Sampling Design for a Farm Household Economy Survey in Korea

Jongwoo Jeon
Department of Statistics, Seoul National University, Republic of Korea

Kyu Seong Kim
Department of Computer Science and Statistics, University of Seoul, Republic of Korea

ABSTRACT: The Farm Household Economy Survey is an important agricultural sample survey in Korea. The primary object of the survey is to estimate gross income and expenditures by province, and the secondary is to estimate them by farming type. To these ends, we propose a stratified two-stage cluster sampling design based on a multivariate stratifying index. Neyman sample allocation is used to determine the size of PSUs. For estimation by province, an expansion estimator is suggested. For estimation by farming type, a ratio estimator after post-stratification of farm households is proposed since farming types of sampled households are not known before the actual survey has been conducted. The proposed sampling scheme and estimation method are evaluated based on 1996 survey data. [This work was funded by the Ministry of Agriculture and Forestry, Republic of Korea in 1997. The design proposed in this paper will be used for the Farm Household Economy Survey in Korea until 2002.]

1. Introduction

The Farm Household Economy Survey (FHES) is a sample survey carried out by the Ministry of Agriculture and Forestry in all provinces of Korea. Since 1953, FHES has furnished information on farm economy to government, farm organizations and farm economy researchers to assist their activities. The main items surveyed in FHES are farm household income, agricultural income, non-agricultural income and debt. They are summarized by province and reported monthly and yearly. Also reported are the items summarized by farming type. In the latter case, only full-time farm households are used for the summary.

2. Sample Design

Sampling units: In FHES, the administrative village is used as the primary sampling unit (PSU) and the farm household as the secondary unit (SSU). There are about 39,000 PSUs and about 1,300,000 SSUs.

Stratification: PSUs are stratified into 9 strata as follows: Paddy, Upland, Orchard, Vegetable, Special crop, Livestock and poultry, Floriculture, Other farming, and Part-time farming. There are several ways of classifying PSUs into one of the 9 strata. A multivariate stratification method is proposed in the sense that we classify PSUs using the numbers of farm households for each of 8 farming types and the number of part-time farm households in the PSUs, which are available from the 1995 Agricultural Census.

Sampling scheme within a stratum: A two-stage cluster sampling scheme is designed in each stratum. We choose a_h PSUs in the hth stratum at the first stage, then b farm households in each selected PSU at the second stage. To obtain the same selection probability of each farm household in a stratum, we use a probability proportional to measure of size (PPS) sampling scheme at the first stage and simple random sampling (SRS) at the second stage.

Sample size determination: Neyman allocation is used to determine a_h using variance estimates based on 1996 survey data. Under a simple linear cost function, we find an optimal value of b which is dependent on the intraclass correlation coefficient. In FHES, intraclass correlation coefficients of main

survey variables are in the range between 0.08 and 0.14. With these values and some of the cost ratios, optimal subsample sizes are found.

3. Estimation

Estimation by province: The following expansion estimator is proposed for estimation by province:

$$\bar{y} = \sum_{h=1}^{L} W_h \frac{1}{a_h} \sum_{i=1}^{a_h} \frac{M_{hi}}{M_h} \frac{\bar{y}_{hi}}{p_{hi}} \tag{1}$$

where W_h is the weight of the hth stratum, M_{hi} the size of ith PSU in the hth stratum and $M_h = \sum_{i=1}^{a_h} M_{hi}$. Also $\bar{y}_{hi} = \sum_{j=1}^{b} y_{hij} / b$ where y_{hij} is the income of a farm household, and p_{hi} is the selection probability of the ith PSU in the hth stratum. In addition, an unbiased estimate of the variance of $\bar{y}$ is given by

$$var(\bar{y}) = \sum_{h=1}^{L} W_h^2 \left[\frac{1}{a_h(a_h - 1)} \sum_{i=1}^{a_h} \left(\frac{M_{hi}}{M_h} \frac{\bar{y}_{hi}}{p_{hi}} - \bar{y} \right)^2 \right]. \tag{2}$$

Estimation by farming type: A combined ratio estimator after post-stratification is of the form

$$\bar{y}_f = \frac{\sum_{h=1}^{L} \sum_{(i,j) \in S_{hf}}^{a_h} w_{hij} \, y_{hij}}{\sum_{h=1}^{L} \sum_{(i,j) \in S_{hf}}^{a_h} w_{hij}} \tag{3}$$

where w_{hij} is the inverse of the selection probability of the hijth SSU and S_{hf} is the post-stratified sample of the fth farming type in hth stratum.

4. Conclusion

In this paper, we have proposed a sampling scheme and estimation method for the Farm Household Economy Survey in Korea. To evaluate our design and estimators, we used the 1996 survey data to estimate coefficient of variations of main items. It was found that our design improves in accuracy over the current one which was designed in 1992.