

Improving Sampling Efficiency for the Annual Survey of Services

By

Andy Chiang, Cui Hui Min, Neo Siok Hoong and Tay Boon Shang
Business Statistics Division
Singapore Department of Statistics

Introduction

The Annual Survey of Services (AS) collects vital business data used for the study of financial structure and performance of the services industries and the compilation of Singapore's economic statistics. The survey sample size for the Annual Survey of Services has grown exponentially since 1998. This article discusses the recent improvement in sampling design of the AS that will be implemented for reference year 2007.

Optimal Sampling for Business Surveys

In business surveys, populations have skewed distributions where a small number of units account for a large

share of the total of the study variable (y). Following sampling theory, it is more efficient to employ a stratified sampling design where y is used as a size stratification variable by constructing boundaries in the range of y . In business surveys conducted by the Singapore Department of Statistics, operating receipts is frequently used for such size stratification as it is positively correlated to several key variables of interest such as operating expenditure, employment, value added and profit.

Survey statisticians have developed various approaches for optimal size stratification. Horgan (2006) and Kozak & Verma (2006) reviewed the best methods, which include the Hidiroglou (1986) and Lavallée-Hidiroglou (1988) methods. These two methods adopt a common strategy of stratifying

the population into two or more size strata. The firms in the large size stratum are selected with certainty whereas the firms in the smaller size strata are sampled using simple random sampling without replacement. Both methods produce optimal sample size and allocation for a given level of precision¹. The difference is that the Hidioglou (1986) method allows for two size strata (large and small) while the Lavallée-Hidioglou (1988) method provides two or more² size strata.

Sampling Design of Annual Survey of Services

Before 1997 : Stratified Systematic Sampling

Before 1997, the AS used a stratified systematic sampling design. All establishments were stratified by SSIC³ industry and operating receipts as follows. Within each SSIC industry, establishments belonging to enterprises with multiple branches or divisions were selected with

certainty. For the remaining establishments, they were selected based on their operating receipts relative to a pre-determined cut-off value. Those above the cut-off value were selected with certainty while those below were selected with a probability of one in ten systematically. Based on this method, the sample size grew proportionately with the population in the survey frame from 22,000 for 1986 to 27,000 establishments for 1996.

1997 to 2006 : Stratified Random Sampling Using Hidioglou Method

The steady increase in sample size has resulted in increased respondents' burden, longer processing time before survey results become available as well as the need for increasingly scarce manpower resources. To manage the situation, a sampling design review⁴ was conducted and resulted in the adoption of the Hidioglou (1986) method for stratified sampling of enterprises with coefficient of variation of 10 per cent at SSIC 5-digit industry level. This brought about a reduction of sample size to 16,400

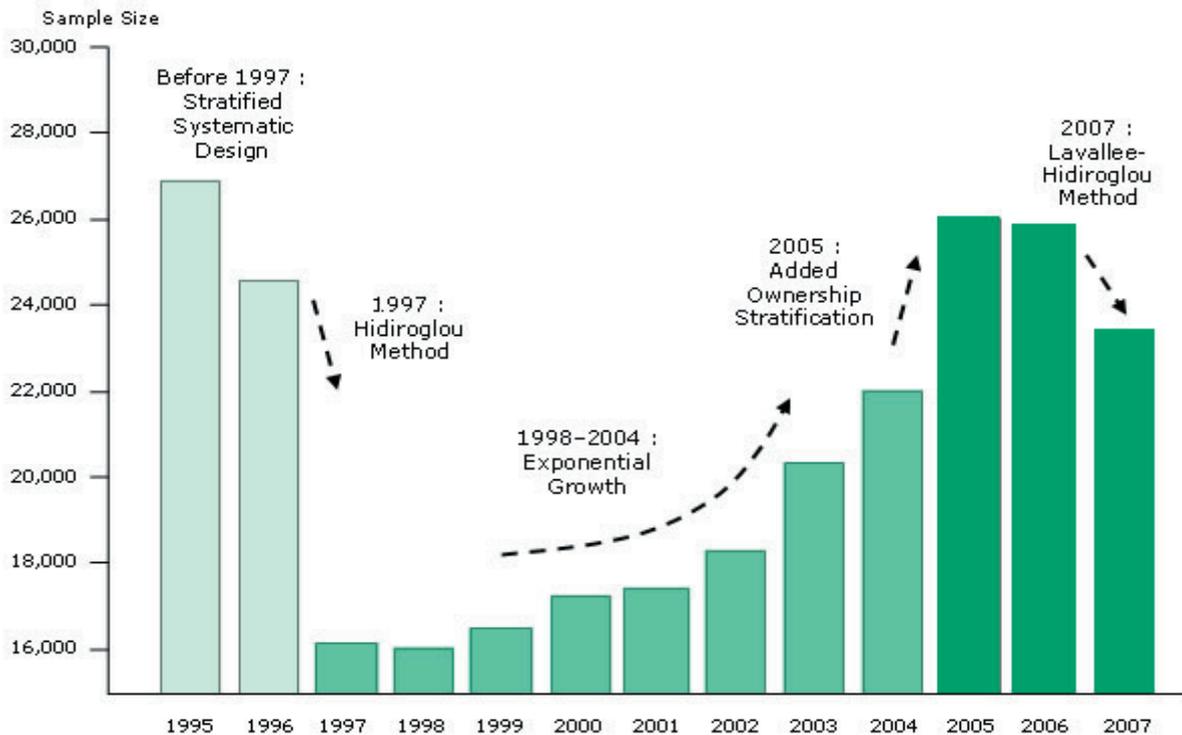
¹ Precision is measured by a coefficient of variation which is expressed in per cent format. A higher (lower) coefficient of variation corresponds to lower (higher) precision for the survey estimate.

