

The Organization of the Agricultural Statistical System Towards the 2010: Problems of Farm Definition, Data Collection and Estimation

*L'Organizzazione del Sistema delle Statistiche Agricole Verso Il 2010:
Problemi di Definizione delle Aziende, Raccolta delle Informazioni e Stima*

Andrea Giommi (*), Massimo Sabbatini (**)

(*) Dipartimento di Statistica "G. Parenti", Università degli Studi di Firenze

(**) Dipartimento Istituzioni, Metodi Quantitativi e Territorio, Università degli Studi di Cassino

Riassunto: Il lavoro ha per oggetto alcuni temi di ricerca affrontati nell'ambito del PRIN: *Innovative methods, organization and contents in sample surveys for agriculture and environment*, recentemente concluso ed è diviso in due parti nelle quali si affrontano distinte problematiche. Nella prima parte, viene discussa l'adeguatezza della definizione di azienda agricola adottata nelle indagini, in rapporto all'attuale contesto evolutivo dell'agricoltura e del carattere di multifunzionalità di numerose aziende. Nella seconda, si affrontano problemi di stima per piccole aree, cioè per domini territoriali di dimensioni inferiori a quello regionale per il quale sono progettate le indagini correnti.

Keywords: Multifunctional farms, Systemic classification, Small area estimation

1. Introduction

This work deals with some issues come out from the PRIN *Innovative methods, organization and contents in sample surveys for agriculture and environment*. Its interdisciplinary feature has suggested to split the paper into two parts: in the first, the capacity of the current system of agricultural statistics to interpret the demand for information from users is analysed in the light of the most recent developments in the food and agricultural policy at both national and European level. Particularly, we focus on the object of the structural surveys for evaluating if the definitions of the survey unit can be still adequate to the needs of the policy makers and of the other operators of the agricultural sector. Up to now, in fact, the debate on this point has focused mainly on the survey costs connected with both a too much extensive definition and the harmonization of the definitions between the countries of the European Union. In this paper an economic policy approach is followed in order to evaluate the impact of the evolution of the agricultural sector and of its policies on the definition of the survey object. The second part, we deals with estimation problems. Structural agricultural surveys have been designed in order to supply accurate estimates at regional level. However there is a growing demand for reliable data also at sub-regional level such as provinces, municipalities and local economy systems (LES). As it is impossible to reach an adequate level of accuracy applying the estimation methods which are designed for the regional level, we have to cope with a small area estimation problem. We have devoted the second part of the paper to a synthetic review of some of the methods that have been developed to face these needs. Finally after some evaluation of the work carried out, we point out some research line towards the 2010 agricultural census.

2. Systemic classification and empirical definitions of farms

The Common Agricultural Policy (CAP) was introduced in the 1950s with the declared objective of guaranteeing food supplies in post-war Europe. In the 80s, the attainment of food self-sufficiency and the imbalances on the domestic and international markets created by the agricultural policies led to the rethinking of the policies, culminating in the scrapping of a production-based approach, in favour of a gradual commitment to the goal of rural development. Agriculture is increasingly no longer viewed as a means for satisfying the primary needs of society, but as an economic activity, capable of contributing to the well-being and development of the rural populations and, indeed, of the community at large, thanks to its capacity to produce income, healthy food and positive externalities.

This new policy guidelines translate directly into a changing demand for information: the emphasis, in fact, is shifted away from the organisation of production towards the competitive capacity of holdings and their relations with the surrounding geographical area and the social, economic and physical environment. Consequently, the priority request of data from the statistical system no longer concerns food supply and security which had characterised the system from its inception, but it is currently increasingly focusing on the aspects related to the competitiveness of holdings, their management criticalities and their capacity to generate positive externalities. These guidelines have translated into a lesser emphasis being placed on structural data, and growing attention being paid to the economic, relational (with regard to districts and supply chains) and environmental impact data.

The new trends in the demand for statistical information directly affect the definition of the subject-matter of the survey, i.e. the decision relating to “which information is relevant when describing agriculture and which isn’t”. In particular, there are two opposing trends: on the one hand, the recognition of the multifunctional role of agriculture broadens the scope of the survey, in order to take into account of all the stakeholders using the management of the biological cycle to produce a personal income and benefits to society as a whole. On the other hand, the scope narrows to exclude the observations relating to stakeholders who, despite the fact that they run a food-production activity, are unable to activate a significant set of relations with the market and the social, physical and economic environment.

