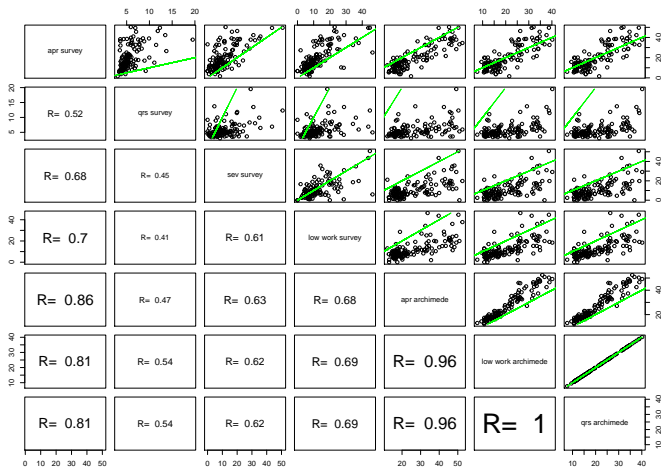
Response Variables ctd.

- ★ **Low work intensity**: Proportion of people living in households with very low work intensity namely household members of working age that have worked during the income reference year less than 20% of the number of months that could theoretically have been worked by the same household members;
- ★ **Income Inequality**: Ratio of total equivalised income received by the 20% of the population with the highest income to that received by the 20% of the population with the lowest income.

Response Variables ctd.

- ★ **Low work intensity**: Proportion of people living in households with very low work intensity namely household members of working age that have worked during the income reference year less than 20% of the number of months that could theoretically have been worked by the same household members;
- ★ **Income Inequality**: Ratio of total equivalised income received by the 20% of the population with the highest income to that received by the 20% of the population with the lowest income.

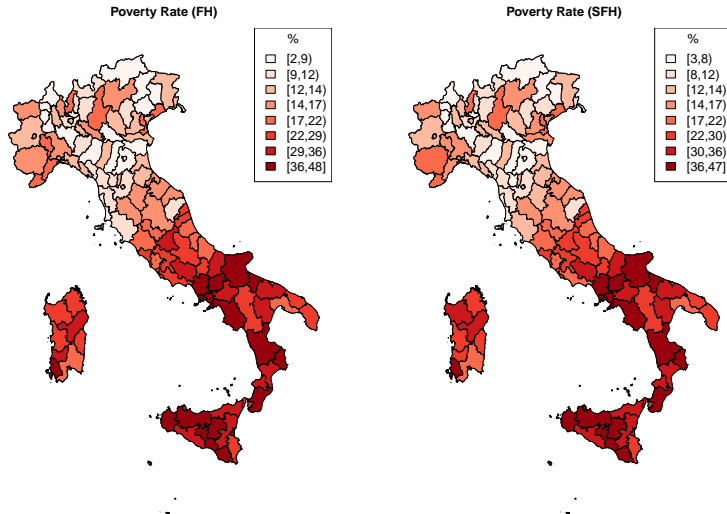
Correlation Plot (Direct Est. vs. Register)



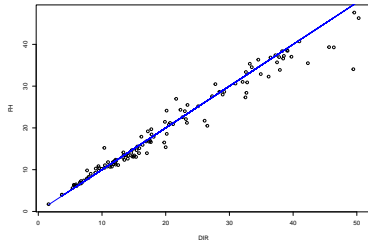
Area-level SAE

- ★ Compare results from:
 - Fay-Herriot model (Fay and Herriot, 1979);
 - Spatial Fay-Herriot Model (Molina et al., 2009; Pratesi et al., 2009);
- ★ Possible covariates:
 - Proxy from ARCHIMEDE for
 - at risk of poverty rate,
 - low work intensity,
 - Income Inequality;
 - Percentage of the population that
 - is foreigner,
 - is employed;
 - belongs to 14 sex × ages classes.

