A Modified Two-Stage Sampling Scheme with Integrated Second Stage Sample

Chia-Liang, Weng* Chang-Tai, Chao†

Abstract

In a large scale sampling survey situation, a two- or multi-stage cluster sampling design is often used to save the sampling effort. Consequently, estimation precision would be sacrificed, and often it would be difficult to estimate the subpopulation of interest since secondary sampling units are independently selected within each selected primary sampling units, hence, the within-subpopulation sample size often cannot be controlled. In order to make balance between the sampling cost and the estimation precision, a modified two-stage cluster sampling design is constructed and investigated in this research.

A set of primary units is selected in the first stage by some probability design, and then the sampling population of the second-stage sampling is composed of the integration of all the secondary units within the selected primary units. Therefore, the second-stage sample can be selected with more flexibility. Various combinations of the first- and second-stage designs are studied together with different estimators to investigate the property of this sampling design. The performance are also compared with other comparable conventional designs.

Key Words: Sampling Strategy, Stratified Sampling, Cluster Sampling, Systematic Sampling, Allocation Method

1. Introduction

The Census of Agriculture, Forestry, Fishery and Animal Husbandry is a census conducted every five years by the *Statistical Office of the Council of Agriculture, Executive Yuan, R.O.C.* that provides information about the size of operator characteristics, production practices and gross income of the agriculture sector in Taiwan. With these information, the associated authorities can measure trends, new development and other related agriculture issues of the current society. Furthermore, it can help the government with effective agriculture planning and policy-making. One feature of the census of agriculture is that it involves the collection of data at the individual holding level, of which is invaluable in providing statistical sound source of agriculture statistics. However, census, is often a time-consuming method which requires high capital investment as it involves the collection and observations of all the values of population. To collect data in a more timely manner and reduce the sampling cost, sampling survey, which need not to observe all the population units, is a practical alternative. Therefore, Primary Farm Household Income Survey has been conducted annually between two consecutive of the Census of Agriculture.

There are many sampling strategies which are commonly used in a large-scale sampling survey, each has the strength and weakness depending upon the situation. In order to have better estimating precision, stratified sampling design is of the preference for the choice of sampling method. In a stratified sampling, a population is partitioned into several strata and a stratified sample is selected by some design along with the allocated sample size within each stratum. By such a way, the selected sample can cover the study region as much as possible, hence minimize the within-stratum variance and make good estimates of the

*National Cheng Kung University, No. 1, Daxue Rd., East Dist., Tainan City 701, Taiwan (R.O.C.)
†National Cheng Kung University, No. 1, Daxue Rd., East Dist., Tainan City 701, Taiwan (R.O.C.)

population of interest. Despite of such property, the dispersion of the sample may cause the difficulty during the execution of the survey. Because most of the observation unit of agriculture household is located at rural or remote area, which is sometimes inaccessible, the travel cost may be intolerable under a limited budget. For the reasons above, two-stage cluster sampling design was proposed to eliminate the problem. When applying a two-stage cluster sampling design in this survey, one can partition the population into districts as the first-stage sampling units, and the primary farm households within each district are seen as the secondary units. Since the number of survey districts is limited due to the first-stage sampling, the survey cost can also be reduced regardless of the sacrifice of the estimation precision.

To keep the estimation precision to a certain level, it is important that the selected sample is as much diverse as the population it is. Nevertheless, it is possible that no unit in the sub-populations of interest, which the sub-population size is relatively small, would be selected into the sample under a classical two-stage sampling. Hence it may cause the increasing of the estimation error. To control the sample sizes of the sub-populations of different production types and/or scale, a modified two-stage sampling design is proposed in this research. The first step of the modified two-stage sampling design is to select the first-stage sample by some probability design. Then integrating the secondary units within the selected primary units as the second-stage sampling population, so that the secondary unit can be selected with more flexibility. With such design, it is possible to make a balance between the estimation precision and the survey cost.

