

The Coverage Survey of the 6th Agricultural Census

Antonella Bernardini, Loredana De Gaetano, Matteo Mazziotta, and Lorenzo Soriani

Abstract The Post-Enumeration Survey (PES) presented in this paper has the specific objective to measure the under-coverage and over-coverage of the agricultural holdings enumerated within the Agricultural Census carried out by Istat in 2010. On the base of the previous experience for the 5th Agricultural census 2000 the survey has been designed so that reliable estimates of under-count or over-count are obtained for geographic areas, using another independent list of units existing in a sample of cadastral maps. Thus, the PES can indicate to census data user's specific coverage problems inherent in the census data and such errors can be quantified. In this paper the main survey phases are described: the methodology and sample design of the primary units to be surveyed (cadastral maps), the survey technique, the Statistical Model and some innovations introduced in its implementation.

Key words: quality evaluation, coverage survey.

1 Introduction

The incidence of non-sampling errors, particularly in complex investigations such as those on agricultural topics in which a considerable effort of memory and knowledge of the phenomenon are required, can seriously affect the reliability of final results. Istat, by two post-Census sample surveys, has documented the statistical information quality of the Italian Agricultural Census. The quality takes on the meaning of precision that is expressed as an inverse function of the error statistics. The aim of the Istat is to provide accurate estimates of the main non-sampling errors, particularly in complex investigations such as the Census. The non-sampling error is a function of many factors: organizational aspects of the survey, the behaviour of a plurality of individuals or Institutions. This paper follows the main steps of the PES of the Italian 6th Agriculture Census, both from the theoretical and practical point of view.

2. Methodology and survey design

In a statistical survey two kind of non sampling errors are presented: measurement (or response) error and coverage error. To evaluate the Census undercount, the real number of agricultural holdings, N , is estimated, using the dual-system estimation approach, which requests an independent enumeration of the agricultural holding universe. In order to obtain the aim, Istat conducts the PES, called the coverage survey. The survey aims are to estimate the number of agricultural holdings really existing on the Italian territory during the 6th General Agriculture Census (24 October 2010) and the coverage rate as the ratio between the number of agricultural holdings pointed out over the Census and the number of agricultural holdings really existing on the Italian territory during the Census itself. The survey was conducted between June and July 2011 in Veneto and from October 2011 until January 2012 in the whole national territory. A total survey is conducted on a portion of areas (sample). Different sources and techniques from those applied by the 6th General Agriculture are used. The coverage rate estimate is based, therefore, on comparison among agricultural holdings surveyed in the Census and agricultural holdings enumerated in the sample survey. The areal sample was approximately 1,500 cadastral maps of the land register (in the autonomous provinces of Trento and Bolzano, respectively, 1,641 and 776 land parcels were extracted) that have constituted the final sampling units. The coverage survey is a two stage stratified sample survey. The primary sampling units are the Italian Municipalities, while the second-stage units are the cadastral maps. Each cadastral map sample is surveyed in order to fix the agricultural holdings having lands in the territories of the cadastral map. The estimates domains are: the whole national territory, the 5 main geographical areas, the Regions, and a further study in detail is extended at the analysis at the provinces using small area estimation techniques.

3. Survey technique

The PES was designed in order to guarantee the independence from the 6th General Agriculture Census: the survey is based on an areal sample and the technique for the identification of agricultural holdings is based on the information of the owner of land parcels. The detection technique requires that the interviewer, from information on the owner of land parcels, discovers the agricultural holding and the holder that are on sampled cadastral map. The interviewer uses the cartography of the cadastral map, the list of the owner of the land parcels included in the cadastral map and demographic information (Mazziotta et al., 2011). For each sampled cadastral map, the data collection consists of two steps:

- Step 1: the identification of the agricultural holdings. The interviewer individuates all agricultural holdings with the agricultural area that at the date of October 24 2010 was totally or partially in the cadastral maps. In this phase the interviewer has to contact the owner of the land parcels included in the cadastral map in order to find out the holder of a possible agricultural holdings, whose land areas are included in the above mentioned parcels.