Agricultural holdings for statistical purposes appear as vaguely defined survey items, because the concepts of multi-functionality and relational capital – although clearly outlined theoretically – are difficult to apply empirically. This translates into a difficulty to identify *a priori* defining schemes, capable of being non-ambiguous, effective and consistent with the economic theory at the same time.

This study is an attempt to overcome the limits of the *a priori* approach, to foster the emergence of the characteristics of a viable definition of agricultural holding based directly on the empirical analysis. The goal is to develop a typology based on the emerging strategies of the holdings, through multivariate analysis methods applied to the broadest possible set of observations and variables drawn from the *Agricultural Holding Questionnaire* attached to the 5th General Agricultural Census (GAC), capable of rebuilding the set of relations existing between the holdings and the external environment. The new definition of agricultural holding will be based on the set of typologies, wherein the agricultural activities of the rural households achieve significant levels of interaction with the market (income production) and positive externalities.

The field of observation consists of the 2,241,921 holdings managed directly by the farm holders, accounting for 93.4% of the total survey holdings, and is, therefore, capable of providing an exhaustive picture of the national primary sector.

The reference framework for classifying the emerging strategies of agricultural holdings is inspired by the systemic theory, which describes an enterprise as an “open system”, i.e. a unit that interacts with the external economic environment. In this context, strategy may be defined as the set of decisions governing relations between an enterprise and the external environment (Ansoff, 1980). The systemic approach, although outdated, is nevertheless a consolidated theoretical basis. It allows to identify a relationship between the strategy, i.e. a set of decisions and objectives, with a set of observable behaviours, i.e. the relations with the environment. The link between the census figures and agricultural holding strategies, therefore, is indirect and is mediated by a twofold interpretational activity by the researcher: the holding data are interpreted to extract relations, which in turns are then used to infer strategic approaches.

In order to make the systemic classification operational, it is necessary to clearly define the concept of external environment, with respect to a holding. The recent literature on multi-functionality has highlighted how agricultural holdings have a large number of exchanges with a wide range of parties, through both food marketing and the production of environmental and social externalities (Mollard, 2003). For the purposes of this study, the external environment of a holding has been broken down into five fundamental ambits, defined as “domains”, namely, production factors, market, rural household, geographical area and natural resources (i.e. the environment proper). This specification is aimed at outlining the set of relations pertaining to the holding, and deriving from both its economic activities and its position inside a specific spatial, social and cultural context. Each environment is described by a set of variables drawn from the *Agricultural Holding Questionnaire*, and which directly or indirectly refer to a specific domain. The analysis makes use of a total of 66 variables, 12 of which appear in two domains, for a total of 78 active variables: 22 for the structural classification, 11 for the commercial, 22 for the household, 10 for the relational and 13 for the multifunctional. The repetition of the variables is due to the relevance of certain particular items in more than one domain: a clear example is the presence of organic productions, which affect both relations with the product markets and the environmental impact of the holdings.

Based on these assumptions, the study comprises two phases: in phase one, the census figures are used to define five typological classifications capable of describing the set of relations between a farm and each of the five reference domains; in phase two, the five typologies are used as active variables in a new cluster analysis, enabling the determination of nine reference behavioural archetypes. These archetypes-aggregates holdings according to a principle of homogeneity in their relations with the five reference domains. The method is fully consistent with the notion of emerging strategy, i.e. a set of relations with the environment that take place even without being explicitly formalised and which manifest themselves – to the eye of an observer – through the actual behaviour of the holdings.

Within the framework of this study, the directly managed holdings have been classified based on their relations with each of the five domains of the economic environment. The adopted methodology is based on the Multiple Correspondence Analysis – Cluster Analysis algorithm, which has produced uniform typologies of agricultural holdings, based on an exploratory approach. The five classifications thus obtained have been

presented in previous publications, to which reference should be made for a more detailed description (Sabbatini, Cardillo and Spigola, 2005; Cocuzzoli and De Rosa, 2005; Adinolfi, Cardillo and Russo, 2005; Palombo and Bartoli, 2005; Cardillo, Grassi and Greco, 2005).