To apply the modified two-stage sampling design, one has to decide the probability design used in each of the two stage. Several design combinations are discussed and examined in the following sections. Consider the most intuitive and the simplest way, simple random sampling at each stage, in section 2.1. The selection probability of sampling units are equal in both of the first- and the second-stage. In addition, five estimations for the population of interest are proposed and the associated statistical evaluations are evaluated correspondingly. However, this method does not guarantee the existence of at least one unit within each sub-population of interest. To collect a better representative second-stage sample, stratified sampling is implemented to take place of simple random sampling as the second-stage sampling design used in section 2.2. The comparison of allocation methods has also been made, and the results are as follows. In fact, the population variance of interest is composed of two parts, the between and the within primary unit variances, where the first one is of the greater portion than the other. That is, it is more advantageous to improve the first-stage sampling design rather than the second-stage sampling design. Certain probability design such as stratified sampling or systematic sampling can be used to select the first-stage sample of the primary sampling units instead of simple random sampling, and the estimation precision can be further better. Nevertheless, only the first-stage design under which the sampling weights are equal for all primary units are considered for now so that the estimations proposed in the previous chapters can be used reasonably. Several combinations of the first-stage and second-stage designs together with different estimators are studied and more detail including the division measurement of the primary units in the first-stage sampling are given in section 2.3. Some classical sampling designs such as overall simple random sampling, stratified sampling and two-stage cluster sampling, are used as baselines for evaluating the performance of the modified sampling designs in this research. In the meantime, a case study using the census data in 2015 is shown in section 3. Finally, conclusion about the sampling strategies and discussion are given in the last section.

2. Sampling Designs

2.1 SRS/SRS

Consider first the simple random sampling (SRS) at each stage, which is rather intuitive and easy, with the equal selection probability of each unit in both of the first and the second stages. Let N denote the number of primary units in the population and M_i denote the number of secondary units in the ith primary unit. $M = \sum_{i=1}^{N} M_i$ is the total number of secondary units in the population. Let y_{ij} denote the value of variable of interest for the jth secondary unit in the ith primary unit, and $y_i = \sum_{j=1}^{M_i} y_{ij}$ denote the total of value of secondary unit in the ith primary unit. The population total is $\tau = \sum_{i=1}^{N} \sum_{j=1}^{M_i} y_{ij}$. The mean per secondary unit in the ith primary unit is $\mu_i = y_i/M_i$. The population mean per primary unit is $\mu_1 = \tau/N$, while the overall population mean is $\mu = \tau/M$.

The first step is to select n primary units by simple random sampling without replacement. Next, integrate the secondary units within the selected primary sampling units as the second-stage sampling population, then selecting m secondary units by simple random sampling from the second-stage sampling population, where the sample size of the ith primary unit is denoted as m_i. The set of the selected primary units is denoted as S_1, and the selected second-stage sample set is denoted as S_2, while denoting S_{1i} as the set of the selected secondary units in the ith primary unit. S_1' is denoted as the set of distinct primary units intersecting by the second-stage sample. The total number of secondary units in the selected primary units is $K = \sum_{i \in S_1} M_i$.

2.1.1 Estimations

Based on the design, we provide five different estimators to estimate the population mean and evaluate them with some simulations.

Arithmetic Average

Make use of the naive sample mean of the second-stage sample as the estimate of the overall population mean. The estimator is defined as

$$\hat{\mu}_{1.1} = \frac{1}{m} \sum_{i \in S_1} \sum_{j \in S_{1i}} y_{ij}$$

Horvitz-Thompson: ssu

Use the Horvitz-Thompson estimator as the estimation of the population mean μ. The estimator is defined as

$$\hat{\mu}_{1.2} = \frac{1}{M} \frac{N}{n} \sum_{i \in S_1} \sum_{j \in S_{1i}} \frac{y_{ij}}{\pi_{ij}}$$

, where $\pi_{ij} = \frac{m}{K}$ is the probability that jth unit in the ith psu is included in the second-stage sampling population, for $i = 1, 2, ..., N;\ j = 1, 2, ...M_i$. Since the Horvitz-Thompson estimator can be an unbiased estimate of the population total τ , it can also be an unbiased estimator for the population mean μ.

Probability Proportional to Size

Consider the selection of secondary unit is in fact a probability proportional to size (PPS) with replacement of primary units in S_1 with inclusion probability p_i, thus an estimator of population mean in S_1 can be formed as i estimate y_i, where the ith primary unit is intersected by the selected ssu, say μ_1.

ii The average of the estimated y_i can be used as an estimator of μ_1.

iii Then define the estimator of overall ssu mean

$$\hat{\mu}_{1.3} = \frac{1}{M}\frac{N}{n}\frac{1}{m}\sum_{i \in S_1} m_i \frac{\hat{y}_i}{p_i} = \frac{1}{M}\frac{N}{n}\frac{1}{m}\sum_{i \in S_1} K \sum_{j \in S_{1i}} y_{ij}$$

where $p_i = \frac{M_i}{K}$, $\hat{y}_i = \frac{M_i}{m_i}\sum_{j \in S_{1i}} y_{ij}$.

