

Small Area Estimation of Latent Economic Wellbeing

Angelo Moretti¹²

Natalie Shlomo¹ and Joseph Sakshaug³

¹Social Statistics Department, University of Manchester, U.K.

²Geography Department, University of Sheffield, U.K.

³Institute for Employment Research and University of Mannheim, Nuremberg, Germany.

11th International Conference of the ERCIM WG on
Computational and Methodological Statistics (CMStatistics 2018)

Pisa, 14-16 December 2018

Topics

- Economic wellbeing
- Measuring latent economic wellbeing at a “local level”
- The use of factor scores as composite estimates
- EBLUP of factor scores mean
- Mean Squared Error estimation of an EBLUP of factor scores mean
- Unit-level approach
- An application

What is latent economic wellbeing

- Wellbeing is a **multidimensional phenomenon** and not directly observable
- A continuing debate about the suitability of using composite estimates based on **averaging social indicators** vs. using a **dashboard** of single indicators
 - Composite indicators lead to a loss of information (Ravallion, 2011)
 - Yalonetsky (2012): composite estimates are necessary when the aim is measuring multiple deprivations (or wellbeing) within the same unit (individual or household)

The use of factor analysis models

- Factor analysis models can be used to provide composite estimates of social phenomena (OECD-JRC, 2008)
- The **factor scores** provides the composite estimates (Moretti, Shlomo and Sakshaug, 2018a,b)
- Why factor scores?
 - Relatively easy to obtain composite estimates for variables measured on different scales simultaneously
 - Easy to interpret: they are linearly related to the observed variables

The Setting

- We assume first **one wellbeing dimension** $M=1$, e.g. economic wellbeing
- These dimensions come from **a priori developed** wellbeing frameworks: single indicators (dashboard) are already grouped into dimensions e.g. Italian BES 2015
- Composite estimates can be produced for the dimension defined as the latent variable
- Moretti, Shlomo and Sakshaug (2018a) compare the use of a dashboard of **univariate** Empirical Best Linear Unbiased Predictors (EBLUPs) of small area means to the case of an EBLUP of a single factor score means;
- A confirmatory factor analysis approach is used
- Moretti A., Shlomo, N and Sakshaug, J. (2018a) Small Area Estimation of Latent Economic Wellbeing. *Sociological Methods and Research (In Press)*.

Simulation study (1)

Generation of the population

- $N = 20,000$, $D = 80$, and $130 \leq N_d \leq 420$. $N_d \sim \mathcal{U}(a = 130, b = 420)$, $\sum_{d=1}^D N_d = 20,000$
- Multivariate nested-error regression model (Fuller and Harter, 1987)

$$\mathbf{y}_{di} = \mathbf{x}_{di}^T \boldsymbol{\beta} + \mathbf{u}_d + \mathbf{e}_{di}, i = 1, \dots, N_D, d = 1, \dots, D$$

$$\mathbf{u}_d \sim \text{iidMVN}(\mathbf{0}, \boldsymbol{\Sigma}_u), \mathbf{e}_{di} \sim \text{iidMVN}(\mathbf{0}, \boldsymbol{\Sigma}_e),$$

\mathbf{u}_d and \mathbf{e}_{di} are independent.

- \mathbf{y}_{di} 3×1 vector of correlated ($r_e = 0.5$) observed responses for unit i belonging to area d
- Two uncorrelated covariates are generated from the Normal distribution:
- Intra-class correlation: 0.1, 0.3, 0.8

		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 FA on the simulated population

Simulation study (2)

Simulation steps

1. Draw $S = 1, \dots, 500$ 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 s and estimate the following for each area d :
 - EBLUP of factor scores means
 - EBLUP of the mean of each observed variable y_k
 - Weighted and simple averages of standardised (across the areas) EBLUPs. The weights are the factor loadings
3. Evaluated the results via bias and empirical RMSE
4. For the case of $\rho = 0.3$ only: evaluation on the RMSE of the EBLUP accounting for the error from the factor analysis models.