The procedure has enabled the assignment of each holding to a specific typology for each of the five domains. These typologies have then been used as active variables for a subsequent cluster analysis, which has allowed the identification of nine uniform groups, with regard to relations with the five environmental domains of reference. -

According to the systemic theory, the clusters may be interpreted as archetypes describing the principal typologies of behaviour emerging from the single national agricultural holdings. Table 1 shows the number of single holdings and the principal structural statistics. Following is a brief description of the classes, while references have been added for further investigation (Russo and Sabbatini, 2006).

The first cluster may be defined as the archetype of *Marginalization Marginal Farm*, describing a class of holdings with the lowest exchange with the external environment. The production unit is fully coincident with the family: the approach is based on the minimisation of externally acquired factors, on dependence from public support and on the use of the output almost exclusively for self-consumption. The management is characterised by a limited time horizon, due, on the one hand, to the old age of the parents and, on the other hand, to the lack of any young family members working in the holding. The picture also highlights a scarcity of human capital, which emphasizes the dependence on public transfers, often consisting of pension benefits. This class includes the highest number of holdings (23.7% of the total), with the lowest factor endowments among all five domains (Table 1).

Table 1: *Classification of typological archetypes*

Strategic archetypes			Average values per holding			Average values per hectare of UAA		
	No. Holdings	% Holdings	UAA (ha)	Work (days)	SGM (euros)	Work (d/ha)	SGM (€/ha)	Mechaniz. (kw/ha)
Marginal	573,974	23.7	1.4	33	1,487	23	1,060	3.1
Subsistence	432,573	17.9	2.2	116	4,567	53	2,089	6.6
Traditional	227,910	9.4	5.2	141	8,143	27	1,565	7.3
Low impact	318,721	13.2	5.4	135	6,586	25	1,227	6.6
Diversified	80,894	3.3	5.3	181	9,275	34	1,744	8.7
Residential	477,194	19.7	1.5	35	1,832	23	1,197	4.9
Value-added	27,440	1.1	3.2	284	23,136	89	7,228	17.4
Relational	154,960	6.4	9.0	265	24,130	29	2,672	7.9
Industrialised	128,255	5.3	26.7	649	55,483	24	2,076	5.7
TOTAL		100.0	4.4	127	8,216	29	1,856	6.4

The strategic archetype of *Subsistence* accounts for 17.9% of the total holdings; these are characterised by a limited interaction with the external environment in all domains. The factor endowments is low (the average UAA is 2.2 hectares), and commercial activities are accompanied by a large amount of self-consumption: only 13.7% of the holdings in this class, in fact, have declared that they sell their entire crop output. The environmental impact of these holdings seems to be defined by two opposing trends: on the one hand, there is the minimisation of factor use, which entails a downscaling of the pressure, on the other hand the short time horizon of the management, characterised by elderly (and often nuclear) families, tends to assign less importance to long-term

agricultural sustainability. In this context, the lower pressure on the environment appears to be the result of reduced investments in production and the objective of self-consumption, rather than a conscious choice, based on the environmental responsibility of the holding management.

Holdings adopting a **Traditional** approach (9.4% of total farms) feature a type of interaction with the economic environment based on the use of local relations for commercial purposes. The holding operations have an average factor endowment and they maximise farm income by selling cash crop. In 37.3% of the cases, the farmers market their entire production. This traditional category possesses the characteristic traits of Mediterranean agriculture, based on small commercial farms, which, however, often make use of income-supplementing activities in order to achieve satisfactory economic results. The strong links with the local community reflects a uniform approach, because it is based primarily on the farm holders' relations (for example, through membership of farmers' associations and cooperatives, and not on the quality of the human capital involved or on the capacity to implement multifunctional strategies, based on quality or on the improvement of the environment.

The **Low Impact** archetype comprises holdings featuring the adoption of organic or extensive farming methods. The predominant cropping systems are seed crops (30.8%), mixed crops (19%) and orchards and vineyards (14.6%). This type comprises two distinct groups of holdings: on the one hand, there are market-oriented farms that adopt a commercial approach based on quality certified by an organic label (13.9%), on the other hand, there are holdings characterised by a significant amount of self-consumption and low intensity of use of production factors. Even the choice of the commercial channel reflects this double approach: alongside a high proportion of holdings that sell their output directly (37.4%), we find a significant amount of farms that use contract marketing channels (15.8%), showing good links with the supply chain.