Horvitz-Thompson: psu

Similar to the second estimator, the fourth estimator is also an Horvitz-Thompson estimation of μ while estimating the overall population mean by

i bringing out the unbiased estimator of y_i, the population total of each primary unit, then

ii the estimated y_i, for which psu i is intersected by the selected ssu, can be formed as an unbiased estimator of the population total in S_1, which is also an sample total of the psu's in the population.

iii Hence, an unbiased estimation of the overall population mean can be carried out at last.

The estimator is defined as

$$\hat{\mu}_{1.4} = \frac{1}{M}\frac{N}{n}\sum_{i \in S_1'} \frac{\hat{y}_i}{\pi_i}$$

, where $\pi_i = 1 - \frac{\binom{K-M_i}{m}}{\binom{K}{m}}$, $\hat{y}_i = \frac{M_i}{m_i}\sum_{j \in S_{1i}} y_{ij}$.

Ratio Type Estimator

The fifth estimator for the population mean, which uses the total number of units in the first stage sample as an auxiliary information, is actually a ratio type estimator, and is defined as

$$\hat{\mu}_{1.5} = \frac{\sum_{i \in S_1} \hat{y}_i}{\sum_{i \in S_1} M_i}$$

2.1.2 Simulation Result

In order to evaluate the performances of the estimators under the proposed design, we generate a set of data to run some simulations to compare the mean square error of the estimators with each other. Also, there are two comparable classical designs, overall simple random sampling and classical two-stage sampling design which are seen as two baselines in the simulation. The simulated sampling population is with the overall population size $M = 30116$, number of primary units $N = 300$, population mean $\mu = 843.32$, and population variance $\sigma^2 = 44007.38$. The data will also be used throughout this research.

The following graph shows the mean square error of the estimators under the sampling design using simple random sampling at each stage and the two classical sampling designs with different sample size of secondary sampling unit. As we can see from Figure 1, the mean square error of the estimator under overall simple random sampling is the lowest among all. The MSE of $\hat{\mu}_{1.5}$ is only higher than the one under overall simple random sampling. After all, there is no significant difference between other estimations.

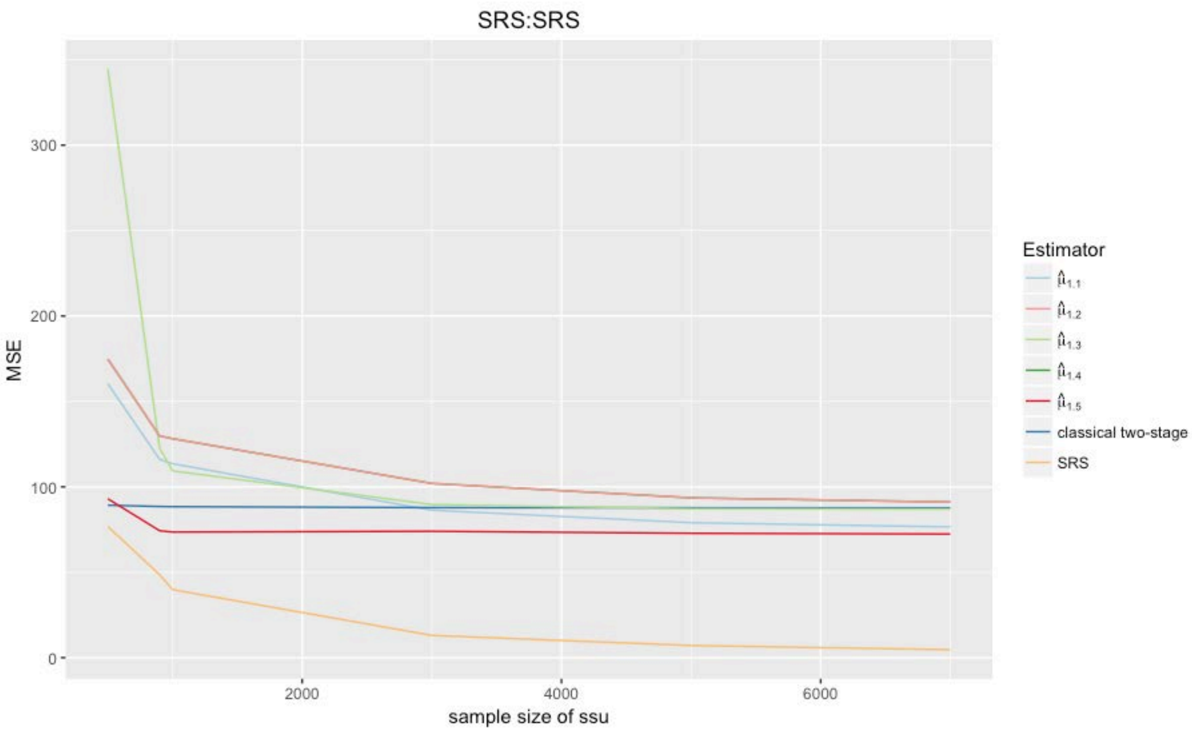

Figure 1: The mse of the proposed estimators (SRS/SRS) and the baselines with different ssu sample sizes.

