Available online at www.sciencedirect.com
ScienceDirect
Procedia Economics and Finance 17 (2014) 99 – 103
Procedia Economics and Finance
www.elsevier.com/locate/procedia

Innovation and Society 2013 Conference, IES 2013

Small Area Models for analysing job placement survey data of the STELLA Consortium

Lucio Masserini[a]*, Monica Pratesi[b]

[a]Statistical Observatory of the University of Pisa, Pisa 56126, Italy
[b]Department of Economics and Management of the University of Pisa, Pisa 56124, Italy

Abstract

Job placement is a very important issue in nowadays governance of universities and data on career of graduates in the labour market are crucial also for evaluating the performance of the courses of study. The University of Pisa is member of the STELLA consortium whose aim is to perform periodic sample and census surveys for investigating and monitoring the career of graduates on the labor market. In this paper the level of satisfaction for the coherence of the employment condition with the studies of graduates one year after the degree is analysed. Small Area Models (SAE) are used to obtain more accurate estimates for the unplanned domains defined by the course of study. Focus is on the Economics and Statistics master's of science or single-cycle degree courses of the University of Brescia and Pisa.

Selection and peer-review under responsibility of the Organizing Committee of IES 2013.

Keywords: Small Area Estimation; job placement; job coherence with studies

1. Small Area Estimation

Estimating quantities of interest with survey data is a common practice both for the population as a whole and for subpopulations (domains or areas). Domain estimators computed using only the sample data from the domain are known as direct estimators but they lack of precision whenever domain sample sizes are small. In these cases we have two choices: oversampling over those domains or applying statistical methods that allow for reliable estimates in those domains. Small Area Estimation (SAE) aims at producing reliable estimates of characteristics of interest for areas or domains for which only small samples or no samples are available. Producing estimates for small areas with

* Corresponding author. Tel.: +039-050-2212181
E-mail address: l.masserini@adm.unipi.it

Selection and peer-review under responsibility of the Organizing Committee of IES 2013.
doi:10.1016/S2212-5671(14)00883-1

an adequate level of precision often requires indirect or model-based estimators that relies on the availability of population level auxiliary information. Model-based methods can be classified into two categories, namely methods based on fixed effects models, i.e. models that explain between-area variation in the target variable using only the auxiliary information, and methods based on mixed (random) effects models that include area-specific random effects to account for between-area variation beyond that explained by the auxiliary information. Mixed effects models are widely used in small area estimation (Rao, 2003). For the purposes of this analysis we refer to unit-level models, using both the Linear Mixed Effects Models to obtain the Empirical Best Linear Unbiased Predictors (EBLUPs) and the M-Quantile approach. The Linear Mixed Effects Models for small area estimation assume that a vector of p auxiliary variables x_{ij} is known for each population unit i in small area j and that information for the variable of interest y (Level of satisfaction with the "coherence between the current job and the studies") is available for units in the sample. The general linear mixed effects model has the form:

$$y_{ij} = \mathbf{x}_{ij}\boldsymbol{\beta} + u_j + \varepsilon_{ij} \qquad (i = 1,\ldots,n; j = 1,\ldots,d) \tag{1}$$

where u_j is a vector of random effects. The model parameters can be estimated by using maximum likelihood (ML) or restricted maximum likelihood (REML), usually under normality assumptions. Domain-specific means are estimated by:

$$\hat{m}_j = N_j^{-1}\left\{\sum_{i \in s_j} y_{ij} + \sum_{i \in r_j}\left(\mathbf{x}_{ij}\boldsymbol{\beta} + u_j\right)\right\} \tag{2}$$

where s_j denotes the n_j sampled units in area *j* and r_j denotes the remaining N_j - n_j units in the area. However, such models depend on strong distributional assumptions, require a formal specification of the random part of the model and not allow for outlier-robust inference. The M-quantile approach to small area estimation has been proposed by Chambers and Tzavidis (2006) and is based on the M-quantile regression model. M-Quantile model determines area effect with M-Quantile values (called M-quantile coefficients, q-values) of the units belonging to the area. This method models quantile-like parameters of the conditional distribution of the target variable given the covariates and prevents the problems associated with specification of random effects, allowing for inter-area differences to be characterised by area-specific M-quantile coefficients. An M-quantile coefficient is calculated for each area (course of study) by suitably averaging the q-values of each sampled individual in that areas. Denoting this area-specific q-value by θ_j, the M-quantile Small Area M-Quantile Model is:

$$y_{ij} = \mathbf{x}_{ij}\boldsymbol{\beta}_{\psi}\left(\theta_j\right) + \varepsilon_{ij} \tag{3}$$

where β_ψ is estimated using the iterative weighted least square and ε_{ij} has a non specified distribution. Given estimates of β_ψ and θ_j, we can obtain the small area mean estimator, as follows:

$$\hat{m}_j = N_j^{-1}\left\{\sum_{i \in s_j} y_{ij} + \sum_{i \in r_j}\left(\mathbf{x}_{ij}\boldsymbol{\beta}_{\psi}\left(\hat{\theta}_j\right)\right)\right\} \tag{4}$$

where a 'hat' represents an estimator of an unknown quantity. Note that alternative definitions of θ_j, and hence estimators θ_j of this quantity, are possible, such as the area-j median of the unit M-quantile coefficients. We refer to θ_j as the M-quantile coefficient of area *j* in what follows.

