

On the Use of Auxiliary Variables in Agricultural Surveys Design

Roberto Benedetti*, Giuseppe Espa**, Federica Piersimoni***

* Università degli Studi "G.d'Annunzio" di Chieti-Pescara, Dipartimento DASTA, Viale Pindaro, 42, Pescara, benedett@unich.it

** Università degli Studi di Trento, Dipartimento di Economia Via Inama, 5 – 38100 Trento, giuseppe.espa@economia.unitn.it

*** Istat, Servizio Agricoltura, Via A. Ravà, 150 – 00142 Roma, piersimo@istat.it

Auxiliary variables, both univariate and multivariate, must be efficiently used to obtain accurate estimates. They are useful *ex ante*, that is when the sample has to be drawn, but also *ex post*, as the weight calibration method.

The classical issue on efficient sample design through a stratification based on auxiliary information will be reviewed in comparison with sampling units selection methods that make an appropriate use of auxiliary variables. We will focus our attention on three approaches: the *model-based* approach, the π ps approach and the ranked set sampling method.

While the π ps is a well known sample design method, the *model-based* approach of sample surveys assumes that a *superpopulation* model is specified. It is based on a distribution of a random vector and the real population is considered as an its realization.

The *model-based* inference assumes that a long series of vector Y realizations are available, for fixed drawn sample s . The inference relies on the specified model and is constrained by s , the sample composed of the effective realization of the variable. Sample design $p(s)$ and inclusion probability do not play any role in the inference.

Ranked set sampling (further on *RSS*) has been introduced by McIntyre (1952). Since the publication of this seminal work, the literature proposed numerous *RSS* extensions for both parametric and non-parametric estimates. In its original formulation, method *RSS* starts with the selection of a simple random sample without reinsertion of n population units. Then the mean must be estimated from those units. The n are ranked in increasing order with respect to an auxiliary variable x , that is, without the effective calculation of the interest variable y .

Which method is the best depends on the application at hand. To show some evidence, we compare the methods on real data on the slaughtering sector, for which the sample strata and the variables estimates are calculated, following the suggestions that are contained in Dorfman and Valliant (2000). Finally a comparison between *ex-ante* and *ex-post* use of the auxiliary information is performed. We will draw, for each selection method applied, samples based on the census list in order to build the sample spaces and on the basis of them we will estimate which value is the "nearest" to the true census one in terms of MSE.

Bibliography

- Dorfman A.H., Valliant R. (2000) Stratification by size revised, *Journal of Official Statistics*, 16, 2, 139–154.
- McIntyre G.A. (1952) A method for unbiased selective sampling, using ranked set, *The Australian Journal of Agricultural and Resource Economics*, 3, 385–390.
- Royall R.M. (1970) On finite population sampling theory under certain linear regression models, *Biometrika*, 57, 2, 377–387.

- Royall R.M., Herson J. (1973a) Robust estimation in finite populations I, *Journal of the American Statistical Association*, 68, 344, 880–889.
- Royall R.M., Herson J. (1973b) Robust estimation in finite population II: stratification on a size variable, *Journal of the American Statistical Association*, 68, 344, 890–893.