# *AN APPROACH TO THE DESIGN OF SAMPLING PROCEDURES FOR ECONOMIC SURVEYS IN TRINIDAD AND TOBAGO*

Christopher Martin Clarke

*A Paper presented at the XXV Conference of the Regional Programme of Monetary Studies Econometric Modelling Seminar on the 17-20 November 1993*

Research Department
Central Bank of Trinidad and Tobago

# AN APPROACH TOTHE DESIGN OF SAMPLING PROCEDURES FOR ECONOMIC SURVEYS IN TRINIDAD AND TOBAGO

## I Introduction

In this note, we examine some of the factors which impact on the design of sampling procedures for economic surveys in small open economies like Trinidad and Tobago. For the purposes of this paper, economic surveys have been defined to include:

> "... surveys of financial, industrial, commodity, employment, capital expenditure and taxation statistics collected on a monthly quarterly, annual or occasional basis". [Colledge and Lussier (1987)].

The paper attempts to examine the arguments for the use of judgment methods of sampling in the context of economic surveys in Trinidad and Tobago. As an alternative, a stratified sampling strategy which makes use of appropriate 'cut-off' rules for defining stratum boundaries is proposed as being more appropriate for such surveys. This method, which is based on the principle of randomization, combines the benefit of allowing for the exercise of some degree of personal judgment in the selection process, while avoiding the problem of a completely model determined design.

The paper is divided into four sections. In the next section we examine the major sampling problems which confront economic survey practitioners in Trinidad and Tobago. We critically examine the remedy usually proposed judgment selection. In section II, we examine the cut-off stratified sampling method and show how it can be utilized for economic surveys. The technique is illustrated with the specific example of the proposed Corporate Financing Survey to be undertaken by the Central Bank in 1991.
While this technique overcomes some of the major weaknesses and limitations of model dependent sampling, several critical concerns remain to confront the survey practitioner. These

are discussed briefly in section III. In the final section of the paper we present our summary and conclusions.

## II Economic Surveys in Trinidad and Tobago

In Trinidad and Tobago, the two major agencies with the legal authority to conduct economic surveys are the Central Statistical Office (CSO) and the Central Bank of Trinidad and Tobago (CBTT). Traditionally, the latter agency confined it's attention to the financial system while the CSO had major responsibility for real-sector economic surveys in the country. However, within recent times, the Bank has attempted to become more involved in the compilation of real-sector data sets, primarily in the areas of international payments and national accounts, currently undertaking economic surveys to assist in the support these activities.

In Table I, we attempt to show the major official economic surveys undertaken by both institutions, the periodicity of the surveys undertaken, the sampling frame utilized and the sampling methods used. An examination of this table reveals a preference for either full enumeration or for judgment methods of selection. In some cases, full enumeration has been justified on the basis of recommendations from international statistical agencies charged with maintaining the quality, integrity and comparability of data sets generated in different member countries. Hence, for example, IMF (1977), recommended that official estimates for balance of payments should be based (whenever possible) on a full enumeration of sectors. In other cases for example when the survey population is small, as in the instance of the domestic commercial banking system (i.e. eight commercial banks), the use of sampling techniques is unnecessary and even dangerous.

**TABLE 1**
**OFFICIAL ECONOMIC SURVEYS IN TRINIDAD AND TOBAGO, 1992**

| Survey | Responsible Agency | Frequency | Sampling Method | Frame Utilized |
|---|---|---|---|---|
| Annual Survey of Establishments | C.S.O. | Annual | Judgement selection/full enumeration for some sectors e.g. Petroleum | Register of Establishments |
| Survey of Retail Prices Building Materials | C.S.O. | Quarterly | Judgement selection of areas and establishments | -do- |
| Survey of Retail Sales | C.S.O. | Quarterly | Establishments stratified by level of sales previous quarter/judgement selection | -do- |
| Survey of Domestic Production | C.S.O. | Quarterly | Judgement Selection | -do- |
| Survey of Minimum Wage Rates May and November | C.S.O. | Semi-annually | Judgement Selection | -do- |
| Survey of Employment and Wages | C.S.O. | Semi-annually May & November | Judgement Selection | CSO's Register of Jobs |
| Survey of Direct Foreign Industry | CBTT | Quarterly | Full enumeration | All firms with 25% or more foreign participation |
| Survey of Insurance Company | CBTT | Quarterly | Full enumeration | All Insurance companies Life and Non-Life |
| QGDP-Survey of Manufacture and Insurance | CBTT | Quarterly | Judgement Selection | C.S.O. estimates of Firm'sContribution to Value-added 1985 |
| Financial Returns System | CBTT | Weekly/monthly/ quarterly/ semi-annually | Full enumeration | All Banking and Non-Banking Financial Institutions |