Simulation study (3)

Some results

	Scenario		
	$\rho = 0.1$	$\rho = 0.3$	$\rho = 0.8$
$\widehat{\mathbf{Y}}^{EBLUP_S_Averages}$	0.780	0.996	0.999
$\widehat{\mathbf{Y}}^{EBLUP_W_Averages}$	0.793	0.996	0.998
$\widehat{\mathbf{F}}^{EBLUP}$	0.986	0.997	0.999

Table 2 Spearman's correlation estimates for the three approaches

- EBLUP of factor scores mean perform always better than weighted and simple averages of standardised EBLUPs
- Weighted and simple averages of standardized EBLUPs perform slightly worse in case of small intra-class correlation (which may be common in real data)

Simulation study (4)

Some results

Approach	Statistics	Scenario		
		$\rho = 0.1$	$\rho = 0.3$	$\rho = 0.8$
$\widehat{Y}^{EBLUP_S_Averages}$	<i>Min</i>	0.590	0.247	0.083
	<i>Mean</i>	1.432	0.336	0.119
	<i>Max</i>	4.566	0.549	0.165
$\widehat{Y}^{EBLUP_W_Averages}$	<i>Min</i>	0.610	0.247	0.083
	<i>Mean</i>	0.793	0.334	0.118
	<i>Max</i>	1.984	0.549	0.165
\widehat{F}^{EBLUP}	<i>Min</i>	0.085	0.094	0.065
	<i>Mean</i>	0.140	0.125	0.090
	<i>Max</i>	0.276	0.262	0.130

Table 3 RMSE estimates: comparison across 500 samples for the three approaches

Simulation study (5)

Is it important to take into account the variability arising from the FA model in the EBLUP of factor scores means?