The co-existence of the two approaches shows how agricultural extensivisation can be both a strategic choice, aimed at exploiting commercial opportunities, and an extreme form of cutting production costs, probably related to maximisation of the political rent. In both cases, the role of the policies emerges in the determination of the strategic approach, through both the rewards given for organic crops and decoupled support and agro-environmental measures. The holdings belonging to this type, therefore, feature a high degree of consumption of policies and may benefit extremely well from the introduction of both the single premium and accentuated forms of aid modulation based on the environment.

The **Diversification** approach concerns 80,894 holdings (3.3%), which are characterised by commercial activities aimed at marketing goods and services other than the basic agricultural products. These holdings, in fact, combine agricultural production with processing activities or other farm-related activities, such as rural tourism or handicrafts. Thus, the holdings included in this typology operate on a number of different end markets, to escape the competitive pressure on the agricultural product market and open up broader opportunities of making profit. Moreover, the holdings are also characterised by the use of local connections to sell their products, either directly (also through catering activities) or through local associations.

The **Residential** type includes 477,197 holdings (19.7%), the common trait of which is the presence of young and pluriactive farm holders. In this typology of holding, agricultural activities do not serve a production purpose, but are an instrument for providing household services. The holding, therefore, is transformed from an

instrumental good to a consumer good, where farming is instrumental to the possibility, for the household, of benefiting from a flow of services, among which there emerge a residential function and the fruition of a rural culture. The holdings of this type are the example of a new use of the rural space, where the residential and social needs prevail over production. By means of a significant off-farm activity, the farm holders manage to reconcile their belonging to a rural cultural and social context with the pursuit of high levels of income from the outside.

The archetype of **Value-added farm** is the smallest of the types of holdings (1.1%) and is geographically concentrated. These holdings base their strategy on the capacity to obtain a high income from limited factor endowments, through product differentiation and high quality. The link with the local community is a fundamental component of this approach; it is achieved through the adoption of typical regional products (in 49.8% of cases) and the high endowment of relational assets (in 86.2% of cases). A similar link enables the holdings to achieve factor profitability much above the national average. The characteristics of these holdings, which are concentrated, geographically, in the regions of Trentino and Emilia Romagna, consist in the fact that all relations with the environment are organised according to production goals with a high added value. The notion of quality adopted by these holdings is not confined to commercial relations (as in the case of traditional or low-impact holdings, for example), but affects all the domains of the external environment.

The **Relational** type comprises 154,960 holdings (6.4%). The elements shared by this type are relational endowments, marketing through contract-based relationships and association channels, and a considerable production vocation. The factor endowments are high, on average: the land area is 9 hectares and the average work days are 265. The typology relates to a highly market-oriented production-based agriculture. The widespread presence of intensive models with a high environmental pressure, and the significant use of leased land (totalling, on average, 2.5 hectares per holding), seem to suggest a large availability of land assets. Despite this, the holdings manage to achieve an economic balance thanks to the high output of the factors involved (including the human capital), and the capacity to exploit relational networks for commercial purposes. The archetype of **Industrialisation** comprises 5.3% of the total survey sample. The distinctive characteristics of this class consists in the high factor endowments, which allows the holdings to benefit from economies of scale: the average UAA is 26.7 hectares and the average number of days worked is 649. The availability of factors translates into the capacity to produce high income flows, capable of adequately compensating the work of an entire family. The large availability of land assets and the capacity to access the labour market enables these holdings to benefit from significant economies of scale, attaining very high levels of income.

The new typologies highlight the great diversity existing within the national primary sector, which is why it is so difficult to define an agricultural holding based on a small number of *a priori* measured parameters, generally consisting of land area (UAA) and Standard Gross Margin (SGM). The *ex post* approach, on the contrary, allows the use of the entire information content to characterise the different holding strategies.

This typological analysis shows that 43.4% of the census universe (i.e. the marginal and residential archetypes) is composed of holdings that substantially have no connection with the production market, being characterised by self-consumption and the fruition of the rural environment as a consumer good. In the face of such a widespread situation, we must ask ourselves whether the two typologies may be considered an extreme form

of off-farm diversification or, since production is not the central focus of the holding, we should exclude these farm structures from the concept of agricultural holding and, consequently, from the field of observation of the census universe itself. This question directly concerns the definition of agricultural holding, with regard to the developments that have taken place in the use of rural space and in farming.