- Step 2: the interview to holder by the PES questionnaire. The interviewer interviews the holder found out in the step 1 in order to collect the selected information on the agricultural holding individualized as indicated in step 1. The list of agricultural

holdings is obtained after these two steps. The main innovative aspect is the organizational use of a computerized system, in fact the interviewer acquires information by Istat, performs the two steps of the survey and returns information to Istat.

4. The Statistical Model

One of the most common Coverage Error Models, the Petersen one, is utilized to obtain an independent estimate of agricultural holding universe total N . This model uses the PES in conjunction with the Census, assuming (Wolter, 1986):

1. both the Census and the survey are referred to the same period (the Census one);
2. Census and survey enumerations are independent of one another, i. e., the probability of a agricultural holding being on the census list (capture probability) is independent of the probability of a agricultural holding being enumerated by the survey and vice versa;
3. it is possible to match the PES sample results to the Census results without error;
4. spurious events, e.g. duplicates, non-existent cases, out-of-scope cases, have been eliminated.

Assumptions 2 and 3 of the model are questionable: the homogeneity of capture probabilities across units is unlikely for the whole population, therefore homogeneous groups are created in such way that assumption holds within each group. On the other side, the assumption of perfect match between Coverage and Census lists can be relaxed.

Under the assumption that, at the end of the PES, the following situation will be presented:

		PES		
		In	Out	
Census	In	N_{11}	N_{12}	N_{1+}
	Out	N_{21}	N_{22}	N_{2+}
		N_{+1}	N_{+2}	N

where the amounts of the table are the holdings observed at the Census and at the PES (N_{11}), or in only one (N_{12} and N_{21}); however N_{22} and N are unknown. If it were possible to know N_{22} , and consequently N , it would be possible to calculate the coverage rate as:

$$\tau = N_{1+} / N$$

To estimate the unknown amount of a specific population are used statistical models of capture-recapture type, dual-systems models or log-linear models (Ding Y., Fienberg SE (1994), Hagenaars, Jacques A. (1993)).

For the 2011 PES the coverage rate will be calculate with Petersen model; it is possible to use other models with different hypothesis (Wolter, 1986). A study was conducted to calculate the coverage rate with the "Equal Catchability Model" and the "Behavioral Response Model" (Mazziotta, Russo, 2004) for the 2001 PES.

5. Concluding remarks

The quality evaluation of the Agricultural Census has been many developments during the time. The fifth edition of Census planned the coverage survey based on an areal sampling (Bernardini et al., 2011). In 2011, the coverage survey presents different innovative aspects, both methodological and organisational. The main innovative aspect is that, for the first time, the survey is conducted in all the national territory, involving many different Census regional offices. The other innovative aspects are the sample size, 1500 cadastral maps, and management system (SGRCOP) for the data collection. These innovative developments will allow obtaining the immediate availability of data and estimates at regional level; besides, the progress of the small area estimation techniques will permit to evaluate the Census quality, for the first time in Italy, at the provincial level.

Acknowledgements

Antonella Bernardini has written section 4; Loredana De Gaetano has written section 3; Matteo Mazziotta has written section 1 and 5 and Lorenzo Soriani has written section 2.

References

1. DING Y., FIENBERG S.E. (1994) Dual system estimation of Census undercount in the presence of matching error, *Survey Methodology*, 20, 149-158.
2. HAGENAARS, JACQUES A. (1993) *Loglinear models with latent variables*. Quantitative Applications in the Social Sciences Series n. 94. Thousand Oaks, CA: Sage Publications.
3. MAZZIOTTA M., BERNARDINI A., SORIANI L. (2011), *Indagine di Copertura del 6° Censimento Generale dell'Agricoltura*. Manuale delle Istruzioni, Istat, Roma.
4. MAZZIOTTA M., RUSSO M. (2004). *Multivariate Models for Estimating Coverage Error in the Italian 5th Census of Agriculture* – European Conference on Quality and Methodology in Official Statistics, Mainz, Germany, 24-26 Maggio 2004.
5. WOLTER K. (1986). Some Coverage Error Models for Census Data, *Journal of the American Statistical Association*, Vol. 81, No. 394, pp. 338-346.