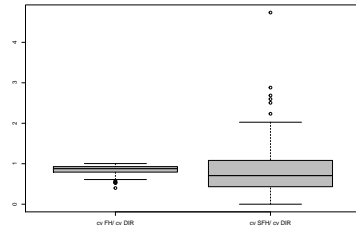
At risk of Poverty Rate



At risk of Poverty Rate



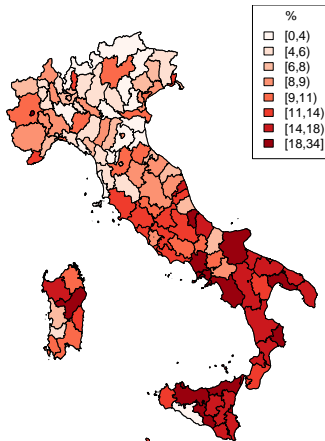
(a) FH vs DIR



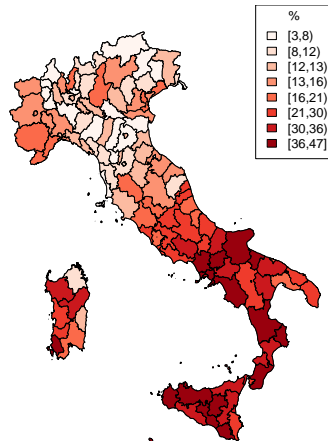
(b) CV ratios: FH-CV/DIR-CV (left), SFH-CV/DIR-CV (right)

Severe Material Deprivation Rate

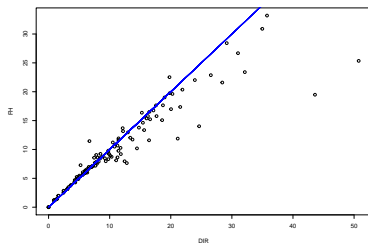
Sev. Material Deprivation Rate (FH)



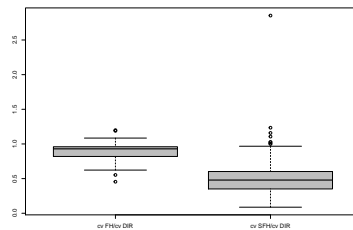
Sev. Material Deprivation Rate (SFH)



Severe Material Deprivation Rate

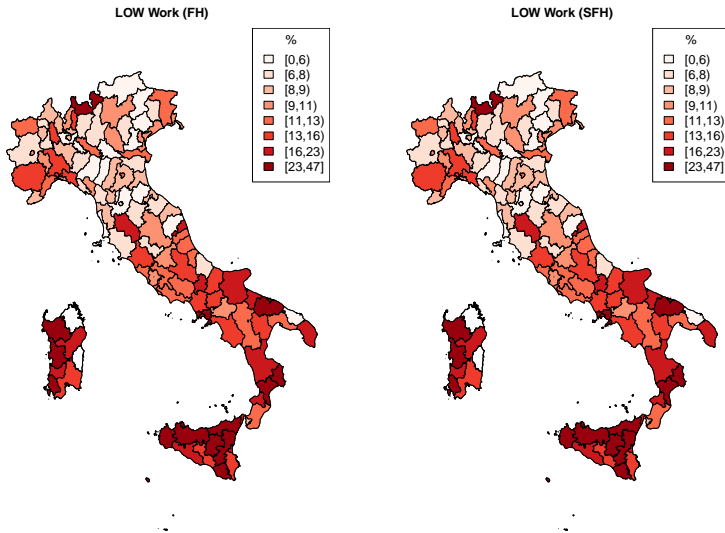


(c) FH vs DIR

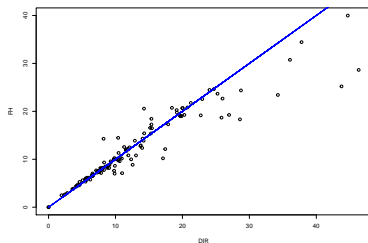


(d) CV ratios: FH-CV/DIR-CV (left), SFH-CV/DIR-CV (right)