Based on this empirical analysis, it emerges that farming, in the case of these holdings, is related to the social, rather than the economic, sphere. The residential and cultural aspects clearly prevail over production: the financial feasibility of these holdings is achieved, first and foremost, through non-holding activities and public transfers, often consisting of pension benefits, therefore not strictly aimed at supporting agriculture. Consequently, these units may be considered a part of the local social fabric, but outside the sphere of enterprise and the agricultural market policy objectives. In these two typologies the farm household establishes relations with the external environment based on circuits that are unrelated to the current agricultural activities. Consequently, these typologies could be excluded from the field of observation of agricultural statistics, without jeopardising the wealth of information available to users.

The *ex post* definition of agricultural holding, although consistent with the new demand for information and with economic theory, is of difficult practical application. The method, in fact, consists in outlining the actual field of observation only after the survey has been conducted: the high intensity of data required to classify holdings makes it impossible to determine the inclusion in or exclusion from the survey of a holding in the design phase. However, the empirical result of this study provides an assessment parameter for the possible *a priori* criteria. Empirical classification, in fact, can be used as an excellent definition for calculating the loss of information due to alternative criteria adopted in the design phase of the survey. -

3. Estimation

As shown in the first part of this work, the numerous survey issues connected with the concept of multi-functionality of large number of the farms can hardly be studied on the sole basis of the information coming from the census and from the structural surveys.

Another important problem is that structural surveys were designed for giving accurate estimates at regional level, while there is a growing demand of accurate estimates also at sub-regional level: provinces, municipalities and other areas such as local economy systems (LES). To make an example, in the farm structure survey (FSS, which corresponds to the acronym SPA in Italian). Provinces, in particular, are the institutions charged to collect data and thus their demand for accurate data in order to design their social and economic policy is absolutely obvious.

The problem we have to cope with is well known in sampling theory as that of small area estimation. Note that in agriculture the small area corresponds frequently with a surface of relatively small size, even if, as known, the term small area refers in general to the small size of the sample in a domain of study.

In some cases the problem of small area estimation is made worse by the complete absence of sample units in the domain of study. Beyond to this, the presence of nonresponse, which is practically unavoidable in both households and economic surveys, is perhaps the main disturbing factor.

In the PRIN: *Innovative methods, organization and contents in sample surveys for agriculture and environment* that has been recently completed and in the new one: *The statistical information in agriculture: present needs and future developments* some of the problem cited above have been studied and will be studied in the next two years.

The parameters of interest for sub-regional levels may be estimated following diverse approaches. First of all we must distinguish between estimation methods based on the available data from those connected with new survey methods that could yield additional information at low cost. Available data mainly come from censuses and structural surveys such as the FSS. Useful information may be found also in administrative registries. A large part of the new research will be dedicated to evaluate the possibility of integrating survey and administrative data by either exact or statistical matching techniques.

In the research group, the most of the studies are based on the analysis of survey and census data and are devoted to the development of small area estimation methodologies. To this purpose we must distinguish between design, model assisted and model based estimation approaches. Design estimation mostly refers to the well known synthetic estimator, though calibration techniques have been recently advanced as a powerful alternative (Estevao and Särndal, 2004). Calibration methods, that in some sense may be viewed also as model assisted techniques, have not been studied yet in the context of agricultural estimation but are object of study in the new phase of the research.

Other techniques that must be classified as design based are represented by the spatial network sampling (SNS) and the ranked set sampling (RSS). These techniques however regard the sample design phase of the survey and not the analysis of available data.

Up to now the model approach has been widely favourite, since almost all the works produced in the research are based on some explicit model.

Among the model based methods it is opportune to make a distinction between parametric, semi-parametric and non parametric methods. In the study line of the research group all the three approaches have been considered.

Further, among the parametric methods we must distinguish between models which include random area effects and models which do not. Finally random errors at area level may be spatially correlated or not.

This wide range of possibilities stimulated a lot of works mostly based on models including random area effects. Since it is impossible to give complete account of all of them, I will limit myself to only introduce some of the most relevant research steps.

Between the parametric models those including random area effects are the relatively most frequent. Differences between parameters of interest in different areas may be due to the influence of both individual and area auxiliary variables. Of course, not all these auxiliary variables are observable. Beyond this, it is custom to think that the variability is also affected by random area effects which may be specified in the models.