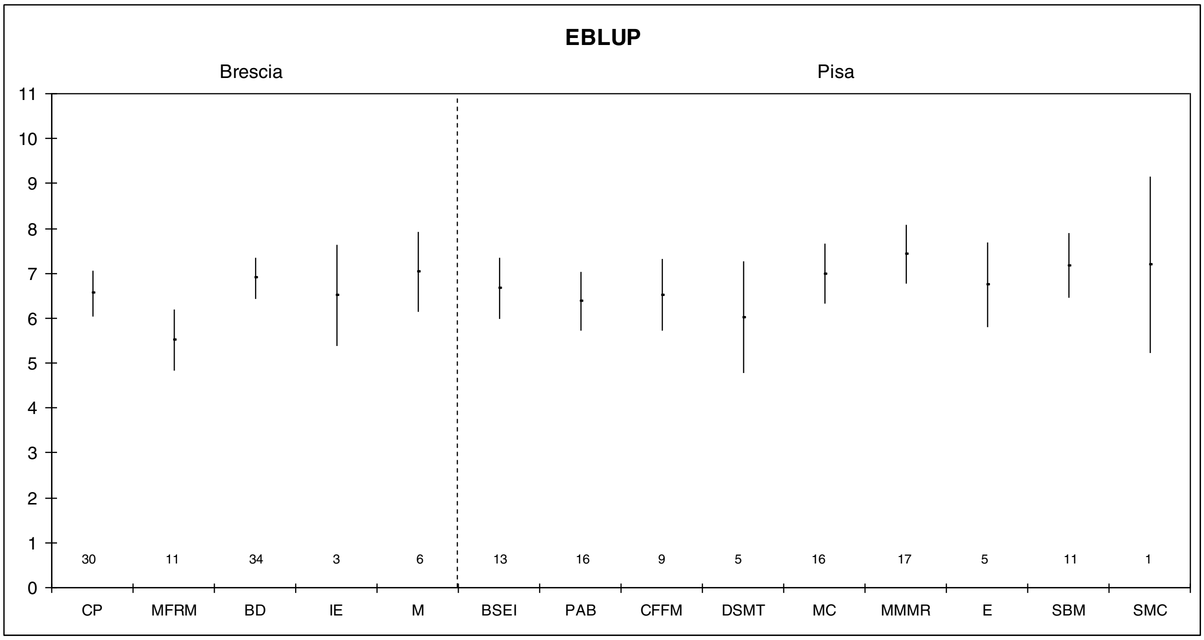
2. Data and variables

Data were extracted from the STELLA archive, limited to the population of graduates (master's degree or single-cycle courses) in 2011 in the universities of Pisa and Brescia (N = 3532). From this population a sample of graduates was interviewed (CATI+CAWI) one year after graduation (n = 2076). However, for the purposes of the analysis, the focus is on graduates in the Economics and Statistics degree courses (N = 448). More specifically, graduates come from the following course of studies: Consulting and Profession (CP; N = 86), Finance and Risk Management (FRM; N = 22), Business Direction (BD; N = 71), International Economy (IE; N = 11) and Management (M; N = 13), for the University of Brescia and Bank, Stock Exchange and Insurance (BSEI; N = 29), Professional Advice to Business (PAB; N = 44), Corporate Finance and Financial Markets (CFFM; N = 26), Development and Sustainable Territory (DST; N = 7), Management and Control (MC; N = 62), Marketing and Market Research (MMR; N = 35), Economics (E; N = 9), Strategies and Business Management (SBM; N = 29) and Strategy and Management and Control (SMC; N = 3), for the University of Pisa. The variable of interest is the Level of satisfaction with the “coherence between the current job and the studies”, measured by questionnaire on a scale from 0 to 10 only for graduates with a job at the moment of interview. The population auxiliary variables introduced into the small area models are: Gender (Female = 0; Male = 1), Residence (Province, North, Center, South, Foreign), High school diploma (Lyceum, Commercial and Technical Institute, Other) and High school mark (0 = 60-89; 1 = 90-100), Time for obtaining a degree (years), Age at graduation and Degree mark.