In contrast, it is more difficult to justify the use of purposive sampling in the other surveys listed in Table 1. The most common argument for the use of purposive methods, in economic surveys is eloquently restated in Farrell, Najjar and Marcelle (1986). The authors in explaining the choice of a judgment sampling strategy in a survey of corporate financing undertaken by the CBTT in 1984, contended, that the appropriate sampling plan had to take into consideration:

> "... the uneven distribution of firms when classified by size and ownership .. and to ensure the selection of a cross section of large firms, as well as firms owned by government and foreigners ..."

Furthermore, they noted that the problem of choosing an acceptable sample was compounded by the small number of firms in many sectors. In such circumstances, Farrell et al. suggested that 'good' samples could only have been drawn by a sampling method which utilized the subjective judgments of the survey practitioners.

However, Farrell et al. also noted that when such procedures were adopted, they were faced with the critical problem of making inferences about the estimates obtained from such samples. The selection probabilities were unknown to the survey practitioners and it is not possible to make objective inferences about survey populations based on such a judgment sample. Of equal importance, is the fact that such techniques leave the survey practitioner open to the charge of introducing personal bias into the selection procedure. The danger of this accusation being leveled against the survey practitioner is particularly worrying when the survey in question produces a controversial result which is supposed to inform the policy decisions of the authorities. Moreover, secondary analysis, including the use of such formal analytical techniques as regression, correlation, discriminant and logic analysis, assume that data in question come from samples drawn with random methods. The use of such techniques in the context of judgment samples cannot be supported by statistical theory.

In reply to these criticisms practitioners in the field argue that these concerns are largely academic and that their extensive knowledge of the economy ensures against such bias. However, one must be extremely wary about accepting such arguments, as Kish (1965) warned, the history of the social sciences is replete with cases where so-called 'expert' knowledge failed to yield acceptable samples. Moreover, Hansen, Madow and Tepping (1979), in a critique of model dependent sampling methods, have noted that while such techniques can dramatically reduce the cost of sampling, the resulting estimators are extremely non-robust, if the super population model is miss-specified, then the resulting model based estimator is unreliable, biased and inefficient. As such, Hansen, Madow and Tepping (1978), have contended strongly that randomization is an necessary (but often not sufficient) prerequisite for ensuring the acceptability of official surveys.

This is not to say that judgment selection is completely without merit. Indeed, Cochran (1979) argued with respect to quota sampling, (a particular form of judgment sampling), that this technique performed as well as probability sampling on questions of attitude and opinion. Moreover, judgment sampling was cheap and easy to apply. Unfortunately, Cochran also noted that this technique produces unreliable estimates on the very characteristics that economic surveys have been usually designed to measure, such as income, occupation, sales, employment and other related variables.

As a final line of defense, the practitioners in the field have contended that the received survey sampling theory fails to provide them with cost effective techniques, which come to terms with the sampling problems peculiar to small open economies. They argue, that while stratified sampling is widely touted as the solution to the problems posed by highly skewed populations for economic surveys, the standard techniques require information on characteristics, usually with regard to measures of size, which are simply not available from standard economic survey frames. Furthermore, this technique fails to address the sampling issues which arise directly as a result of the small size, both of the potential survey population and of the domestic economy.

