PISA, 8-10 MAY 2018

Multivariate Small Area Estimation of Multidimensional Economic Well-being Indicators: a focus on Housing Quality

Angelo Moretti[12], Natalie Shlomo[1] and Joseph Sakshaug[3]

[1]**Social Statistics Department, University of Manchester – United Kingdom**
[2]**Geography Department, University of Sheffield – United Kingdom**
[3]**German Institute for Employment Research, Nuremberg, Germany.**

FINAL EVENT OF THE JEAN MONNET CHAIR SAMPLEU

Outline

1. Data dimensionality reduction

2. Small area estimation via multivariate EBLUP

3. Simulation study

4. Application based on EU-SILC data

5. Future work

Data dimensionality reduction

- Multivariate statistical analysis techniques
  - Depending on variable types: continuous, categorical, ordinal
- The problem of **weighting and aggregation** (OECD-JRC, 2008)
  - Big debate on the methodology
  - Statistical models taking into account for the **degree of correlation** among the observed variables (dashboard of indicators)
  - Different type of variables
- **Factor analysis models – the use of the factor scores**
  - Possible for any type of variables (transformations)
  - Good rankings, smaller variability compared to the traditional weighted averages
  - FA model variability needs to be take into account in MSE estimation (Moretti, Shlomo and Sakshaug, 2017a)

<u>The aim of this paper</u>

- Evaluate **EBLUP approach** under BHF model (Battese et al. 1988) and **<u>multivariate EBLUP</u>** (MEBLUP) under Fuller and Harter model (Fuller and Harter, 1987) in data dimensionality reduction for latent well-being indicators
- Extension of Moretti, Shlomo and Sakshaug (2017a): univariate latent well-being
- Case of observed indicator <u>a priori</u> selected and studied in well-being frameworks
- The case of **averages of standardised EBLUPs** compared to the EBLUP of **factor scores means (multivariate and univariate case)**
- Simulation study and application using European Union Statistics on Income and Living Conditions (EU-SILC) Italian data.

Multivariate EBLUP approach

Target parameter: a vector of K means for small area d:

$$\bar{\boldsymbol{Y}}_d = N_d^{-1}\sum_{i=1}^{N_d}\boldsymbol{y}_{di},$$

$$\bar{\boldsymbol{Y}}_d = N_d^{-1}\left(\sum_{i\in s_d}\boldsymbol{y}_{di} + \sum_{i\in r_d}\boldsymbol{y}_{di}\right)$$

Multivariate nested-error model (Fuller and Harter, 1987)

$$\boldsymbol{y}_{di} = \boldsymbol{x}_{di}\boldsymbol{\beta} + \boldsymbol{u}_d + \boldsymbol{e}_{di}, d = 1, \ldots, D, i = 1, \ldots, N_d,$$

$$\boldsymbol{u}_d \sim N_K(\boldsymbol{0}, \boldsymbol{\Sigma}_u), \qquad \boldsymbol{e}_{di} \sim N_K(\boldsymbol{0}, \boldsymbol{\Sigma}_e)\ \ \boldsymbol{u}_d \textit{ and } \boldsymbol{e}_{di} \textit{ are independent}$$

Model parameters estimated via ML (Datta et al. 1999)

Prediction (Fuller and Harter, 1987; Datta et al. 1999)

$$\hat{\bar{\boldsymbol{Y}}}_d^{MEBLUP} = \bar{\boldsymbol{X}}'_{dp}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{u}}_d = \bar{\boldsymbol{X}}'_{dp}\hat{\boldsymbol{\beta}} + (\bar{\boldsymbol{y}}_d - \bar{\boldsymbol{x}}'_d\hat{\boldsymbol{\beta}})[(\hat{\boldsymbol{\Sigma}}_u + n_d^{-1}\hat{\boldsymbol{\Sigma}}_e)^{-1}\hat{\boldsymbol{\Sigma}}_u], d = 1, \ldots, D$$