3. Results

The results are shown in Fig. 1 for the EBLUP (a) and for M-Quantile (b) estimators, respectively. For each course of study for the University of Brescia and Pisa, the estimated average score is indicated by a point while 95% confidence intervals are represented with bars; the number of observations for each course is reported close to the x axis. In general, EBLUP and M-Quantile estimators provides more accurate estimates than direct estimators (not reported here), justifying the use of small area methods. In particular, EBLUP performs better than M-Quantile, as shown by the narrower confidence intervals. As regard the results at course-level, EBLUP estimator indicates that the courses of Marketing and Market Research (University of Pisa), Management and Control (University of Pisa) and Business Direction (University of Brescia) have a higher mean level of satisfaction with coherence compared to Money, Finance and Risk Management (University of Brescia). Instead, because of the wider confidence intervals, M-quantile estimator shows that only Marketing and Market Research (University of Pisa) has a higher mean level of satisfaction with coherence compared to Money, Finance and Risk Management (University of Brescia).

a



b

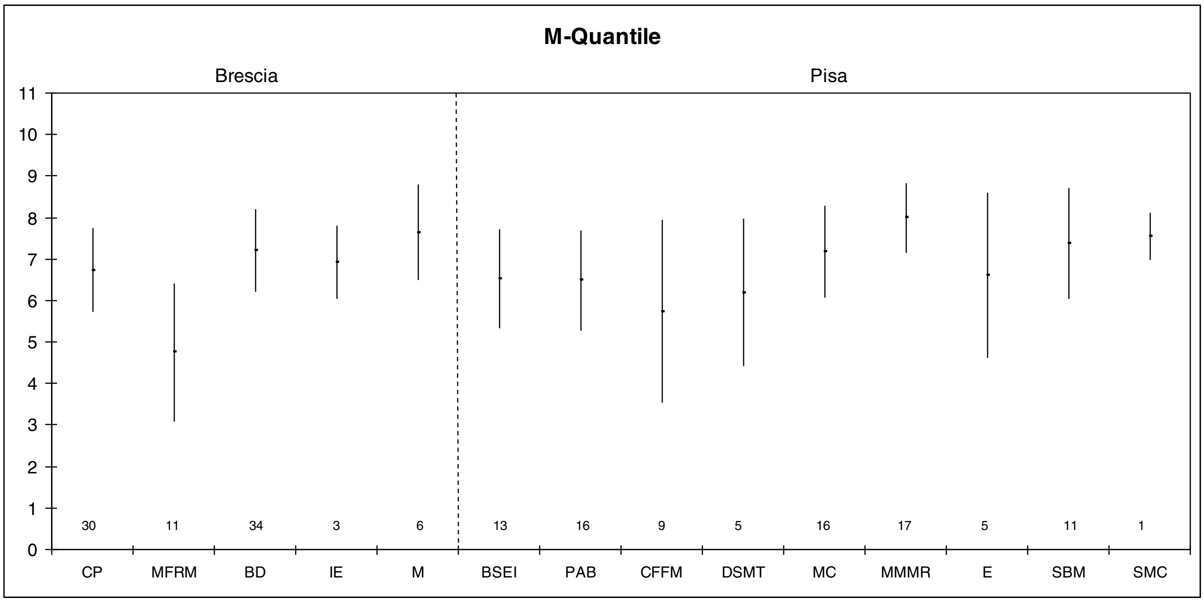


Fig. 1. (a) EBLUP estimators; (b) M-Quantile estimators.

Acknowledgements

The authors are particularly grateful for the support given by Professor Maurizio Carpita of the University of Brescia and wish to acknowledge Dr Montalbetti of the STELLA Consortium and Dr Manuela Marini of the Statistical Office of the University of Pisa for their assistance with the collection of the data.

References

Breckling, J., Chambers, R., 1988. M-quantiles. Biometrika 75, 761-771.
Carpita, M., Vezzoli, M., 2012. Statistical evidence of the subjective work quality: the fairness drivers of the job satisfaction. Electronic Journal Of Applied Statistical Analysis 5(1), 89-107.
Chambers, R., Tzavidis, N., 2006. M-quantile models for small area estimation. Biometrika 93, 255-268.
Chambers, R.L., Chandra, H., Salvati, N., Tzavidis, N., 2013. Outlier robust small area estimation. Journal of the Royal Statistical Society: Series B, 75, 1-23.
Ciavolino, E., Nitti, M., 2013. Simulation study for PLS path modelling with high-order construct: A job satisfaction model evidence. In Advanced Dynamic Modeling of Economic and Social Systems (pp. 185-207). Springer Berlin Heidelberg.
Hall, P., Maiti, T., 2006. On parametric bootstrap methods for small area prediction. Journal of the Royal Statistical Society: Series B, 68, 2, 221-238.
Rao, J.N.K., 2003. Small Area Estimation. New York, Wiley.
Tzavidis, N., Salvati, N., Pratesi, M., Chambers, R., 2007. M-quantile models for poverty mapping. Statistical Methods & Applications 17, 393-411.
Tzavidis N., Marchetti S., Chambers, R., 2010. Robust estimation of small area means and quantiles. Australian and New Zealand Journal of Statistics 52, 2, 167-186.
Vezzoli, M. (2011). Exploring the facets of overall job satisfaction through a novel ensemble learning. Electronic Journal Of Applied Statistical Analysis, 4, 1, pp. 23-38.