Let m be the number of small areas and p the number of covariates. The basic area level model assumes that the $m \times p$ matrix of the area-specific auxiliary variables, $\mathbf{X}^T$, is related to the $m \times 1$ vector of small area parameters of inferential interest $\boldsymbol{\theta}$ (mean or/and total) as:

$$\boldsymbol{\theta} = \mathbf{X}^T \boldsymbol{\beta} + \mathbf{Z} \mathbf{v}$$

where $\mathbf{Z}$ is a $m \times m$ matrix of known positive constants, $\mathbf{v}$, the $m \times 1$ vector of independent random area specific effects, with zero mean and $m \times m$ covariance matrix

$\sigma_u^2 \mathbf{I}$ and $\mathbf{I}$, the $m \times m$ identity matrix. The link model is combined with the sampling model $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta} + \mathbf{e}$, where $\hat{\boldsymbol{\theta}}$ is the $m \times 1$ vector of direct estimators of $\boldsymbol{\theta}$ and $\mathbf{e}$ the vector of independent sampling errors with mean $\mathbf{0}$ and $m \times m$ diagonal covariance $\boldsymbol{\Phi}$. It is customary to assume that the matrix $\boldsymbol{\Phi}$ is known. The combined model is:

$$\hat{\boldsymbol{\theta}} = \mathbf{X}^T \boldsymbol{\beta} + \mathbf{Z}\mathbf{v} + \mathbf{e}$$

and it is a special case of linear mixed model.

Under this model, the Best Linear Unbiased Predictor (BLUP) $\tilde{\boldsymbol{\theta}}_i^{\text{th}}(\sigma_u^2)$ is extensively used to obtain model based indirect estimators of small area parameters $\boldsymbol{\theta}$ and associated measures of variability. This approach allows a combination of the survey data with other data sources in a synthetic regression fitted using population area-level covariates.

Small area estimation models usually assume uncorrelated random area effects. An interesting possibility is to extend these models with spatial correlation between the small area random effects, modelling through the Simultaneously Autoregressive (SAR) process (Pratesi and Salvati, 2005a).

The assumption of spatial independence is not unchallenged in the literature. Pfeiffermann (2002) shows that, with many areas and large cross-sectional correlations, the loss in efficiency from ignoring correlations between areas can be substantial. However, in his study the correlation between pairs of areas is inserted without considering the spatial location of the areas themselves.

Under the SAR model, the Spatial Best Linear Unbiased Predictor (Spatial BLUP) estimator of $\boldsymbol{\theta}_i$ is:

$$\begin{aligned} \tilde{\boldsymbol{\theta}}_i^S(\sigma_u^2, \rho) = & \mathbf{x}_i \hat{\boldsymbol{\beta}} + \mathbf{b}_i^T \sigma_u^2 [(\mathbf{I} - \rho \mathbf{W})(\mathbf{I} - \rho \mathbf{W}^T)]^{-1} \mathbf{Z}^T \times \\ & \times \left\{ \text{diag}(\boldsymbol{\Phi}_i) + \mathbf{Z} \sigma_u^2 [(\mathbf{I} - \rho \mathbf{W})(\mathbf{I} - \rho \mathbf{W}^T)]^{-1} \mathbf{Z}^T \right\}^{-1} (\hat{\boldsymbol{\theta}} - \mathbf{X} \hat{\boldsymbol{\beta}}) \end{aligned}$$

where ρ is the spatial dependence parameter, $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \hat{\boldsymbol{\theta}}$, $\mathbf{W}$ is the proximity matrix and $\mathbf{b}_i^T$ is a $1 \times m$ vector $(0, 0, \dots, 0, 1, \dots, 0)$ with value 1 in the i -th position.

The estimator $\tilde{\boldsymbol{\theta}}_i^S(\sigma_u^2, \rho)$ 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 $\tilde{\boldsymbol{\theta}}_i^S(\hat{\sigma}_u^2, \hat{\rho})$ is obtained and it is called Spatial EBLUP (Pratesi and Salvati, 2005a). Following this approach the EBLUP estimator has been applied on FSS data for estimating the grape production at municipality level.

The traditional EBLUP estimators do not allow the estimation in the areas where no sample unit is available but using regression estimators which cannot include random area effects. Information on spatial correlation between areas or individual units can allow small area estimation also for non sampled domains of study. There is not enough room in this note to give a detailed description of the method. Synthetically, the key solution is a Spatial EBLUP estimator which includes model based “synthetic estimation” for all the non sampled small area.