2.2 SRS/Stratified

In practice, the sampling design which applies simple random sampling at each stage, seems too naive since every possible combinations of sample may be selected with equal probability; hence, the selected sample may not have similar data structure to the population. In order to select a better "representative" sample, another modified sampling design which uses simple random sampling and stratified random sampling as the first- and the second-stage sampling design respectively are introduced in following.

Assume that the population is partitioned into H strata. Let N and M denote the number of primary and secondary units in the population respectively. Denote M_{hi} as the number of units in the ith psu intersecting by the hth stratum, and $M_h = \sum_{i=1}^{N} M_{hi}$ as the number of units in the hth stratum for $i = 1, ...N$, $h = 1, ..., H$. The value of variable of interest of jth ssu in the ith psu intersecting with the hth stratum is denoted as y_{hij}, and $\sum_{j=1}^{M_{hi}} y_{hij} = y_{hi}$. The total value of variable of interest of psu is denoted as y_i, where $y_i = \sum_{h=1}^{H} y_{hi}$, while the population total is τ, where $\tau = \sum_{i=1}^{N} y_i$. The overall population mean is denoted as μ, while the standard deviation is denoted as σ.

First, select n primary units by simple random sampling without replacement as the first-stage sample. Denote S_1 as the set of primary units selected in the first stage sample, and K as the number of the secondary units in the first-stage sample. Next, integrating the secondary units in which the selected primary units, as the population for the second-stage sampling. Then grouping the units by the corresponding stratum to which each of the unit belongs, that is decided beforehand. Finally, for each of the stratum, select m_h secondary sample units by simple random sampling without replacement, where $m = \sum_{h=1}^{H} m_h$ is the sample size of demand. Moreover, let S_{hi} denote the set of sample in the ith primary unit adjacent to the hth stratum, and the set of the second-stage sample is denoted as S_2.

2.2.1 Estimations

Two estimators are given to estimate the population mean base on the modified sampling design using stratified sampling in the second stage. Consider making the overall population mean estimation by obtaining an unbiased estimate of the total of the first-stage sample in the beginning, a desired unbiased estimate of the parameter of interest can be derived then, as simple random sampling is applied in the first stage.

For the first estimator, estimates the total of the first-stage sample, $\sum_{i \in S_1} y_i$ in first place, then estimating the overall population total, τ, accordingly, consequently dividing the overall population total estimate by the overall population size M as the estimate of the overall population mean.
The estimator is defined as

$$\hat{\mu}_{2.1} = \frac{1}{M} \frac{N}{n} \sum_{i \in S_1} \sum_{h=1}^{H} \frac{M_h}{m_{hl}} \sum_{j \in S_{hi}} y_{hij}$$

The second estimator estimates the total of the first-stage sample, $\sum_{i \in S_1} y_i$, in the first, then dividing it by the number of secondary units in the first-stage sampling population, K, as the estimate of the overall population mean. Actually, the second estimator is also a population mean of the first-stage sampling population, and the estimator is defined as

$$\hat{\mu}_{2.2} = \frac{1}{K} \sum_{i \in S_1} \sum_{h=1}^{H} \frac{M_h}{m_{hl}} \sum_{j \in S_{hi}} y_{hij}$$

Since the estimator is also a sample ratio, hence, it can be seen as a ratio type estimator.

2.2.2 Simulation Result

The following graph (Figure 2) shows the MSE of the two proposed estimators under the modified sampling design using simple random sampling/ stratified sampling design in the first/ second second stage. The overall simple random sampling, classical two-stage sampling and overall stratified sampling (proportional/ Neyman allocation) designs are also taken as baselines to evaluate the performance of the proposed modified sampling design. The MSE of $\hat{\mu}_{2.1}$ is in fact, nearly the same as the one under classical two-stage sampling. The performance of $\hat{\mu}_{2.2}$ is better than $\hat{\mu}_{2.1}$ and the one under classical two-stage sampling design.

