Disaggregating agricultural statistics by small area models

Monica Pratesi – University of Pisa
monica.pratesi@unipi.it
Alessandra Petrucci – University of Firenze
alessandra.petrucci@unifi.it

Agricultural Statistics: what are they?

- They can be intended in a broader sense:

context indicators (among others: Territory, Population density, Employment rate, Self-employment rate, Unemploymen rate, GDP per capita, Poverty rate),

environment indicators (among others: Land cover, Less favoured areas, Farming intensity, Protected forestry, Water quality)

Agricultural Statistics: what are they?

- …or in a more specific sense:

sectorial indicators: (among others: Employment by economic activity, Agricultural holdings (farms), Agricultural area, Agricultural area under organic farming, Irrigated land , Livestock units , Farm labour force)

Reference: The EU's common agricultural policy (CAP) for our food, for our countryside, for our environment

Disaggregation: what does it mean?

- Obtaining quantities of interest – known at a population level - for subpopulations (also known as domains): domains can be defined by any characteristics that partition the population into a set of mutually exclusive subpopulations.

  - Broad areas – Population level: Area under organic farming by EU member states, livestock units by country. Per farm average production by state,
  - Domains – Subpopulation level : Area under organic farming by Regions, livestock units by Regions. Per farm average production by Regions or local administrative units,

Disaggregation: simply division or estimation?

- Estimation for many of the previous indicators!!!!!

- Data Sources: sample surveys, censuses, registers (see data sources for CAP policy at EU level), satellite imagery, GIS data
- Direct domain estimators are computed using only the sample data for the domain

  - Example: per farm average production (crop or livestock) by Region from Farm Structure Survey (Eurostat)

Disaggregation: SAE estimation!

- Statistical Problem: veracity, accuracy of domain estimates
- Direct estimates often lack precision when domain sample sizes are small. Domains for which direct estimates of adequate precision cannot be produced are known as small areas.
- Survey designs usually focus on achieving a particular degree of precision for estimates at a much higher level of aggregation than that of small areas; therefore, the sample sizes for small areas are typically small.
- Producing estimates for small areas with an adequate level of precision often requires Model based - indirect estimators (SAE models!) that use auxiliary data or values of the variable of interest from related areas, or both.

SAE models

- The traditional indirect estimators, such as synthetic and composite estimators, rely on implicit linking models.
- Synthetic estimators for small areas are derived from direct estimators for a large area that covers several small areas under the assumption that the small areas have the same characteristics as the large area.
- Composite estimators are essentially weighted averages of direct estimators and synthetic estimators.
- Both synthetic and composite estimators can yield estimates that provide higher precision compared to direct estimators, also including spatial GIS data.
- However, both types of estimators share a common tendency to be design-biased, and the design bias does not necessarily decrease as the sample size increases.
- More details in Pratesi and Petrucci, FAO: http://www.gsars.org/spatial-disaggregation-and-small-area-estimation-methods-for-agricultural-surveys-solutions-and-perspectives/)

Learning by an example

- Sectorial indicator: per farm average production of grapevine by Region (population) and by Agrarian Region (domain)
- Data sources: FSS and Census data + GIS data
- Statistical problems: accuracy, point mass at zero, highly skewed distribution of the non-zero values, strong spatial structure of the data
- Benchmarking: data from expert evaluation

Data sources: more details

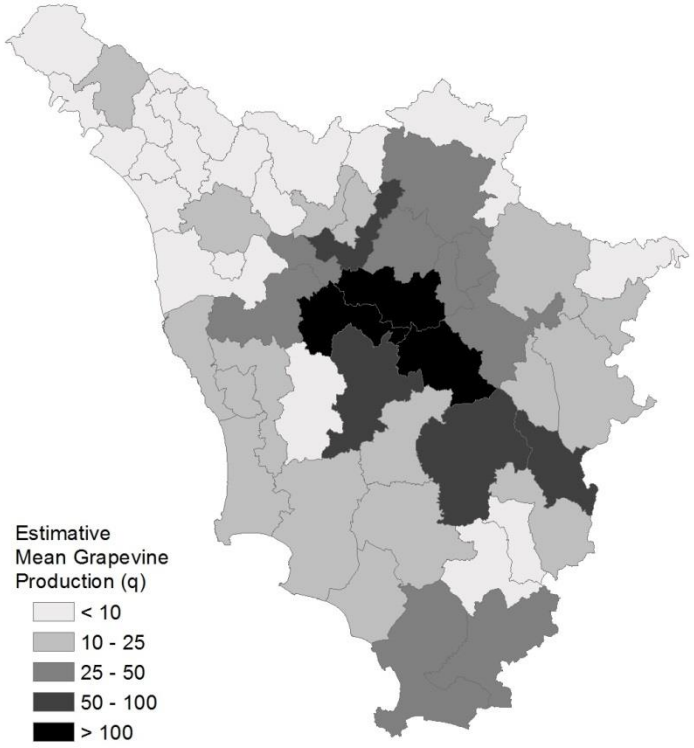
- Agricultural Census ten-yearly (ISTAT).
- Farm Structure Survey (FSS) two-yearly (ISTAT).
- The unit of observation is the farm and the data of the surface areas allocated to different crops are registered for each farm.
- The FSS survey is designed to obtain estimates only at regional level, therefore to obtain estimates at sub-regional levels it is necessary to employ indirect estimators that “borrow strength” from related areas.
- The indirect estimators can be based on regression models that use the variables collected at the census time as auxiliary variables, known for all the population units, and that can incorporate specific random area effects to account for the residual between area variation.
- A particularly useful information: the farms location from the Fifth Agricultural Census (2000).