Beyond parameters like the mean or the total, also features of the study variable distribution may represent elements of great interest. Under the normality assumption it is possible to estimate the quantiles of the distribution of the study variables in each small area, but this assumption is hardly suitable in agricultural context.

Many study variables show an asymmetric distribution requiring either the specification of some models or the resort to a non parametric approach.

Small area non parametric estimation has been recently addressed in international literature. A new approach is based on modelling quantile-like parameters of the conditional distribution of the variable of interest y given the covariates $\mathbf{x}$. Unlike traditional random effects models, this technique does not depend on strong distributional assumptions, and it is robust against outlying values. In our research this new methodology has been applied for estimating the average production of olives per farm for each of the 42 LESs of the Tuscany, using data of the FSS sample.

Other approaches have been taken into account in the research. Beyond non parametric methods, also semi-parametric methods have been explored. Non parametric methods cannot include any spatial correlation between areas. In the light of the gain in efficiency due to the introduction of the spatial correlation in the EBLUP predictor, the possibility of inserting the spatial information in the M-quantile regression approach has been explored by proposing an alternative semi-parametric method of estimation. This method is applied to the sample quantiles and allows to obtain an average quantile for each area. An application has been carried out for estimating the mean of olive production per farm in Tuscany using data of the FSS.

Another approach explored in the research is based on the “geographic weighted regression” (GWR). In this approach the small area estimation is again based on a regression model, but the use of the spatial information is quite different from that discussed above in the strategy of parametric models and in EBPLUP or Spatial EBLUP estimators.

GWR extends the traditional regression framework by involving local rather than global parameter estimation. The regression model can be written as follows:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i) x_{ik} + \varepsilon_i, \quad i = 1, \dots, n$$

where (u_i, v_i) denotes the coordinates of the i -th point in the space, $\beta_k(u_i, v_i)$ is a realisation of the continuous function $\beta_k(u, v)$ at point i , $x_{i1}, x_{i2}, \dots, x_{ip}$, are the explanatory variables at point i and ε_i are error terms.

For a given data set, local parameters $\beta_k(u_i, v_i)$ are estimated using the weighted least square procedure. The weights w_{ij} , for $j = 1, \dots, n$, at each location (u_i, v_i) are obtained as a continuous function of the proximity between the point i and the other data points.

The parameters β must be estimated for each point i and the w_{ij} indicate the proximity of each data point to the location of i : data points closer to i carrying more weight in the estimation of the parameters than those farther away. In our research, this approach has been applied for estimating the agricultural surface allocated to grapevines for almost all the municipality of the province of Florence.

Finally I would cite a methodology belonging to those which must be classified as design based methodology: the spatial network sampling (SNS). SNS does not refers to the analysis phase of the survey but to the design phase. The purpose of this method is

to enlarge the original small area sample with a set of units belonging to the same area, on the basis of information collected while observing the initial sample units.

The SNS has been applied in the study of multifunctional farms in a province of Tuscany. The set s of multifunctional farms selected and surveyed by FSS in the province is considered a random sample s of size n^s . The selected multifunctional farms are localized using the cadastral sheet map. From each sample farm, information on the presence in the surrounding area of other non sample multifunctional farms is gathered. Then a set, c , of the neighbouring multifunctional farms (network) is defined by means of the cadastral sheet map computing the Euclidean distances between sample and non sample (census) farms. Let indicate with n^c the size of the set c . In such a way the initial sample of the multifunctional farms is enlarged adding the farms in the network.

From a technical point of view, the correspondence between the sample farms and the other farms of the census can be represented by a link matrix $\mathbf{WN}_{sc} = [wn_{ji}^{sc}]$ of size $n^s \times n^c$ where each element $wn_{ji}^{sc} \geq 0$. That is, the unit (farm) j of s is related to the unit i of c provided that $wn_{ji}^{sc} > 0$, otherwise the two units are not related to each other. It is assumed that for any unit j the values of the link matrix can be obtained. That is, in a real life situation, it is possible to collect all the values of $\mathbf{WN}_{sc}$, by direct interview or by some administrative sources, for any sample unit j . Once collected the values of the study variable y_i for each unit of the set c , then it is possible to define the unbiased Horvitz and Thompson estimator $\hat{Y}^c$ of the parameter of interest, say the population total Y^c (Pratesi, Salvati, 2005b). Furthermore, the variance of Y^c can easily be obtained.

