	Area-level models	

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

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

SIS 2018, Palermo, June 22nd, 2018

	Area-level models	

< □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □

The presentation at a Glance

LabInn

Introduction

Data

Area-level models

Conclusions and Future Work

Labinn		
•		

Why are we together?



◆□▶ ◆□▶ ◆国▶ ◆国▶ ◆□▶

<u>Bertarelli,</u> Casacci, D'Alò, Ercolani, Guandalini, Fasulo, Ranalli, Solari

	Introduction ●○○		
Introd	uction		

- ★ 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.
- $\star\,$ Need to provide solutions that exploit the availability of new sources of information.

◆□▶ ◆母▶ ◆ヨ▶ ◆ヨ▶ ヨヨ シタペ

Introduction		
000		

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.

< □ > < 同 > < 三 > < 三 > < 三 > < □ > < □ > < ○ </p>

Introduction		
000		

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	
	00000	

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:

< □ > < 同 > < 三 > < 三 > < 三 > < □ > < □ > < ○ </p>

- 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)

	Data	
	00000	

EU-SILC sample sizes



· • ㅁ • • @ • • 토 • • 토 • 토 = · 이익은

Bertarelli, Casacci, D'Alò, Ercolani, Guandalini, Fasulo, Ranalli, Solari

	Data 00●00	

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;

< □ > < 同 > < 三 > < 三 > < 三 > < □ > < □ > < ○ </p>

- * 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 meet or fish every second day,
 - a week holidays away from home,
 - a car,
 - a washing machine,
 - a colour TV,
 - a telephone.

	Data 00●00	
_		

★ 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;

< □ > < 同 > < 三 > < 三 > < 三 > < □ > < □ > < ○ </p>

- ★ 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 meet or fish every second day,
 - a week holidays away from home,
 - a car,

Response Variables

- a washing machine,
- a colour TV,
- a telephone.

	Data	
	00000	

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.

	Data	
	00000	

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.

	Data	
	00000	

Correlation Plot (Direct Est. vs. Register)



Bertarelli, Casacci, D'Alò, Ercolani, Guandalini, Fasulo, Ranalli, Solari

	Area-level models ●00000000	

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);

< □ > < 同 > < 三 > < 三 > < 三 > < □ > < □ > < ○ </p>

★ 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;
 - $^\circ$ belongs to 14 sex imes ages classes.

Bertarelli, Casacci, D'Alò, Ercolani, Guandalini, Fasulo, Ranalli, Solari

	Area-level models	
	00000000	

At risk of Poverty Rate

Poverty Rate (FH)







	Area-level models	
	00000000	

At risk of Poverty Rate



(a) FH vs DIR

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

< ロ > < 同 > < 回 > < 回 > < 回 > <

EL SQA

Bertarelli, Casacci, D'Alò, Ercolani, Guandalini, Fasulo, Ranalli, Solari

	Area-level models	
	00000000	

Severe Material Deprivation Rate

Sev. Material Deprivation Rate (FH)



Sev. Material Deprivation Rate (SFH)



	Area-level models	
	00000000	

Severe Material Deprivation Rate





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

< 日 > < 同 > < 回 > < 回 > < 回 > :

EL SQA

Bertarelli, Casacci, D'Alò, Ercolani, Guandalini, Fasulo, Ranalli, Solari

	Area-level models	
	00000000	

Low Work Intensity

LOW Work (FH)





	Area-level models	
	000000000	

Low Work Intensity



(e) FH vs DIR

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

Bertarelli, Casacci, D'Alò, Ercolani, Guandalini, Fasulo, Ranalli, Solari

	Area-level models	
	000000000	

Income Inequality (SFH)

= nac

Income Inequality Rate

Income Inequality (FH)



	Area-level models	
	00000000	

Income Inequality Rate



(g) FH vs DIR

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

Bertarelli, Casacci, D'Alò, Ercolani, Guandalini, Fasulo, Ranalli, Solari

	Area-level models	Conclusions and Future Work
		00

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
		0

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.

< □ > < 同 > < 三 > < 三 > < 三 > < □ > < □ > < ○ </p>

	Area-level models	Conclusions and Future Work
		00

Thank You!!!!

My work is supported from InGRID-2

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

◆□▶ ◆母▶ ◆ヨ▶ ◆ヨ▶ ヨヨ シタペ

References |

- 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.