IN SUPREMÆ DIGNITATIS 1343

Centro Camilo Dagum
Tuscan Interuniversity Centre
Advanced Statistics for Equitable and Sustainable Development

MAKSWELL
MAKing Sustainable development and WELL-being frameworks work for policy analysis

Small Area Methods for Monitoring of Poverty and Living Conditions in EU
Jean Monnet Chair

Simulation study (1)

Generating the population

- From the multivariate nested-error model $\boldsymbol{y}_{di}$ with *k=1,…4* are generated with:

$$N = 20{,}000, \quad D = 80, \quad 130 \leq N_d \leq 420,$$

$$N_d \sim \mathcal{U}(a = 130, b = 420)$$

$$X_1 \sim \mathcal{U}(a = 145, b = 459)$$

$$X_2 \sim \mathcal{U}(a = 55, b = 345)$$

- Different correlation structure in $\boldsymbol{\Sigma}_e$ *and* $\boldsymbol{\Sigma}_u$:

  $r_u = r_e = 0.2,$

  $r_u = 0.2$ and $r_e = 0.7,$

  $r_u = -0.2$ and $r_e = 0.7$

- Intra-class correlation: $ICC_k = \{0.05,0.1,0.3\}$ so we generate $\boldsymbol{\Sigma}_u$ matrices as functions of *ICC*.

Centro Camilo Dagum
Tuscan Interuniversity Centre
Advanced Statistics for Equitable and Sustainable Development

MAKSWELL
MAKing Sustainable development and WELL-being frameworks work for policy analysis

SMALL AREA METHODS FOR MONITORING OF POVERTY AND LIVING CONDITIONS IN EU
Jean Monnet Chair

Simulation study (2)

Simulation steps

1. Draw $S = 500$ samples using simple random sampling without replacement n=1000: allowing for **unplanned domains**

2. Estimate a *one-factor*, and *two-factors analysis* models and estimate the EBLUP and MEBLUP **factor score means** for each area d in each sample

3. EBLUP and MEBLUP on each of the original observed variables y and vectors $\boldsymbol{y}$ are also estimated: construct **simple averages** of the standardized small area EBLUPs and MEBLUPs' components and **weighted averages** using the factor loadings

4. Results are evaluated via RMSE, percentages of reduction in terms of RMSE, and relative bias.

Simulation study (3)

Case two-factor only

Table 1: Percentage relative reduction (%) in terms of RMSE of simple and weighted averages of standardised MEBLUP over EBLUP , two-factor CFA model .

		Scenario					
ICC_k		$r_e = 0.7, r_u = 0.2$		$r_e = 0.7, r_u = -0.2$		$r_e = 0.2, r_u = 0.2$	
		Simple	*Weighted*	*Simple*	*Weighted*	*Simple*	*Weighted*
0.05	*Factor 1*	-5.26	-5.29	-5.56	-6.52	-2.86	-2.94
	Factor 2	-3.90	-4.74	-20.99	-21.82	-0.43	-0.22
0.1	*Factor 1*	-8.89	-9.30	-26.67	-22.73	-0.50	-0.25
	Factor 2	-9.00	-11.08	-25.53	-26.67	0.00	0.00
0.3	*Factor 1*	-11.67	-11.88	-16.67	-17.92	-0.28	-0.29
	Factor 2	-13.89	-12.86	-19.87	-17.05	0.00	0.00

Simulation study (4)

Table 2: Percentage relative reduction (%) in terms of RMSE of MEBLUP over EBLUP, factor scores, two-factor CFA model

		Scenario		
ICC_k		$r_e = 0.7, r_u = 0.2$	$r_e = 0.7, r_u = -0.2$	$r_e = 0.2, r_u = 0.2$
0.05	Factor scores 1	-2.44	-2.50	0.00
	Factor scores 2	-2.50	-2.56	0.00
0.1	Factor scores 1	-2.56	-3.13	0.00
	Factor scores 2	-3.33	-2.86	0.00
0.3	Factor scores 1	-4.48	-5.56	0.00
	Factor scores 2	-5.56	-6.67	0.00