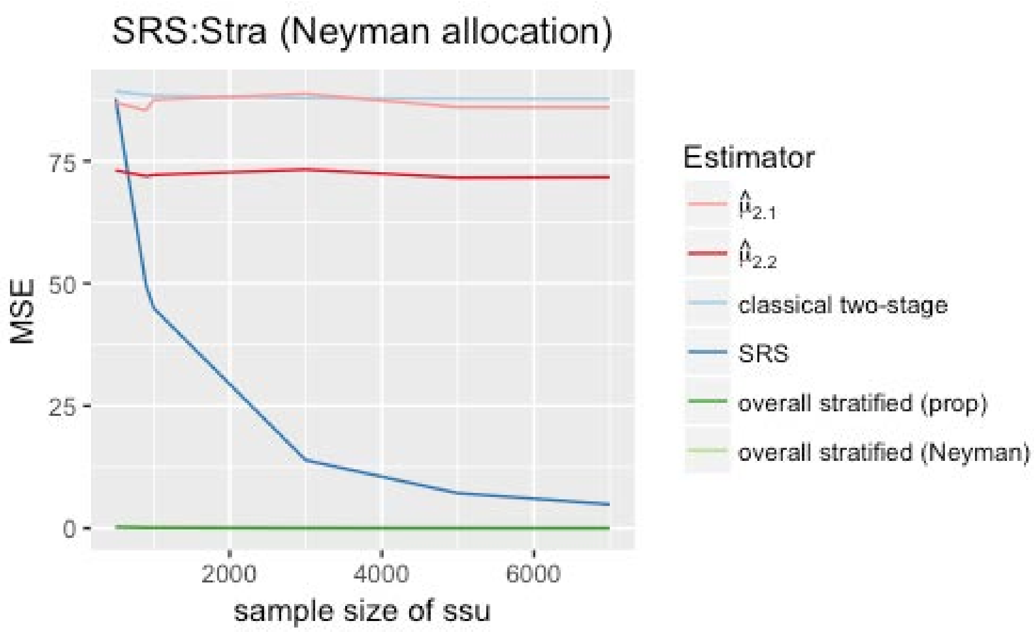

Figure 2: The mse of the proposed estimators (SRS/Stratified) and the baselines with different ssu sizes.

2.3 Improved First-Stage Sampling Design

The original purpose to apply the stratified sampling in the second stage design, is to take advantage of stratified sampling, so the sub-population can be aggregate properly, moreover, the estimation precision can be improved in the second stage of stratified sampling. As a matter of fact that the variance of estimator is composed of two parts, one is due to the between primary unit variance in the first stage and the other is the within stratum variance in the second stage, where the first one is of the major portion comparing to the second one. Hence, it is not surprise that the improvement is not significant when applying the stratified sampling in the second stage.

In contemplation of making improvement in the first-stage sampling, we proposed two more sampling design which applies stratified and systematic sampling in the first stage respectively. On the other hand, the inclusion probability of each primary unit in the previous two sections where simple random sampling is applied, are identical. For a better comparison with former sections, and problem simplification, both of the two sampling designs are made with equal inclusion probability in the first stage sampling. In order to have the inclusion probability of each primary unit preserve identical, using *proportional allocation* for the stratified sampling and assigning equal inclusion probability for the systematic sampling in the first stage so that the same estimators as in section 2.2 can be applied as well.

Consider partitioning the primary units into L strata by the ranking of y_i , called the first-stage strata, and dividing the secondary units of the population into H strata, which are the second-stage strata, by y_{ij} value rank. Let N and M denote the number of primary and secondary units in the population correspondingly. Denote N_l as the number of primary units in the lth first-stage strata, for $l = 1, 2, ..., L$. Let M_{hi} be denoted as the number of units in the ith psu intersecting with the hth stratum, $M_h = \sum_{i=1}^{N} M_{hi}$ as the number of units in the hth stratum and $M_i = \sum_{h=1}^{H} M_{hi}$ as the number of secondary units in the ith primary unit for $i = 1, ...N$, $h = 1, ..., H$, where $\sum_{h=1}^{H} M_h = \sum_{i=1}^{N} M_i = M$. The value of variable of interest of jth ssu in the ith psu intersecting with the hth stratum is denoted as y_{hij}, and $\sum_{j=1}^{M_{hi}} y_{hij} = y_{hi}$. The total value of variable of interest of psu is denoted as y_i, where $y_i = \sum_{h=1}^{H} y_{hi}$, while the population total is τ, where $\tau = \sum_{i=1}^{N} y_i$. The overall population mean is denoted as μ, while the standard deviation is denoted as σ. In the first-stage sampling, select n out of N primary units in the population.

