Thesis Title	Estimation of Population Total from Simple Random Samples Containing Some Very Large Units Using Regression Estimator
Name	Miss. Anotai Trevanich
Thesis Advisor	Associate Professor Suchada Kiranandana, Ph.D.
Department	Statistics
Academic Year	1986

ABSTRACT

The problem considered in this thesis is the estimation of the population total from a simple random samples containing some very large units which are actually present in the population. The objective of this study is to investigate the property of suggested estimators for population total $(\hat{Y}_k$; k = 1,2,3) which are designed to increase efficiency by changing weight for very large observations and use of an auxiliary variate X that is linearly related with Y. Then compare them with the simple mean estimator $\hat{Y}_o = N\bar{y}$ and the suggested estimators by Michael & Kadaba $(\hat{Y}_{mkt}$; t = 1,2,3,4) using the relative efficiency.

The data were obtained through simulation using Monte Carlo technique. A computer program was designed to calculate the relative efficiency of $\hat{Y}_k$ or $\hat{Y}_{mkt}$ to $\hat{Y}_o$ for the case of Lognormal and Gamma distributions, with correlation coefficient between Y and X being 0.1,0.3,0.5,0.7 and -0.1,-0.3,-0.5,-0.7. Population of sizes 500 and 1000 were used, having percents of very large observations in population of size 500 as 1.8% 2.8% and 3.2% and in population of size

1000 as 1.8%, 2.8% and 3.3%. The sample sizes considered were 50, 100 and 200 and the number of very large observations in the sample started from 2 to the number of very large observations in population. For each predicament of the experiment was repeated 100 times in population size 500 and 50 times in population size 1000.

The Results of the study as classified by inference forms can be summarized as follows.

For conditional inference, when the number of very large observations in the population is known, the relative efficiency of $\hat{Y}_1$ to $\hat{Y}_o$ is highest and decreases when the percent of very large observations in the population increases. If variable Y is linearly related with variable X or the percent of very large units in the sample increases, then $\hat{Y}_1$ is obviously more relative efficiency than $\hat{Y}_{mk4}$. Particularly when the correlation coefficient between Y and X is 0.7 or -0.7, the relative efficiency of $\hat{Y}_1$ to $\hat{Y}_o$ is about 2 times of the relative efficiency of $\hat{Y}_{mk4}$ to $\hat{Y}_o$. For the case when the number of very large observations in the pupulation is unknown, the relative efficiency of $\hat{Y}_3$, $\hat{Y}_{mk1}$ or $\hat{Y}_{mk3}$ to $\hat{Y}_o$ decrease as the percent of very large observations in the population increases. But the relative efficiency of $\hat{Y}_2$ to $\hat{Y}_o$ increases and the relative efficiency of $\hat{Y}_{mk2}$ to $\hat{Y}_o$ decreases when the percent of very large samples is small. However, when the linear relationship between Y and X is high or the percent of very large units in sample is large, the relative efficiency of $\hat{Y}_3$ to $\hat{Y}_o$ is the highest.

For unconditional inference, when the number of very large observations in the population is known, the relative efficiencies of the estimators to $\hat{Y}_o$ are the same as in the case of conditional inference. When the number of very large observations in population increases, the relative efficiency of $\hat{Y}_1$ to $\hat{Y}_o$ is highest. However, the relative

efficiency decreases for Lognormal distribution but increases for Gamma distribution as the percent of very large observations in the population increases. Particularly if the linear relationship between Y and X becomes stronger, the relative efficiency of $\hat{Y}_1$ to $\hat{Y}_o$ is obviously higher than $\hat{Y}_{mk4}$. For the case when the number of very large observations in the population is unknown, as in the conditional inference, the relative efficiency of $\hat{Y}_i$; i = 2,3, or $\hat{Y}_{mkj}$; j = j = 1,2,3 to $\hat{Y}_o$ decrease as the number of very large observations in the population increases. When the linear relationship between Y and X is low, the relative efficiency of $\hat{Y}_{mk2}$ to $\hat{Y}_o$ is the highest. However, when the linear relationship between Y and X is high, the relative efficiency of $\hat{Y}_2$ to $\hat{Y}_o$ is the highest.