Small size has several implications for the design of sampling methods for economic surveys. Foremost in this respect is the fact that many industrial sectors may contain a relatively few firms, with production and sales concentrated in a few dominant entities [see Farrell and Crichton (1984)]. The extent of the problem is even more evident when we consider that the dominant form of ownership is the private limited liability company. These companies are often family owned concerns and management may often harbor deep suspicions about the intentions of official statistical agencies that are perceived to be operating on behalf of the Government. In such an environment, the information requirements of many official economic surveys may be considered by many firms to be too compromising, representing requests for closely guarded trade secrets. Anecdotal evidence seems to suggest that these perceptions may be most acute among the smaller firms in survey populations, but may not be confined exclusively to them.

Adding to the problem is a general lack of awareness among the survey audience of the importance of business surveys, either as tools for informing general economic policy making or as market research devices. Indeed, the comparative youth of the private market research industry, suggest that in the private sector, strategic decision making based on statistical analysis, is a relatively new idea. As such, respondents may typically regard economic survey response forms as no more than examples bureaucratic red tape.

The two official statistical institutions must also share some of the blame for the observed disinterest and the relative lack of sophistication of the survey audience. Forde (1989), in an examination of the adequacy of the statistical data base to inform the planning and monitoring mechanism, expressed serious concerns about the timeliness, periodicity, accuracy and relevance of socio-economic statistics in Trinidad and Tobago.

Finally, one has to consider the problem of respondent fatigue which is generally not regarded as a sampling issue. However, in the context of a small open economy, one cannot escape from the fact that the absolute size of the population for business surveys is small. The same firms tend to be repeatedly sampled for different purposes by both the CSO and the CBTT. Survey respondents may find this situation confusing and burdensome and this may serve to lower response rates. In such an institutional environment, response rates are generally low such, many practitioners contend that the only way to avoid the pitfalls discussed above is by using purposive methods of sample selection based on an expert knowledge of the survey population.

In any case, it is usually assumed that the alternate probability sampling strategy is bound to be costly in terms of the limited resources of official survey agencies. Moreover, it is conventional wisdom among practitioners in the field, that the existing expert samples yield results that are comparable to those derived from randomized designs. However, very little empirical evidence is adduced to support these claims. Indeed, the users of statistical data derived from such surveys continue to express serious concerns about the accuracy, coverage and general validity of such data.

While we agree that the smallness of the population for economic surveys, the highly concentrated industrial organizational structure, and the relative lack of sophistication among the survey audience may impose serious constraints on all aspects of general survey practice, including the choice of sampling methods, we are of the view that the continued use of judgment methods of sampling is unacceptable and that it is imperative that random sampling methods be adopted for official surveys. In the next section we explore how such a random sampling design can be applied to a practical economic survey.

## III Cut-off Rules For Stratified Sampling with a Self-Representing Stratum

It is well known that stratification improves the precision of estimators by dividing up a heterogeneous population into sub-populations or start, each of which are internally homogeneous in terms of the measurements obtained for the variables of interest. Indeed, Cochran (1977), demonstrates that the largest gains in precision are obtained when institutions vary widely in size and the variables of interest are closely related to the size of institutions. Moreover, Cochran (1977) shows that in terms of allocating a given sample size 'n' among different strata, an optimal strategy is to allocate more of the sample to a particular stratum if that stratum is larger, more variable or cheaper to sample. Furthermore, Hidiroglou and Srinath (1981), have demonstrated in the case of populations which exhibit a high degree of positive skewness, i.e. a few large units and many small units, that simple random sampling may lead to an overestimate of sample characteristics.

One approach to the problem of highly skewed populations, is to divide the survey population into two major strata: a *take-all* strata which contains the largest elements in the population and is surveyed entirely and a *take-some* strata which is sampled with simple random sampling. This technique is often referred to as *cut-off* sampling and was originally proposed by Hansen, Hurwitz and Madow (1953). Glasser (1962), suggested rules for delineating the boundary of take-all and take-some universe. However, these cut-off values were derived on the assumption that the optimum sample size was known in advance. More recently, Hidiroglou (1986), derived both exact and approximate cut-off rules for the more common situation where one tries to determine the optimum sample size when the desired level of precision of the estimators is known in advance. The Hidiroglou cut-off rule is derived in terms of an auxiliary variable which is highly correlated with the variables of interest and known in advance.