Stratified/Stratified

In order to maintain the property of equal inclusion probability of the primary units as the same as it is in section 2.1 and 2.2. The proportional allocation is used in the first stage sampling throughout this section. Hence, the allocated sample size of the lth first-stage stratum is

$$n_l = \frac{nN_l}{\sum_{l=1}^{L} N_l}$$

Systematic/Stratified

Selecting a random starting point then the other $n - 1$ members are chosen subsequently, where the inclusion probability of each primary unit are equal.

Second-Stage Sampling Design

After the first-stage sample is selected, integrating the secondary units within the selected primary units as the second-stage sampling population. Consequently, apply stratified sampling design with Neyman allocation in the second stage.

The sample size of the hth stratum is defined as

$$m_h = \frac{mM'_h\sigma'_h}{\sum_{h=1}^{H} M'_h\sigma'_h}$$

Then randomly select m_h secondary sampling units within the hth stratum.

2.3.1 Estimations

Followings are the estimations for the population mean associated to the sampling design mentioned above the section.

- $$\hat{\mu}_{k.1} = \frac{1}{M}\frac{N}{n}\sum_{i\in S_1}\sum_{h=1}^{H}\frac{M'_h}{m_h}\sum_{j\in S_{hi}} y_{hij}$$
- $$\hat{\mu}_{k.2} = \frac{1}{K}\sum_{i\in S_1}\sum_{h=1}^{H}\frac{M'_h}{m_h}\sum_{j\in S_{hi}} y_{hij}$$

where the index $k = 3/4$ indicates that stratified/ systematic sampling design is applied in the first stage.

2.3.2 Simulation Result

Figure 3, 4 show the mean square error of the estimators under the modified sampling design using stratified sampling and systematic sampling design in the first stage. The comparable classical designs, which are seen as baselines, used here are overall simple random sampling, classical two-stage sampling and overall stratified sampling design (proportional/ Neyman allocation). As we can see from Figure 3, the mean square error of $\hat{\mu}_{3.1}$ and $\hat{\mu}_{3.2}$ are lower than both of the estimators under overall simple random sampling and classical two-stage sampling design when the secondary sampling unit size is not greater than 4,500. Further more, the mean square error of $\hat{\mu}_{3.1}$ is very close to the one under overall stratified sampling design. It is quite an obvious improvement when applying stratified sampling design in the first stage rather than simple random sampling. Similar results can be found in Figure 4. It can be concluded that the improved first-stage sampling designs with some proper estimation can surly make a better estimate of the population mean.

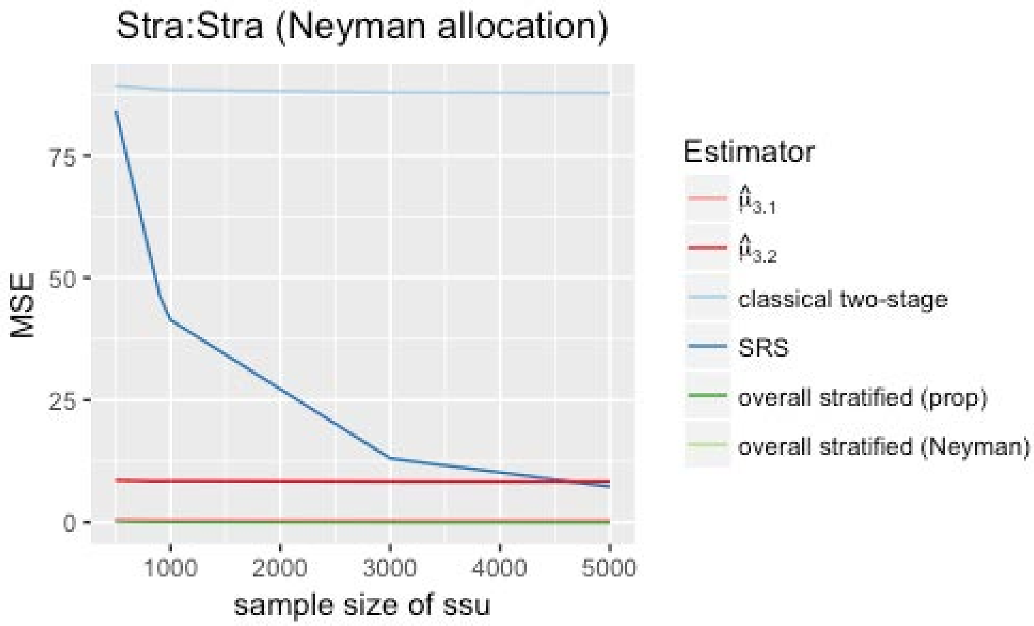

Figure 3: The mse of the proposed estimators (Stratified/Stratified) and the baselines for different ssu sizes.