The SAE model

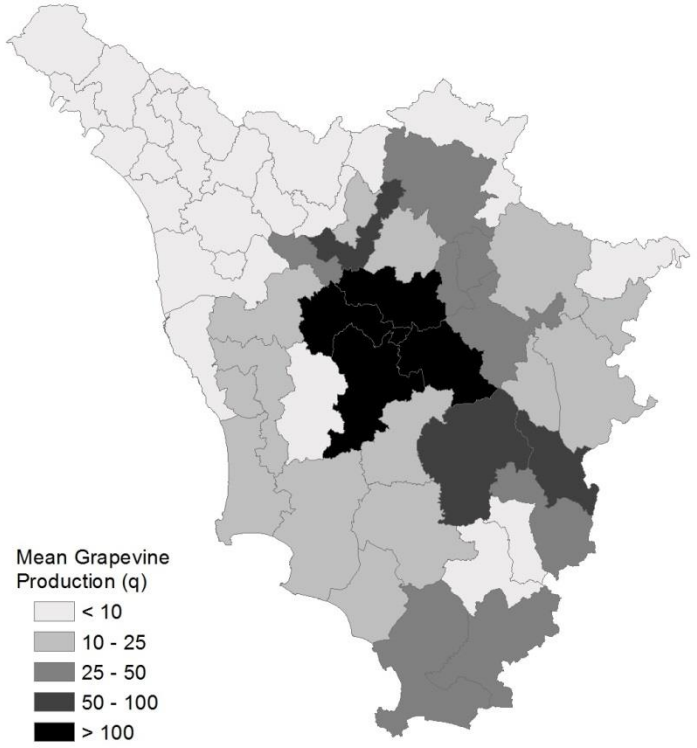
- Example: a two-part geoadditive SAE model to estimate the per farm average grapevine production in Tuscany at Agrarian Region level.
- The nature of the study variable does not allow the use of classic small area methods that assume a linear mixed model and don’t take into account the spatial structure of the data.
- A large number of farms don’t cultivate grapevines, and a few produce the majority of the total region production.
- The production of grapevines for each farm depends on the characteristics of the territory in which the farm is located.
- The quantity of grapevine produced by the same allocated surface may change, depending on the soil productivity and on the production choices of the farms (relative to the typology and quality of the produced grapevine).

The results -1

Agrarian region level estimates of the mean grapevine production



Expert's estimates of the mean grapevine production at agrarian region level.



The results -2

- The map of the estimated agrarian region means presents an evident geographical pattern, with the higher values in the areas belonging to the provinces of Florence and Siena (the well known zone of Chianti corresponding to the ‘Medio Valdarno’, ‘Val d’Elsa Inferiore’, ‘Colline Greve e Pesa’, ‘Alta Val d’Elsa’ and ‘Colline del Chianti’ ARs) and and ‘Montalcino area’ (‘Val d’Arbia’ AR), ‘Montepulciano area’ (‘Alta Val di Chiana’ AR) and ‘Scansano area’ (‘Colline dell’Albenga’ AR).
- The lower values in the north mountainous area of the provinces of Massa Carrara and Lucca
- This pattern is confirmed by the expert’s estimates means produced by ISTAT which are obtained by determination of a crop specific coefficient of soil productivity and are released at provincial level. (figure at the right)

The results -3

- The results confirm the expected pattern of grape wine productions in Tuscany: higher values in the parts where the famous ‘Chianti’, ‘Brunello di Montalcino’, ‘Nobile di Montepulciano’ and ‘Morellino di Scansano’ wines are produced and lower values for the parts of Tuscany where high mountains are located.
- The disaggregation process worked well, it was an interdisciplinary work, consistent with the CAP policy indicators.

THANKS

Acknowledgments: Elisabetta Carfagna – FAO Global Strategy ; Emanuela Dreassi, Chiara Bocci, Emilia Rocco, Università di Firenze; Istituto di Neutraceutica Università di Pisa; Regione Toscana, ASESD “Camilo Dagum”, ISTAT