To detail the Hidiroglou cut-off rules, consider first an ordered population on N units, where 'y' is the stratifying variable (employment) assumed to be highly correlated with the variables of interest:

$$y_{(1)}, y_{(2)}, \ldots, y_{(N)},$$

with $y_{(i)} \le y_{(i+1)}$ for $i = 1,2,\ldots,N-1$. Of the population, $t$ members are designated "large" units and placed in the "take-all" universe and $N$-$t$ units placed in the "take-some" stratum, the total is

$$Y = \sum_{i=1}^{N-t} y_{(i)} + \sum_{i=N-t+1}^{N} y_{(i)}.$$

We wish to draw a sample of size $n(t)$, from the total population where n(t), is composed of $t$ large units and $n(t)$-$t$ small units selected by simple random sampling from the remaining $N$-$t$ units in the population. An estimator for the total Y will be given by

$$\hat{Y} = \frac{N-t}{n(t)-t} \sum_{i=1}^{n(t)-t} z_i + \sum_{i=N-t+1}^{N} y_{(i)},$$

where $z_{(i)}$ is defined to be a member of the take-some universe i.e. $y_{(1)} \le z_{(i)} \le y_{(N-t)}$ for $i = 1,2,\ldots,n(t)-t$. The variance of this estimator is given by

$$V(\hat{Y}) = \frac{N - t\{N - n(t)\}}{n(t) - t} S^2_{[N-t]}, \qquad (1)$$

where

$$S^2_{[N-t]} = \frac{1}{N-t-1} \sum_{i=1}^{N-t} (y_{(i)} - \mu_{[N-t]})^2, \qquad (2)$$

and

$$\mu_{[N-t]} = \frac{1}{N-t} \sum_{i=1}^{N-t} y_{(i)}. \qquad (3)$$

Assume that the desired coefficient of variation is $c$ and may be defined as

$$c = \frac{\sqrt{V(\hat{Y})}}{Y}.$$

Thus the variance of the total $V(\hat{Y})$ can be expressed in the terms of the coefficient of variation $c$ and if we substitute in equation (1) we can obtain an expression for the overall sample size

$$n(t) = N - \frac{(N-t)^2 S^2_{[N-t]}}{c^2 Y^2 + (N-t) S^2_{[N-t]}}. \qquad (4)$$

In other words, the overall sample size is obtained by adding to the number of take all units, the required sample size of the take-some sub-population for a specified coefficient of variation under simple random sampling. Additional manipulation leads to

$$n(t) = N - \frac{(N-t) c^2 Y^2}{c^2 Y^2 + (N-t) S^2_{[N-t]}}, \qquad (5)$$

which has fewer terms and is simpler to work with.

Hidiroglou(1986) demonstrated that for C, Y, and N fixed, there existed a unique minimum for $n(t)$. His demonstration rested on specifying a continuous analog for (5)

$$n(x) = N - (N-x)A / [A + (N-x)f(x)], \qquad (6)$$

where $A = c^2 Y^2, 0 \le x \le N-2,$ and $(N-x)^2 f'(x) \ge (N-x-b) f(x+b)$ for $b \ge 0$. and proving that the derivative of $n(x)$ with respect to $x$, $n'(x)$ is an increasing function with respect to $x$. In this approach the optimum sample size $n(t)$ was not determined in advance but depended on the (unknown ) number of take-all units, the desired coefficient of variation $c$, the variance of the take-some strata $S^2_{[N-t]}$ and the total $Y$.

