Agricultural Use of Tuscan Landscape: Small Area Estimation and Sub-Regional Mapping

Uso agricolo del territorio in Toscana: stima per piccole aree e m appatura a livello sub-regionale

Chiara Bocci, Alessandra Petrucci
Dipartimento di Statistica "G. Parenti,
Università degli Studi di Firenze, viale Morgagni, 59 – 50134 Firenze
bocci@ds.unifi.it alex@ds.unifi.it

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

Riassunto: Gli attuali disegni di campionamento utilizzati in ambito agricolo sono stati progettati per avere stime a livello regionale. Per ottenere stime per domini sub-regionali si rende quindi necessario applicare metodologie specifiche, basate su modelli che utilizzano informazioni ausiliarie relative anche alle aree "vicine". Tali stimatori indiretti considerano, oltre alla classica variabilità spiegata dalle variabili ausiliarie del modello, anche quella indotta da specifici effetti di area. Seguendo tale approccio, il lavoro descrive l'applicazione, sia a dati simulati che a dati reali (SPA), di uno stimatore EBLUP combinato con un modello a effetti casuali di area spazialmente correlati.
I risultati ottenuti permettono la mappatura dell'uso agricolo del territorio regionale con attenzione a domini geografici più fini di quelli attualmente disponibili.

Keywords: Small area estimation, Spatial correlation, Spatial EBLUP, Simultaneous autoregressive model.

1. Introduction

In the context of Italian agricultural surveys the term "small area" generally refers to a local geographical area, such as province or municipality. Predicting average or total crop area for municipalities has not often been attempted, both due to a lack of available survey data and the interest focused on Italian provinces (Benedetti *et al.*, 2004). When traditional municipality-specific direct estimator does not provide adequate precision it is possible to employ indirect estimators that "borrow strength" from related areas. The indirect estimators can incorporate specific random area effects that account for between areas variation beyond what is explained by auxiliary variables included in the model. Traditionally the random area effects are considered independent, but in practice, basically in most of the applications on environmental data, it should be more reasonable to assume that the random area effects between the neighbouring areas (for instance the neighbourhood could be defined by a contiguity criterion) are correlated and the correlation decays to zero as distance increases.

The aim of this work is to estimate the average production of olives in each Local Economy System (LES) of the Tuscany, using the EBLUP estimator under a model with spatially correlated errors (Salvati, 2004). The LES are aggregations of municipalities but they are different from provinces.
The paper is organized as follows: section 2 recalls the Spatial EBLUP procedure. Section 3 discusses the results of the application of Spatial EBLUP to the estimation of the average production of olives at LES level and reports some final remarks.

2. Spatial area level random effect models

Let ϑ be the parameter of inferential interest (small area total y_i, small area mean $\bar{y}_i$ with $i = 1 \ldots m$) and assume that the direct estimator $\hat{\vartheta}$ is available and design unbiased

$$\hat{\vartheta} = \vartheta + \mathbf{e} \tag{1}$$

with e independent sampling error with mean 0 and variance φ. The spatial dependence among small areas is introduced specifying a linear mixed model with spatially correlated random effects for the ϑ parameter:

$$\vartheta = \mathbf{X\beta} + \mathbf{Zv} \tag{2}$$

where $\mathbf{X}$ is the matrix of the area specific auxiliary covariates $\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{ip})$, $\boldsymbol{\beta}$ is the regression parameters vector $p \times 1$, $\mathbf{Z}$ is a matrix of known positive constants, $\mathbf{v}$ is the second order variation. The deviations from the fixed part of the model $\mathbf{X\beta}$ are the result of an autoregressive process with parameter ρ (spatial autoregressive coefficient) and proximity matrix W (Cressie,1993; Anselin, 1992):

$$\mathbf{v} = \rho\mathbf{Wv} + \mathbf{u} \Rightarrow \mathbf{v} = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{u} \tag{3}$$

where $\mathbf{u}$ is a vector of independent error terms with zero mean and constant variance σ_u^2 and $\mathbf{I}$ is $m \times m$ identity matrix.