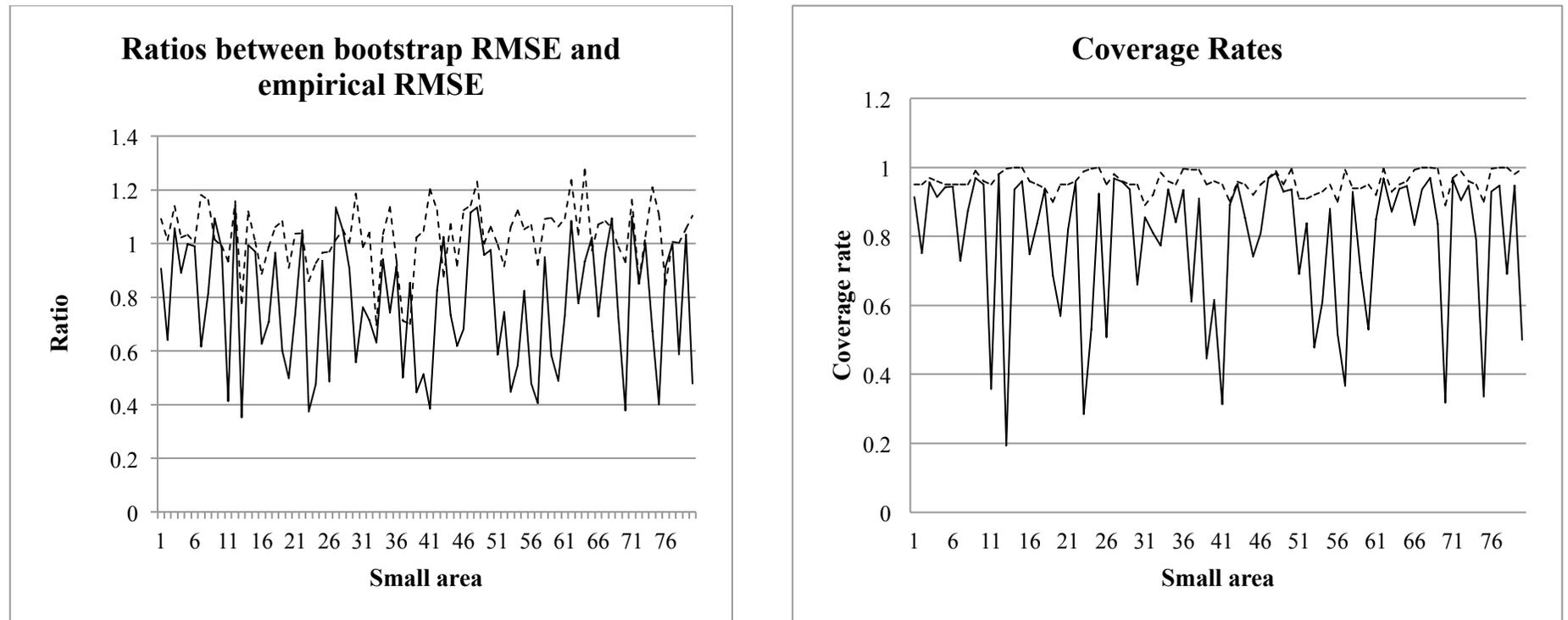


Figure 1 Taking into account the factor analysis model variability (---) vs. bootstrap ignoring the factor analysis model variability (___) EBLUP of Factor Scores case of $\rho = 0.3$.

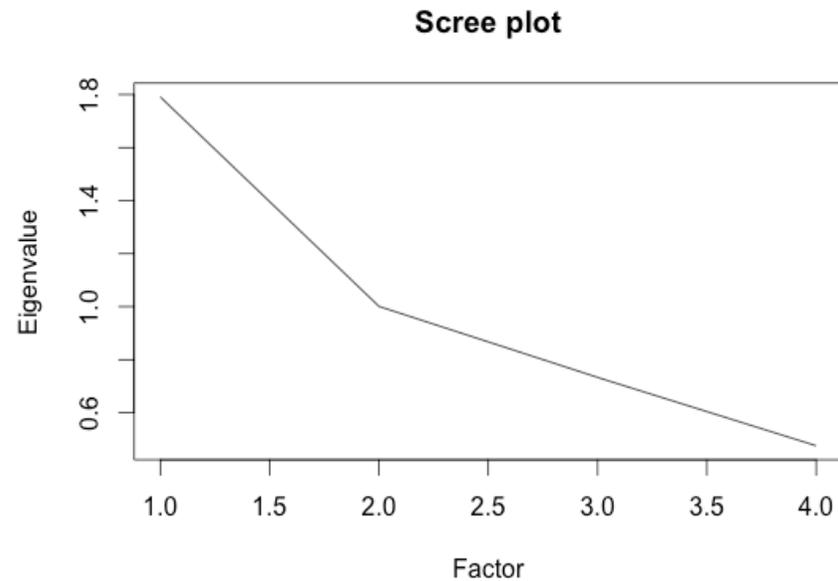
Current extension of this approach

- We study the case of $M > 1$ wellbeing dimensions in:
- Moretti A., Shlomo, N and Sakshaug, J. (2018b) Multivariate Small Area Estimation of Multidimensional Latent Economic Wellbeing Indicators. Revisions to the *International Statistical Review*.
- The use of a multivariate EBLUP is studied (Fuller and Harter, 1987; Datta et al., 1999)
- Same comparisons but in a multivariate small area estimation setting
- The MSE of the estimators for Multivariate Small Area Estimation published in
- Moretti, A., Shlomo, N and Sakshaug, J. (2018) Parametric Bootstrap Mean Squared Error of a Small Area Multivariate EBLUP. *Communications in Statistics-Simulation and Computation* (Dec. 2018)
DOI: 10.1080/03610918.2018.1498889

Application (1)

- The Italian Equitable and Sustainable Wellbeing Framework (BES 2015):
 - 12 dimensions – 134 indicators
- Economic wellbeing in Tuscany:
- Many indicators in the BES economic wellbeing dimension;
 - We chose four of them as strongly correlated and due to data availability:
 - Severe material deprivation according to Eurostat (dichotomous)
 - Equivalised disposable income (continuous)
 - Housing ownership (dichotomous)
 - Housing density as rooms per household component (continuous)
- Small areas: 287 Tuscany municipalities – LAU 2 (ex NUTS 5).