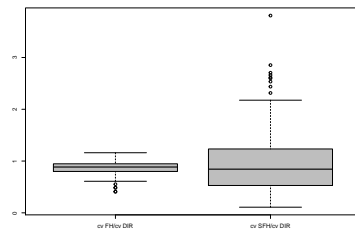
Low Work Intensity



Low Work Intensity

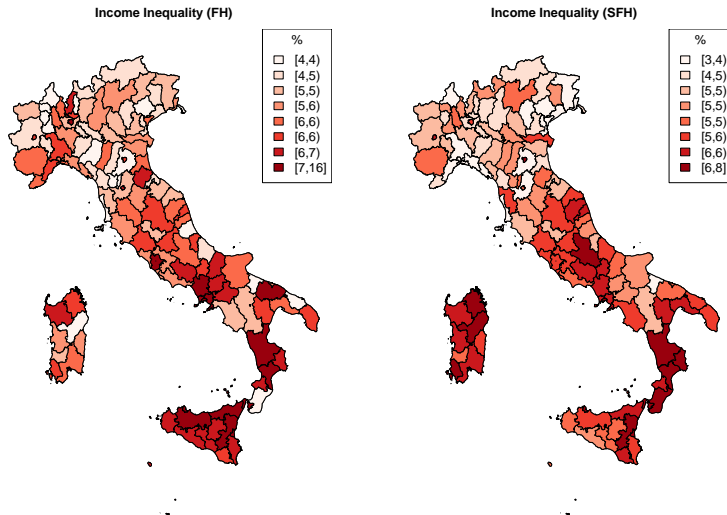


(e) FH vs DIR

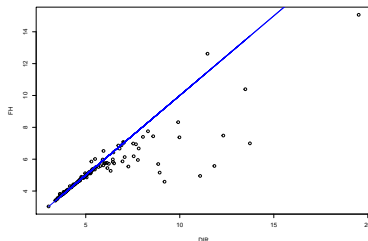


(f) CV ratios: FH-CV/DIR-CV (left), SFH-CV/DIR-CV (right)

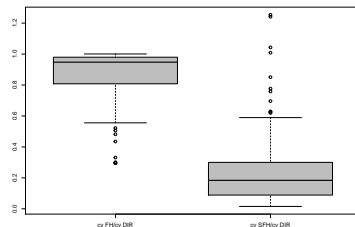
Income Inequality Rate



Income Inequality Rate



(g) FH vs DIR



(h) CV ratios: FH-CV/DIR-CV (left), SFH-CV/DIR-CV (right)

Conclusions and Future Work

- ★ Administrative Data is an important tool to improve efficiency in SAE
- ★ Incorporating the spatial structure provides larger gains in efficiency for most indicators
- ★ Move to multivariate FH models
- ★ Move to area-level models with categorical latent variables (univariate and multivariate, Bertarelli et al., 2018)
- ★ Move to unit-level models
- ★ Move to projection estimation (Kim and Rao, 2011) from EU-SILC 2016 to Register and from EU-SILC 2016 to Labour Force Survey.

Conclusions and Future Work

- ★ Administrative Data is an important tool to improve efficiency in SAE
- ★ Incorporating the spatial structure provides larger gains in efficiency for most indicators
- ★ Move to multivariate FH models
- ★ Move to area-level models with categorical latent variables (univariate and multivariate, Bertarelli et al., 2018)
- ★ Move to unit-level models
- ★ Move to projection estimation (Kim and Rao, 2011) from EU-SILC 2016 to Register and from EU-SILC 2016 to Labour Force Survey.

Thank You!!!!

My work is supported from

InGRID-2

Integrating Research Infrastructures from expertise on Inclusive growth from data to policy

References I

- Ceccarelli, C., M. Di Marco, and C. Rinaldelli (2008). L'indagine europea sui redditi e le condizioni di vita delle famiglie (eu-silc). *Metodi e Norme*.
- Fay, R. E. and R. A. Herriot (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association* 74, 269–277.
- Garofalo, G. (2014). Il progetto archimede obiettivi e risultati sperimentali. Technical report, ISTAT Working papers, 9.
- Kim, J. K. and J. N. Rao (2011). Combining data from two independent surveys: a model-assisted approach. *Biometrika* 99(1), 85–100.
- Molina, I., N. Salvati, and M. Pratesi (2009). Bootstrap for estimating the mse of the spatial eblup. *Computational Statistics* 24(3), 441–458.
- Pratesi, M., N. Salvati, et al. (2009). Small area estimation in the presence of correlated random area effects. *Journal of Official Statistics* 25(1), 37.
- Wallgren, A. and B. Wallgren (2007). *Register-based statistics: administrative data for statistical purposes*, Volume 553. John Wiley & Sons.