Combining (1) and (2), with $\mathbf{e}$ independent of $\mathbf{v}$, the model with spatially correlated errors is:

$$\hat{\vartheta} = \mathbf{X\beta} + \mathbf{Z}(\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{u} + \mathbf{e}. \tag{4}$$

The error terms $\mathbf{v}$ and $\mathbf{e}$ have respectively $m \times m$ covariance matrices:

$$\mathbf{G} = \sigma_u^2\left[(\mathbf{I} - \rho\mathbf{W})(\mathbf{I} - \rho\mathbf{W}^T)\right]^{-1} \text{ and } \mathbf{R} = diag(\varphi_i) \tag{5}$$

Under the model, the Spatial Best Linear Unbiased Predictor (Spatial BLUP) estimator of ϑ_i is:

$$\tilde{\vartheta}_i^S(\sigma_u^2,\rho) = \mathbf{x}_i\hat{\boldsymbol{\beta}} + \mathbf{b}_i^T\sigma_u^2\left[(\mathbf{I}-\rho\mathbf{W})(\mathbf{I}-\rho\mathbf{W}^T)\right]^{-1}\mathbf{Z}^T \times \times\left\{diag(\varphi_i) + \mathbf{Z}\sigma_u^2\left[(\mathbf{I}-\rho\mathbf{W})(\mathbf{I}-\rho\mathbf{W}^T)\right]^{-1}\mathbf{Z}^T\right\}^{-1}(\hat{\vartheta}-\mathbf{X}\hat{\boldsymbol{\beta}}) \quad (6)$$

where $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\hat{\boldsymbol{\vartheta}}$ and $\mathbf{b}_i^T$ is a $1\times m$ vector $(0,0\ldots 0,1,\ldots 0)$ with value 1 in the *i-th* position.

The estimator $\tilde{\vartheta}_i^S(\sigma_u^2,\rho)$ depends on the unknown variance components σ_u^2 and ρ. Replacing the parameters with asymptotically consistent estimators $\hat{\sigma}_u^2,\hat{\rho}$, a two stage estimator $\tilde{\vartheta}_i^S(\hat{\sigma}_u^2,\hat{\rho})$ is obtained and it is called Spatial EBLUP.

The estimators can be obtained iteratively using the “Nelder-Mead” algorithm (Nelder and Mead, 1965) and the “scoring” algorithm in sequence. The Mean Squared Error and its estimation are not reported, more details are in Salvati (2004).

3. Results and final remarks

The Tuscany region is divided in 42 LES. The objective of inference is the average production of olives ($\vartheta = \bar{y}$) for each of the 42 small areas (LES). Auxiliary data at the LES level were available for this study. The major determinants of the average crop area are utilized: the surface area for production of olives (ha).

Hence, the Spatial EBLUP method is implemented to this data to estimate the average production of olives in each of the 42 small area within the study region using a SAR spatial model.

The neighbourhood structure **W** is defined as follows: spatial weight, w_{ij}, is 1 if area *i* shares an edge with area *j* and 0 otherwise. For an easier interpretation, the general spatial weight matrix is defined in row standardized form, in which the row elements sum to one. Sampling variances, φ_i are estimated smoothing the sampling error associated to the population level estimator (Rao, 1998). The estimated variance $\hat{\varphi}_i$ is then treated as a proxy to φ_i. As result the $mse[\tilde{\vartheta}_i^S(\hat{\sigma}_u^2,\hat{\rho},\hat{\varphi}_i)]$ is greater than $mse[\tilde{\vartheta}_i^S(\hat{\sigma}_u^2,\hat{\rho},\varphi_i)]$.