- The factor analysis model modified the correlation structure therefore percentage relative reductions change considerably

Application (1)

- Housing quality: dimension in multidimensional well-being Equitable and Sustainable Well-being **BES Italian framework**
- Determinant of well-being (Andrews et al. 2011)
- Data: **European Union Statistics on Income and Living Conditions** (EU-SILC) 2009 and **Census** of population and households 2001 (Censimento della popolazione e delle abitazioni)
- Small areas: Tuscany municipalities, unplanned domains in EU-SILC

Application (2)

Table 3: Factor loadings for two latent factors using explanatory factor analysis on EU-SILC observed variables

Variable	Factor 1	Factor 2
Severe material deprivation	0.010	**0.733**
Smog	**0.757**	0.025
Noise	**0.617**	0.154
Crime	**0.659**	0.130
Housing ownership	0.096	**-0.589**
Presence of humidity	0.010	**0.596**
Darkness inside the house	-0.002	**0.551**
Absence of rubbish in the street	**-0.843**	0.084
Absence of damages in public buildings	**-0.810**	0.012
Log equivalised disposable income	0.139	**-0.398**

Factor 1: residential area deprivation

Factor 2: housing material deprivation

Goodness-of-fit two-factor

RMSEA: 0.040

CFI: 0.925

TLI: 0.901

Data dimensionality reduction

Factor scores: FA allowing for different types of observed variables

No multivariate EBLUP on the dashboard: variables are on different scales

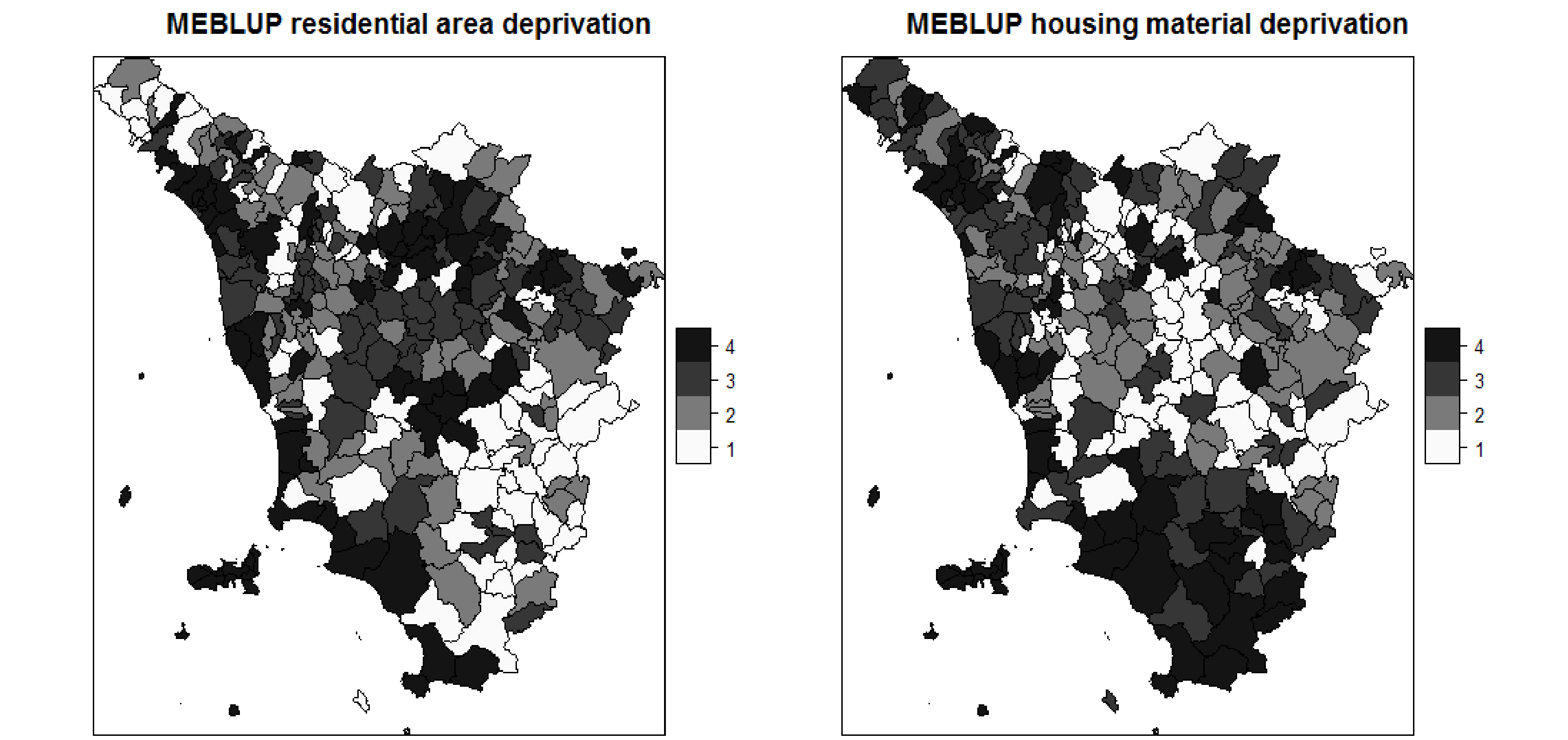
IN SUPREMÆ DIGNITATIS 1343

Centro Camilo Dagum
Tuscan Interuniversity Centre
Advanced Statistics for Equitable and Sustainable Development

MAKSWELL
MAKing Sustainable development and WELL-being frameworks work for policy analysis

Small Area Methods for Monitoring of Poverty and Living Conditions in EU
Jean Monnet Chair

Application (3)



Percentile	0%	25%	50%	75%	100%