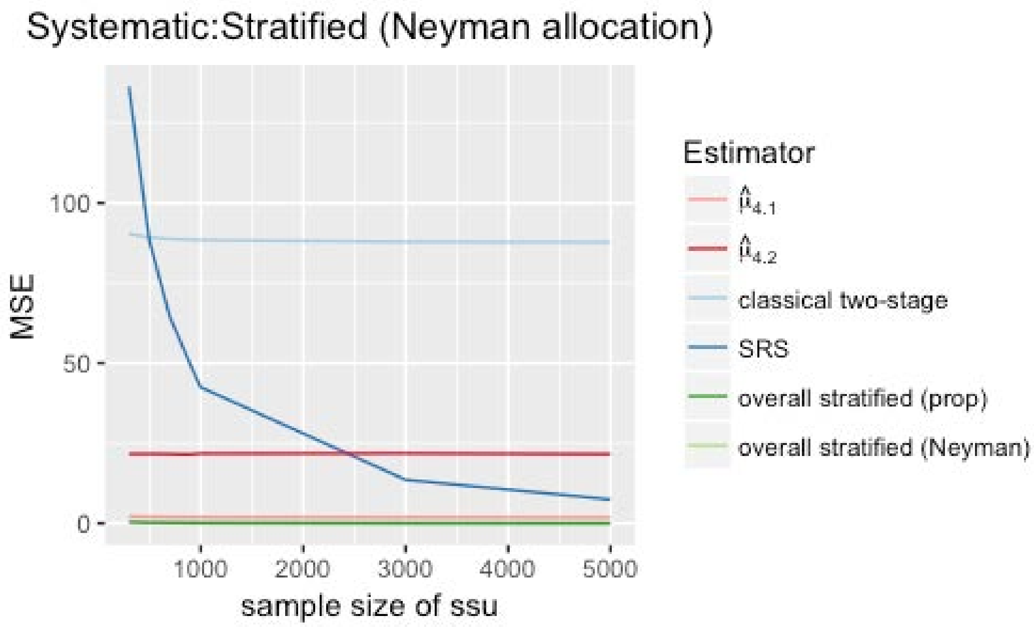

Figure 4: The mse of the proposed estimators (Systematic/Stratified) and the baselines for different ssu sizes.

3. Case Study

This section aims to study the modified two-stage sampling design to which different probability design combinations are applied.

The data used in this research is the 2015 Taiwanese Agriculture Census data. The total population size of agriculture and animal husbandry in Taiwan area is 781,518 and the agricultural gross income is chosen to be the population variable of primary interest. Since some of the farm households are not active, such the households with small agriculture scale and/or all the household agriculture workers are older than 65 years old. Therefore, in order to comprehend the households which are able to dedicate in the further development of agriculture industry, we define the target population to be the farm households whose annual gross income are between 200,000 and 50 millions New Taiwan dollars (NTD) and at least one household member under the age of 65 is currently engaged in the agriculture work. Such a target population is referred as ”primary farm households”.

Consequently, the target population size of the census data is 217,747 with 319 townships, which is refer to the primary sampling unit. To demonstrate the simulation, we determine the primary unit sample size to be 200, the secondary unit sample size to be 1600, and simulate 1000 times for each design.

The results are shown in Table 1. As we can see, the result are similar when using the real census data and the pseudo data for the simulation. Among all the estimators under the proposed modified sampling designs, $\hat{\mu}_{4.1}$ and $\hat{\mu}_{3.1}$ has the lowest mean square error. It shows that if a proper sampling design is applied in the first stage, the performance would be greatly improved. Though the MSE of $\hat{\mu}_{4.1}$ and $\hat{\mu}_{3.1}$ is still higher than that of the overall stratified sampling design, in consideration of the sampling cost, the intersected psu number under stratified sampling design is 280, which is a lot higher than the sample size of psu under the proposed design. Therefore, the sampling cost can be saved significantly when applying the proposed design.