² A three size strata may be labeled as large, medium and small respectively.

³ Singapore Standard Industrial Classification (SSIC) is the official classification of economic activities undertaken by economic units and used for censuses of population, household and establishment surveys and administrative records. SSIC was established in 1958 and last revised in 2005.

⁴ Singapore Department of Statistics (1999). New sample design for the annual surveys of commerce and services. *Occasional Paper on Business Statistics*.

CHART 1 1995-2007 ANNUAL SURVEY OF SERVICES DESIGN SAMPLE SIZE



establishments for 1997 (see Chart 1). Subsequently, due to increases in the size and variability in the business population, AS sample size grew exponentially from 16,400 for 1997 to 22,000 establishments for 2004. Since 2005, in order to produce more precise estimation at ownership level, AS population was stratified by industry and ownership (local and foreign). Each industry-ownership stratum was size-stratified by operating receipts using the Hidiroglou method. The addition of ownership stratification raised sample size to 26,000 for 2005 as the number of strata before Hidiroglou size stratification was doubled.

Improving Sampling Efficiency Using Lavallée-Hidiroglou Method

We will use two examples to illustrate how the new sampling design using Lavallée-Hidiroglou method for size stratification improves over the old sampling design using Hidiroglou method. We will examine two populations (A and B) that are representative of large and small populations in services industries. Chart 2 shows the skewness in the distribution of operating receipts, which is typical in services industries.

Size Stratification in Population A

First, consider population A from a transport industry in Table 1. Assuming 10 per cent coefficient of variation for estimating total industry operating receipts, the Hidiroglou method will stratify the population by operating receipts into two size strata : large (over S\$5.3m) and small (under S\$5.3m). Lavallée-Hidiroglou method stratifies the population more efficiently into three operating receipts size strata : large (over S\$11.2m), medium (between S\$1.3m to S\$11.2m) and small (less than S\$1.3m). By having medium size strata, fewer firms are sampled : the medium and small size strata had a combined sample of 29 units, compared with 65 units for the small size stratum in the old design. Together with the reduction in large

(take-all) units due to higher cut-off for operating receipts, the overall sample size is reduced from 110 in the old design to only 44 in the new design. This represented a 60 per cent reduction in sample size for the same level of 10 per cent coefficient of variation.

Size Stratification in Population B

Next, we examine population B from a wholesale industry in Table 2. Hidiroglou method has two size strata : large (over S\$3.1m) and small (under S\$3.1m) while Lavallée-Hidiroglou method gives three : (large = over S\$13.1m, medium = S\$1.8m to S\$13.1m and small = under S\$1.8m). As the industry distribution of operating receipts is highly skewed (see Chart 2), the Hidiroglou cut-off of S\$3.1m is too low to be efficient. Consequently,

