

Nonparametric estimation of cereals production at Agrarian Region level in Tuscany

Stima non parametrica del livello di produzione di cereali nelle Regioni Agrarie della Toscana

Monica Pratesi, Nicola Salvati

Dipartimento di Statistica e Matematica Applicata all'Economia,
Università di Pisa, via Ridolfi, 10 – 56124 Pisa,
m.pratesi@ec.unipi.it salvati@ec.unipi.it

M. Giovanna Ranalli

Dipartimento di Economia, Finanza e Statistica,
Università di Perugia, via Pascoli – 06100 Perugia – giovanna@stat.unipg.it

Riassunto: La regressione M-quantilica è stata proposta recentemente come nuovo approccio nella stima per piccole aree in contrapposizione all'impiego dei classici modelli misti. Il modello di descrizione della relazione fra le variabili ausiliarie e la variabile di interesse può essere anche di tipo semiparametrico basato su *splines* penalizzate. Il metodo è stato qui applicato alla stima della produzione media di cereali per azienda in ciascuna Regione Agraria della Toscana. I risultati sono presentati e discussi con attenzione all'efficienza delle stime della produzione e confrontando le loro diverse distribuzioni territoriali con le stime dirette ottenute dalla SPA2003.

Keywords: M-quantile regression, robust estimation, penalized splines, small area estimation.

1. Introduction

In small area estimation the interest is usually on the estimation of a parameter of a *small area* such as a mean or a total of a variable y . Unfortunately, sample sizes tend to be too small, sometimes non-existent, to provide reliable direct estimates for these small areas. Mixed effects models assume that the variability associated with the conditional distribution of y given a vector of covariates $\mathbf{x}$ can be at least partially explained by a pre-specified hierarchical structure, such as that induced by the small areas of interest. An alternative approach to modelling the variability of this conditional distribution is via linear M-quantile regression, which does not depend on a hierarchical structure of the data (Chambers and Tzavidis, 2006). M-quantile regression integrates the concepts of quantile regression and expectile regression within a common framework defined by a “quantile-like” generalisation of regression based on influence functions.

Pratesi *et al.* (2006) extend the M-quantile small area estimation approach to the setting in which the relationship between y and $\mathbf{x}$ is described by a semiparametric regression model (Ruppert *et al.*, 2003). In particular, the q -th M-quantile of y given $\mathbf{x}$ can be described as an additive model in which some covariates enter the model parametrically and some others nonparametrically. For the latter, the relationship is left unspecified and learnt from the data through penalized splines.

2. Nonparametric small area estimation

2.1. The data

Data come from the Farm Structure Survey (2003) and the Agricultural Census 2000. The Farm Structure Survey is carried out once every two years and collects information on farms' land by type of cultivation, the amount of breeding, the kind of production, the structure and the amount of farm employment. The sample is selected by a stratified design and the optimal allocation of sample size to the strata is obtained by minimizing the sampling error at regional and national level (see Ballin and Salvi (2004) for details on the sampling design). The target parameter is the per farm average cereal production at Agrarian Region level (AR). Accurate estimates at sub-regional level require either the enlargement of the sample in provinces or municipalities or the application of small area estimation models. The major determinant of per farm average cereal production is the utilized surface area for cereal production (ha) (Pratesi et al., 2006).

2.2. The small area estimation procedure

The method can be summarized as follows. Firstly, the *M-quantile coefficient* is estimated for each farm in the probabilistic sample: this is done defining a fine grid of values on the interval (0,1) and using the sample data to fit the semiparametric model at each value q on this grid. If a data point lies exactly on the q -th fitted curve, then the coefficient of the corresponding sample unit is equal to q . Otherwise, use linear interpolation over the grid to obtain it. If a hierarchical structure does explain part of the variability in the population data (y), we expect units within clusters (Agrarian Regions) defined by this hierarchy to have similar *M-quantile coefficients*. Secondly, an estimate of the *mean area quantile* for a small area (AR) is obtained by taking the corresponding mean value of the sample *M-quantile coefficient* of each unit in that area. In case of out of sample areas, the estimate can be set to 0.5. Thirdly, for each AR, predicted values of the cereal production for the unobserved units are obtained through the semiparametric model at the estimated mean area quantile. The estimator of the small area y mean is finally given by the average of the observed values and the predicted ones. The small area estimates are provided with a bootstrap estimate of their Mean Squared Error.

References

- Ballin, M., Salvi, S. (2004) *Nota metodologica sul piano di campionamento adottato per l'indagine "Struttura e produzione delle aziende agricole 2003"*, Istat.
- Chambers, R., Tzavidis, N. (2006). *M-quantile Models for Small Area Estimation*, *Biometrika*, 93, 255-268.
- Pratesi Monica, Salvati Nicola, Salvatore Renato (2006). "Metodi di analisi multivariata nella stima per piccole aree: possibili applicazioni in ambito agricolo". Atti del Convegno "Le statistiche agricole verso il Censimento del 2010: valutazioni e prospettive", 26-27 Ottobre 2006, Cassino.
- Pratesi, M., Ranalli, M.G., Salvati, N. (2007). Nonparametric *M-quantile* small area estimation via penalized splines, *Working Paper 279, Dipartimento di Statistica e Matematica Applicata all'Economia, Università di Pisa*.
- Ruppert, R., Wand, M., Carroll (2003). *Semiparametric Regression*. Cambridge University Press.