Residential area deprivation	0.000	0.261	0.266	0.270	1.000
Housing material deprivation	0.000	0.418	0.457	0.502	1.000

- **MSE** estimated via parametric bootstrap (Moretti, Shlomo and Sakshaug, 2018**)**
- **Percentage relative reduction (averaged)**

  6.41% and 7.90%

$$\hat{r}_e = 0.10, \hat{r}_u = 0.78$$

$$\widehat{ICC}_{f_1} = 0.21, \widehat{ICC}_{f_2} = 0.0.09$$

Centro Camilo Dagum
Tuscan Interuniversity Centre
Advanced Statistics for Equitable and Sustainable Development

MAKSWELL
MAKing Sustainable development and WELL-being frameworks work for policy analysis

Small Area Methods for Monitoring of Poverty and Living Conditions in EU
Jean Monnet Chair

Conclusion and future work

- Confirmed what found previously in the multivariate SAE literature: **gains in efficiency** of multivariate EBLUP over the univariate EBLUP depend on *ICC*, sign and magnitude of r_u and r_e
- Since factor analysis models may change the **correlation (and variance) structure in the observed data**, this needs to be considered in the multivariate modelling: multivariate EBLUP may not provide large gains
- **Current and future work** is about the use of multivariate generalised mixed models in small area estimation: allowing for different types of responses
- Factor analysis models allow for different types of responses in data dimensionality reduction.

Thank you

IN SUPREMÆ DIGNITATIS 1343

Centro Camilo Dagum
Tuscan Interuniversity Centre
Advanced Statistics for Equitable and Sustainable Development

MAKSWELL
MAKing Sustainable development and WELL-being frameworks work for policy analysis

SMALL AREA METHODS FOR MONITORING OF POVERTY AND LIVING CONDITIONS IN EU
Jean Monnet Chair

Contact information

Dr Angelo Moretti
Researcher in Small Area Estimation Methods
University of Manchester - University of Sheffield (U.K.)
angelo.moretti@manchester.ac.uk or a.moretti@sheffield.ac.uk