4. Conclusions

Rather than to discuss the issues reviewed in the previous sections, we prefer to point out some possible lines of future research.

The implementation of new definitions of agricultural holding in the statistic system is a key issue for the 2010 Census. Our research will focus on the comparison of alternative *a-priori* definition and on the assessment of budget gains and information loss due to the adoption of more restrictive criterion. Also, we will address the design of the most efficient survey about the excluded farms.

Regarding the estimation, we have to face one more problem which has not been seriously studied till now: the non response and its impact on small areas estimation methodologies. Furthermore design based methods and model assisted methods for small area estimation have not been sufficiently studied yet, at least in the agricultural field. Finally, the need of auxiliary information especially in small area estimation carries us towards the exploration of all the administrative sources and of the possibility to link them to the survey and census data.

References

- Adinolfi F., Cardillo C., Russo C. (2005) Le strategie di mercato delle aziende agrarie del Lazio alla luce dei dati censuari: un'analisi multivariata delle aziende a conduzione diretta del consumatore, *Atti del convegno "MIAPA"*, Pisa, 21-22 ottobre 2004.
- Ansoff H. (1980); *Management Strategico* Etas Libri, Milano.
- Bocci C., Petrucci A., Pratesi M., Salvati N. (2005) Agricultural use of Tuscan landscape: small area estimation and sub-regional mapping, *Atti del Convegno Intermedio SIS 2005 "Statistica e Ambiente"*, Messina 21-23 settembre.
- Bocci C., Petrucci A., Rocco E. (2006) L'utilizzo della vicinanza spaziale a fini predittivi mediante la regressione geografica, *Atti del Convegno "AGRI@STAT"*, Firenze 30-31 maggio 2005.
- Cardillo C., Grassi A., Greco M. (2006) La multifunzionalità, punto di forza o debolezza del sistema agricolo nazionale? *Atti del Convegno "AGRI@STAT"*, Firenze 30-31 maggio 2005.
- Chambers, R., Tzavidis, N. (2005) M-Quantile Models for Small Area Estimation, *Working Paper, Southampton Statistical Sciences Research Institute*, 33pp.
- Cocuzzoli G., De rosa M. (2005) La "virtù" nelle aziende agricole della regione Lazio, *Atti del convegno "MIAPA"*, Pisa, 21-22 ottobre 2004.
- Estevao M. V., Särndal C. E., (2004) Borrowing Strength Is Not the Best Technique Within a Wide Class of Design-Consistent Domain Estimators, *Journal of Official Statistics*, 20, 4, 645-669.
- Mollard A. (2003) Multifunctionnalité de l'agriculture et territoires : des concepts aux politiques publiques, *Cahiers d'économie et de sociologie rurale*, 66, 27-54.
- Palombo L., Bartoli L. (2006) Struttura e comportamenti lavorativi delle famiglie agricole, *Atti del Convegno "AGRI@STAT"*, Firenze 30-31 maggio 2005.
- Pfeffermann D. (2002) Small Area Estimation-New Developments and Directions, *International Statistical Review*, 70, 1, 125-143.
- Pratesi M., Salvati N. (2005a) Small area estimation under a spatial model: an experiment for average olive grove area in municipalities of Tuscany, *Atti del convegno "MIAPA"*, Pisa, 21-22 ottobre 2004.
- Pratesi M., Salvati N. (2005b) Sampling strategies and multifunctionality in agricultural surveys, *Atti del Convegno Intermedio SIS 2005 "Statistica e Ambiente"*, Messina 21-23 settembre 2005.
- Russo C., Sabbatini M. (2005) Analisi esplorativa delle differenziazioni strategie nelle aziende agricole, *Rivista di Economia Agraria* n.4.
- Sabbatini M., Cardillo C., Spigola M. (2005) La componente strutturale nella determinazione della performance delle aziende del Lazio alla luce dei dati censuari, *Atti del convegno "MIAPA"*, Pisa, 21-22 ottobre 2004.
- Salvati N. (2005) M-Quantile Geographically Weighted Regression for nonparametric small area estimation, *Report n.264, Dipartimento di Statistica e Matematica Applicata all'Economia*, Pisa.
- Viviani A., Sabbatini M., Salvatore R. (a cura di) (2004) L'informazione statistica e le politiche agricole, *Atti del convegno "ISPA"*, Cassino 6 maggio 2004.