CHART 2 DISTRIBUTION OF OPERATING RECEIPTS IN POPULATION A AND B

Table 1

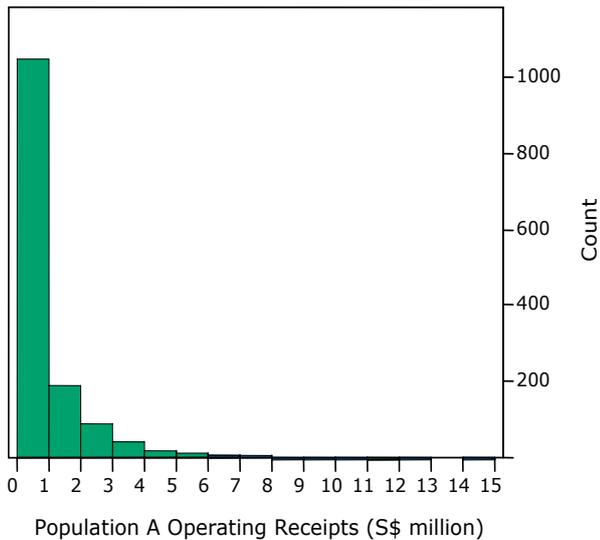


Table 2

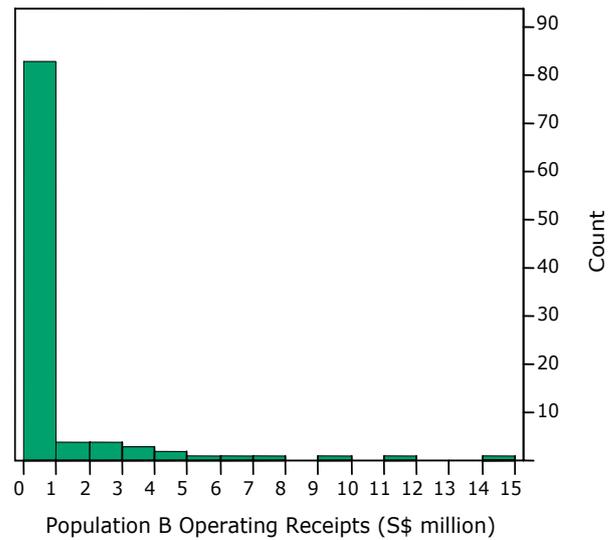


TABLE 1 POPULATION A CUT-OFF VALUES AND SAMPLE SIZES

	Size Strata	Cuf-Off Value (S\$) [1]	Population Size [2]	Sample Size [3]	Sampling Fraction (%) [4]
	Total		1,517	110	7
<i>Hidiroglou Method</i>	Large	5,296,186	45	45	100
	Small	0	1,472	65	4
	Total		1,517	44	3
<i>Lavellée-Hidiroglou Method</i>	Large	11,188,698	15	15	100
	Medium	1,307,368	273	17	6
	Small	0	1,229	12	1

TABLE 2 POPULATION B CUT-OFF VALUES AND SAMPLE SIZES

	Size Strata	Cuf-Off Value (S\$)	Population Size	Sample Size	Sampling Fraction (%)
	Total		103	16	16
<i>Hidiroglou Method</i>	Large	3,081,631	13	13	100
	Small	0	90	3	3
	Total		103	13	13
<i>Lavellée-Hidiroglou Method</i>	Large	13,057,417	5	5	100
	Medium	1,832,494	11	5	45
	Small	0	87	3	3

[1] Cut-off value refers to the lower boundary value of operating receipts for the size strata.

[2] Population size refers to the number of units in the respective strata in the population.

[3] Sample size refers to the number of units in the sample for the respective strata.

[4] Sampling fraction refers to the percentage of units in the population selected into the sample, i.e. [4] = [3] / [2] x 100%.

13 firms are treated as large and to be selected with certainty. In contrast, the Lavallée-Hidiroglou method requires only the top 5 large firms (with over S\$13.1m in operating receipts) to be selected with certainty and samples another 5 out of 11 possible medium firms. Both methods sample only 3 out of 87-90 small firms. The new design using Lavallée-Hidiroglou method achieves a 19 per cent reduction in sample size while maintaining the same level of precision of 10 per cent coefficient of variation.

Final Sample Design for Annual Survey of Services 2007

The Lavallée-Hidiroglou size-stratification for operating receipts with three size strata will be implemented in the AS for reference year 2007. While there

is potential to reduce sample size dramatically (as high as 60 per cent in the example for Population A) for fixed coefficient of variation at 10 per cent, part of the reduction in sample size will be used to trade-off for improvement in estimation precision at broader groupings such as SSIC 2-digit industry divisions⁵ via a reduction in coefficient of variation from 10 per cent to 8 per cent (typically) and as low as 5 per cent in some cases.

Overall, the new sampling design will reduce total sample size by about 10 per cent (23,500 in 2007 compared to 26,000 in 2006) while improving estimation precision: SSIC 5-digit industry coefficient of variation by at least 2 per cent (5-8 per cent for 2007 compared to 10 per cent for 2006).

References

Hidiroglou, M.A. (1986). The construction of a self-representing stratum of large units in survey design. *The American Statistician* 40(1), 27-31.

Horgan, J.M. (2006). Stratification of skewed populations: a review. *International Statistical Review* 74(1), 67-76.

Kozak, M., Verma, M.R. (2006). Geometric versus optimization approach to stratifica-

tion: comparison of efficiency. *Survey Methodology* 32(2), 157-163.

Lavallée, P. & Hidiroglou, M. (1988). On the stratification of skewed populations. *Survey Methodology* 14, 33-43.

Singapore Department of Statistics (1999). New sample design for the annual surveys of commerce & services. *Occasional Paper on Business Statistics*.

⁵ SSIC 2-digit industry division refers to a broad grouping of 5-digit industries with the same first 2-digit code. For instance, "51" retail trade industry division contains the industries ranging from "51101" (Non-specialised retail trade in department stores) to "51609" (Other non-store retail sale). For AS 2007, coefficient of variation is set at 5 per cent for estimation at SSIC 2-digit industry division level.