Application (2)



Factor	Eigenvalue
1	1.791
2	1.001
3	0.733
4	0.475

Table 4 Eigenvalues of EFA

Figure 2 Scree plot of EFA

- We estimated a FA model with **one factor**: (RMSEA=0.047; CFI=0.966)

Application (3)

EBLUP

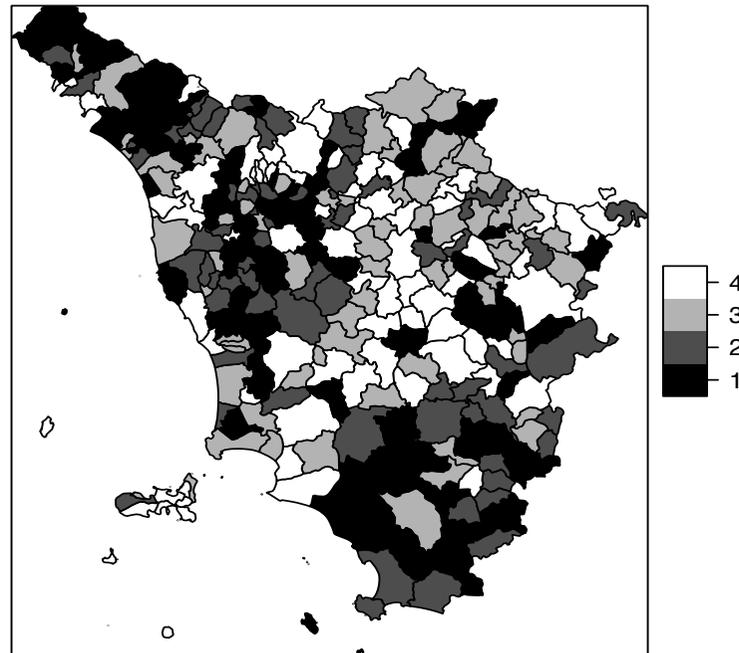


Figure 1 EBLUP of factor scores mean [1=1st quartile; 2=2nd quartile; 3=3rd quartile; =4th quartile] – lighter colour wealthier

Percentile	0%	25%	50%	75%	100%
<i>EBLUP of factor scores mean</i>	0.0000	0.5110	0.5468	0.5819	1.0000

Table 5 Percentiles in Figure 2

Conclusions and current work

- Factor scores provide more accurate and precise composite indicators at small area level (compared to the use of weighted averages) even when the intra-class correlation is small
- The variability arising from factor analysis models **must** be taken into account in estimating RMSE for model-based estimators
- Current work is related to more complex multivariate mixed-effect models in small area estimation, such as the use of multivariate generalized mixed-effect models (e.g. for binary or count data, or binary and count data all together)

References

- Datta, G. S., Day, B., and Basawa, I. (1999). Empirical best linear unbiased and empirical Bayes prediction in multivariate small area estimation. *Journal of Statistical Planning and Inference* 75: 269-279.
- Fuller, W. A., and Harter, R. M. (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.
- Moretti, A., Shlomo, N. and Sakshaug, J. (2018a). Small Area Estimation of Latent Economic Wellbeing. *Sociological Methods & Research*. In press.
- Moretti, A., Shlomo, N. and Sakshaug, J. (2018b). Multivariate Small Area Estimation of Multidimensional Latent Economic Wellbeing Indicators. Revisions to the International Statistical Review.
- Moretti, A., Shlomo, N., Sakshaug, J. (2018c). Parametric Bootstrap of a small area multivariate EBLUP. *Communications in Statistics – Simulation and Computation*. (Dec. 2018) DOI: 10.1080/03610918.2018.1498889.
- OECD-JRC (2008). Handbook on constructing composite indicators: methodology and user guide. OECD Statistics report.
- Ravallion, M. (2011). On multidimensional indices of poverty. *Journal of Economic Inequality* 9(2): 235-248.
- Yalonetzky, G. (2012). Conditions for the most robust multidimensional poverty comparisons using counting measures and ordinal variables. ECINEQ working paper, 2012-2257.

Research funded by the U.K. Economic and Social Research Council (ESRC) [grant number ES/J500094/1]

Conference participation funded by the U.K. Economic and Social Research Council National Centre for Research Methods (NCRM) [grant number ES/N011619/1].

Contact:

Dr. Angelo Moretti

Social Statistics, University of Manchester

Geography, University of Sheffield

angelo.moretti@manchester.ac.uk