Hence, the problem can be stated in terms of finding a cut-off value which would minimize the sample size $n(t)$ for a given level of precision $c$. Hidiroglou also demonstrated that a necessary condition for the option point is that (5), with $t = m$, should not exceed (5), with $t = m - 1$ or $t = m + 1$. This implies the option cut off value y* is found when

$$n(m-1) \geq n(m) \quad \text{and} \quad n(m) \leq n(m+1) \qquad (7)$$

This stopping rule is an exact one and is not expressed in terms of $y$. However, after some additional discussion about the necessary and sufficient conditions for an optimum, Hidiroglou obtained the following approximate cut-off (Approx. 1)

$$\text{Approx.(1) } y^* \leq \mu_{[N]} + \{c^2Y^2 / N + S^2_{[N]}\}^{1/2}, \quad (8)$$

which may prove moderately successful when the population of interest is moderately skewed.

Moreover, a simple iterative procedure can be utilized to obtain a cut-off point that approaches the exact solution. This is done by computing $\mu_{[N-t_1]}$ and $S^2_{[N-t_1]}$ where $t_1$ is the number of take-all units obtain from Approx.(1) and substituting these values into the following expression which gives the $(j+1)$th level of approximation

$$\text{Approx.(j+1)} = \mu_{[N-t_j]} + \{\frac{N-t_j-1}{(N-t_j)^2}c^2Y^2 + S^2_{[N-t_j]}\}^{1/2} \qquad (9)$$

for $j \geq 1$, $t_j, \mu_{[N-t_j]}$ and $S^2_{[N-t_j]}$ where have been determined by the previous approximation. Hidiroglou also proposed the following stopping rule: compute $n(t)$ and $n(t-1)$; choose Approx. (1) if, $0 < 1 - n(t_1)/n(t_2) < 0.10$. Otherwise, continue the process until, $0 < 1 - n(t_{j+1})/n(t_j) < 0.10$, if this condition holds then Approx. (j) is the best approximation to utilize.

After examining some practical illustrations based on the Retail Trade Survey conducted at Statistics Canada, Hidiroglou concludes that once populations are skewed the reduction in the sample size required for a given degree of precision is dramatic. Moreover, those gains increase as the skewness of the population increases. In the case of moderately skewed population Approx. (1) may be more adequate while for the more skewed population Approx. (2) and (3) are recommended.

**An Illustration from The National Sample Survey of Corporate Financing 1992**

In Table and 2 , we present the results of and application of the cut-off sampling procedure using data drawn from the sampling frame developed for 1992 National Sample Survey on Corporate Financing. The tables contain the following information:

1. The population size N
2. The cut-off rule: The stratifying variable utilized is the number of persons employed in the establishment. None, refers to simple random sampling selected from the whole population. Approx. (i), stands for the level of approximation;
3. the boundary value corresponding to the cut-off rule utilized, where simple random sampling is utilized, none refers to the maximum value for the population;
4. the size of the take-all strata 't' n(t) is the corresponding optimum sample size required to achieve a coefficient of variation of 5 of 10 per cent;
5. the skewness of the remaining elements in the population

The data reveal that population of interest is highly skewed and that even the use of Approx.(1) leads to a dramatic fall in the sample size required of a given level of precision. Table 1 shows that for a coefficient of variation of 5 per cent simple random sampling (SRS) would require a sample size of 791 establishments or 88 per cent of the establishments in the frame. By contrast even if we were to use cut-off sampling utilizing the non-optimum Approx.(1), the sample size required for the same level of precision is only 192 establishments or 24.4 per cent of the number of establishments required for the SRS methodology. In both instances Approximation 3 appears to have performed the most adequately leading to a sample sizes of 124 and 66 establishments respectively for the five (5) and ten (10) per cent levels of precision. In each case the cut-off sampling results in a situation where the take-all universe represents more than half the final sample size.