Table 1: Comparison of the MSE of the proposed estimators

Estimator	MSE
Stratified Sampling	343.72
Classical Two-Stage Sampling	12297.24
Simple Random Sampling	18519.09
$\hat{\mu}_{1.5}$	6886.08
$\hat{\mu}_{2.1}$	3737.19
$\hat{\mu}_{2.2}$	1607.04
$\hat{\mu}_{3.1}$	849.18
$\hat{\mu}_{3.2}$	1103.94
$\hat{\mu}_{4.1}$	678.91
$\hat{\mu}_{4.2}$	1025.07

4. Conclusion

The purpose of the modified two-stage sampling studied in this research is to balance the estimation precision and the survey cost. In this research, we apply several probability design combinations to the modified two-stage sampling, and associated candidate estimators are proposed and examined also. The performance, from the perspectives of both of the estimation precision and survey cost, of the proposed modified two-stage sampling designs are evaluated and compared to the comparable classical designs, such as simple random sampling, two-stage sampling and stratified sampling designs.

In section 2.1, consider the modified two-stage sampling design with simple random sampling at each stage. Five estimators are proposed to estimate the population parameter of interest based on the design. The results indicate that the estimator of the ratio type, $\hat{\mu}_{1.5}$, is the most precise among the five candidates. In addition, this modified two-stage sampling design together with $\hat{\mu}_{1.5}$ is better than the classical two-stage sampling design with associated unbiased estimator under equal primary unit sample size. On the other hand, its performance is always worse than the classical simple random sampling and stratified sampling in terms of the mean square error under the same secondary unit sample size, nevertheless, the associated primary unit sample size of the proposed design is always much less than the above mentioned two classical sampling designs, hence the sampling cost can be reduced.

In order to improve the sample selection of the second stage, simple random sampling design is substituted by stratified sampling in section 2.2. Estimators associated with the sampling designs are proposed, where the second estimator, $\hat{\mu}_{2.2}$, is better than the classical two-stage sampling in terms of lower mean square error. Hence $\hat{\mu}_{2.2}$ is suggested. For the same reason as in section 2.1, the estimators in section 2.2 are less precise than the classical simple random sampling and classical stratified sampling in terms of lower mean square error as well.

For a better representative sample selection, not only the second-stage sampling but also the first-stage sampling to which we try to make effort. Despite of simple random sampling, stratified sampling and systematic sampling are applied to the first stage in section 2.3. For a better comparison with previous chapters and problem simplification, using *proportional allocation* for the stratified sampling and assigning equal inclusion probability for the systematic sampling in the first stage so that the same estimators as in section 2.2 can be used as well. The simulation results indicate that $\hat{\mu}_{k.1}$ are the second-best estimator of all in terms of lower mean square error for $k = 3, 4$.

The modified two-stage sampling scheme proposed in this research is usually better than the classical two stage design in terms of lower mean square error, and the sampling cost is expected to be much less than the classical simple random sampling/stratified sampling design since the primary unit sample size intersected to these two classical designs are much more than which of the proposed modified two-stage design. Therefore, we conclude that the sampling scheme proposed in this research can successfully balance the estimation precision and sampling cost.

To construct a complete statistical inference, it is necessary to have the variances and the associated variance estimations of the estimators proposed in this research and related research is currently under investigation. However, the closed forms of the exact or approximated variances of most of the estimators might not available, therefore certain alternative estimation method, such as the Bootstrap/Jackknife method, will also be considered. On the other hand, rather than a single sampling design, the modified two-stage sampling designs described in this this research are in fact designs of a family of a sampling scheme. Different combinations of first and second stage designs can be used as required according to the survey situation. However, the first-stage sampling design considered in this research are restricted to the designs with equal sampling weights, so that the property of unbiased or asymptotically unbiased of the associated estimators can be preserved. It is also sensible to utilize other sampling designs in the first-stage sampling for the selection of the first-stage sample of primary units to further improve the estimation, and also the flexibility of the proposed sampling scheme. How to properly estimate the population quantity of interest under a general probability first stage sampling design is hence of both of theoretical and practical interest, and related research is also under investigation.

REFERENCES

Dalenius, T. and Hodges, J. (1959), "Minimum variance stratification," *Journal of American Statistical Association*, 54, 88–101.

Fattorini, Lorenzo (2006), "Applying the Horvitz-Thompson criterion in complex designs: A computer-intensive perspective for estimating inclusion probabilities," *Biometrika*, 93, 269–278.

Lohr, Sharon L. (2010), *Sampling: Design and Analysis* (2nd ed.), Australia, Boston, Mass: Brooks/Coke.

Thompson, S.K. (2012), *Sampling* (3rd ed.), Hoboken, NJ: John Wiley & Sons.