Small domain estimation of business parameters using skew-normal models*

Abstract

Small area estimation has been till now applied especially to poverty and employment parameters, however the demand for reliable information for restricted domains is growing also with reference to business parameters. The aim of this work is to provide estimates of business parameters for small domains defined by cross-classifying geographical areas, economic activities and firms’ size. We use data on the Small and Medium Enterprises sample survey (1-99 employees) conducted by the Italian National Statistical Institute (Istat) in 2008. We focus on the manufacturing sector and on two basic outcomes, value added and labor cost, which are used to calculate important economic competitiveness indicators. As for many of the considered domains the number of sampled firms is too low to obtain reliable estimates using the “direct” estimation strategy currently employed by Istat, a small area method estimation is advisable. We use model-based small area estimators relying on area level models, that may be represented as a sampling model, which accounts for the sampling variability of the direct survey estimates, and a model linking the small domain parameters of interest to area-specific auxiliary data. We adopt a Hierarchical Bayesian approach to estimation.

Some peculiar issues arise in business surveys (Cox et al., 1995; Rivière, 2002). One of the most relevant is the asymmetry of outcomes’ distributions due to the presence, above all in Italy, of a majority of small firms. To take into account of this peculiarity, we relax the normality assumption of the classic Fay-Herriot model (Fay and Herriot, 1979) and consider the skew-normal distribution. The specification of a skew-normal distribution offers some advantages with respect to other non-symmetric distribution, such as allowing for modeling zero and negative values. The skew-normal distribution has been already considered by Ferraz and Moura (2011) for the sampling error. We propose: i) to assign a skew-normal distribution also to the random area effects; ii) a bivariate extension of such skew-normal models, which enables us to take into account of the high correlation between the target variables. Moreover, we control for firms’ heterogeneity by including the information on the size (available from the ASIA administrative archive) as auxiliary variable. To estimate area level models direct estimates as well as their associated variance are necessary. Direct estimates are obtained by using a Horvitz-Thompson estimator because the domains of interest are collections of strata. Instead, we use a bootstrap strategy to estimate variances and covariance between direct estimates, because the unavailability of some design information hinders the replication of the calibration procedure used by Istat.

Results highlight the importance to take into account of the asymmetry of data, as the skew-normal specification allows for a considerable gain in efficiency, about 30% on average with respect to the direct estimator. Moreover, we find that it is not useful to specify a non symmetric distribution for the random effects when it is specified for the sampling errors, while it is more effective to specify a bivariate distribution for the random effects then for the sampling errors.

1. University of Bologna, Italy, silvia.pacei@unibo.it
2. University of Bologna, Italy, maria.ferrante@unibo.it

*The work is conducted within the BLUE-ETS research project financed by the European Commission under FP7 (cf. http://www.blue-ets.eu).
References