**Table 2**
**Cut-off Boundaries for Population Utilized in the N.S.S.C.F 1992**
**(coefficient of variation 5 per cent)**

| Cut-off Rule | Boundary Value | t | N-t | n(t) | Per cent Contribution | Skewness |
|---|---|---|---|---|---|---|
| None | 8498 | 0 | 895 | 791 | 0.0 | 15.37 |
| Approx.(1) | 494 | 21 | 874 | 192 | 10.9 | 2.85 |
| Approx.(2) | 209 | 58 | 837 | 126 | 46.0 | 1.94 |
| Approx.(3) | 191 | 71 | 824 | 124 | 57.3 | 1.83 |

**Table 3**
**Cut-off Boundaries for Population Utilized in the N.S.S.C.F 1992**
**(coefficient of variation 10 per cent)**

| Cut-off Rule | Boundary Value | t | N-t | n(t) | Per cent Contribution | Skewness |
|---|---|---|---|---|---|---|
| None | 8498 | 0 | 895 | 586 | 0.0 | 15.37 |
| Approx.(1) | 559 | 13 | 882 | 84 | 15.5 | 3.32 |
| Approx.(2) | 347 | 35 | 860 | 66 | 53.0 | 2.30 |

## IV Some Qualifications

While we have utilized an 'objective' cut-off rule in the above illustration, such rigid procedures are not really necessary for sorting the population into the strata. Kish (1965), notes that the entire sorting procedure is an area

> "... par excellence for the exercise of personal judgment, based on expert knowledge of the list and subject matter." [Kish (1965) p. 100]

He further cautioned that the stratifying variables should be only used where they are meaningful, and denote important sources of variation, when such conditions are absent so called objective procedures are unlikely to produce homogeneous strata.

## V Conclusion

In this paper we examined current official economic survey sampling practice and concluded that there is little justification for the continued popularity of purposive methods of selection. A randomized design was suggested as an alternative. In this technique the survey population is divided up into two strata: a take-all strata which comprises of all the large elements and a take-some strata comprising of small elements. The elements in the take-all start is sampled with certainty while a simple random sample is drawn from the take-some strata. We illustrate this method in the context of a proposed Corporate Financing Survey where use is made of cut-off rules due to Hidiroglou (1986) and data from the CSO's Establishment Register. This approach, unlike the purposive designs currently favoured by practitioners, has the advantage of protecting the researcher from accusations of personal bias in the choice of sample while still allowing for the exercise of some degree of judgment in sample selection. The survey practitioner can also make inferences about population statistics solely by reference to the sample parameters. Such surveys also provide rich grounds for secondary analysis using regression and other formal modeling techniques. In the absence of randomized designs the use of such techniques represents an abuse of the statistical method.

# REFERENCES


Balance of Payments Manual, Fourth Edition, International Monetary Fund, 1977.

Cochran, W.G. (1977), Sampling Techniques (3 ed.), New York: John Wiley.

Colledge, M., and Lussier, R. (1988), "A Generalized Methodology for Economic Surveys", Proceedings of the Business and Economic Statistics Section, American Statistical Association, 131 - 140.

Crichton, N., and Farrell, T.W.F. (1988), "Market Structure and Concentration in the Manufacturing Sector in Trinidad and Tobago", Social and Economic Studies, 37, 151 - 192.

Farrell, T.W.F., Najjar, A., Marcelle, H. (1986), "Corporate Financing and Use of Bank Credit in Trinidad and Tobago", Social and Economic Studies, 35, 1 - 167.

Forde, P. (1989), The Adequacy of the Present Statistical Data Base to Inform the Planning and Monitoring Mechanism: A User's Perspective. Paper presented at the Colloquium on Statistics, UNELAC.

Glasser, G.J. (1962), "On The Complete Coverage of Large Units in a Statistical Study", Review of the International Statistical Institutional Institute, 30, 28 -30.

Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953), Sample Survey Methods and Theory (Vol. 2), New York: John Wiley.

Hansen, M.H., Madow, W.G., and Tepping, B.J. (1978), "Foundations of Inference in Survey Sampling", Proceedings of the Survey Research Section, American Statistical Association, 82 - 107.

Hidiroglou, M.A. (1986) "The Construction of a Self-Representing Stratum of Large Units in Survey Design", The American Statistician, 40, 27 - 31.

Hidiroglou, M.A. and Srinath, K.P., (1981), "Some Estimators of Population Totals from Simple Random Samples Containing Large Units", Journal of the American Statistical Association, 76, 690 - 695.