The value of the estimated spatial autoregressive coefficient $\hat{\rho}$ is 0.859 ($s.e. = 0.113$) with ML procedure and suggests small spatial relationship. Figure 1 displays the map of the Tuscany with the Spatial EBLUP estimates for the average production of olives per LES.

The Spatial EBLUP method provides estimates with smaller average estimated standard errors than the EBLUP estimators (Table 1). Table 1 shows also the average estimated of *MSE* and its decomposition in g_1, due to the random effects, g_2, which accounts for the variability in the estimator $\hat{\boldsymbol{\beta}}$, g_3 due to estimate ρ and σ_u^2.

To assess the accuracy of the Spatial EBLUP estimator a simulation study was carried out. The results (not reported here for lack of space) show a slight improvement of the MSE estimates (Pratesi and Salvati, 2005).

Figure 1: The average production of olives estimated for the 42 LES of Tuscany.

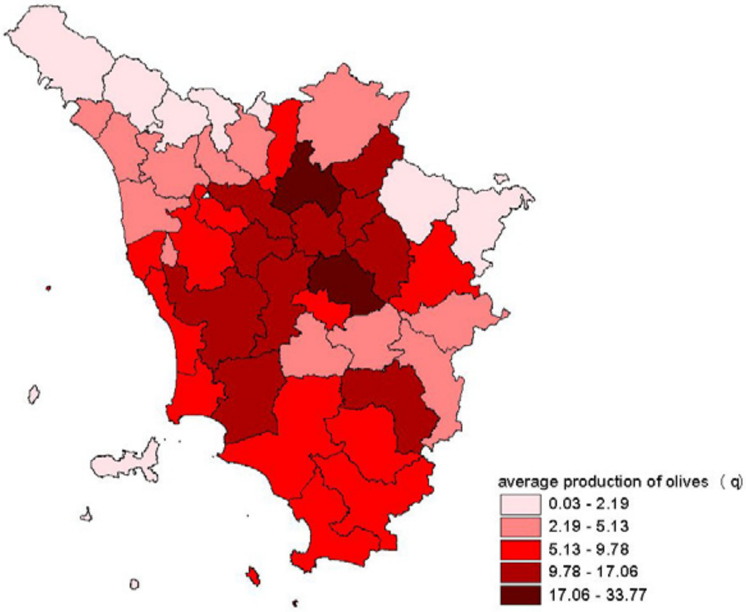

Table 1: *Average Estimated Standard Errors (A.E.Se.) of EBLUP and Spatial EBLUP estimators.*

Estimator	A.E.Se.	A.E.MSE	A.E.(g_1)	A.E.(g_2)	A.E.(g_3)
Spatial EBLUP	1.14	1.61	1.43	0.65	0.03
EBLUP	1.34	2.15	0.90	0.57	0.06

In conclusion, considering the case study, the use of Spatial EBLUP methodology, which takes into account the SAR spatial model in the small area estimation, reduces the confidence interval.

References

Anselin, L. (1992) *Spatial Econometrics: Method and Models*, Kluwer Academic Publishers, Boston.

Benedetti, R., Espa, G., Piersimoni, F. (2004): Esperienze di Stime provinciali indirette di variabili aziendali. *Contributo al Convegno ISPA 2004*, Università di Cassino.

Cressie, N. (1993) *Statistics for Spatial Data*, John Wiley & Sons, New York.

Pratesi, M., Salvati, N. (2005) Small Area Estimation: the EBLUP estimator with autoregressive random area effects, *Working Paper 261*, *Dipatimento di Statistica e Matemtica Applicata all'Economia*, *Università di Pisa.*

Rao, J.N.K. (2003) *Small Area Estimation*, John Wiley & Sons, New York.

Salvati, N. (2004) *La correlazione spaziale nella stima per piccole aree: metodi proposti e casi di studio*, Tesi di Dottorato in Statistica Applicata, Dipartimento di Statistica "G. Parenti", Università degli Studi di Firenze.