Development of multistage designs for statistical surveys in the Philippines

Burton Torres Oñate

Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Statistics and Probability Commons

Recommended Citation

Oñate, Burton Torres, "Development of multistage designs for statistical surveys in the Philippines" (1960). Retrospective Theses and Dissertations. 2624.

https://lib.dr.iastate.edu/rtd/2624
DEVELOPMENT OF MULTISTAGE DESIGNS
FOR STATISTICAL SURVEYS IN THE PHILIPPINES

by

Burton Torres Oñate

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subject: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University
Of Science and Technology
Ames, Iowa

1960
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION.</td>
<td>1</td>
</tr>
<tr>
<td>II. REVIEW OF LITERATURE.</td>
<td>9</td>
</tr>
<tr>
<td>A. Related Specifications of the Sampling System</td>
<td>9</td>
</tr>
<tr>
<td>1. Sampling unit and hierarchy of sampling unit</td>
<td>9</td>
</tr>
<tr>
<td>2. Some procedures of selection of sample</td>
<td>13</td>
</tr>
<tr>
<td>3. Some methods of estimation</td>
<td>17</td>
</tr>
<tr>
<td>B. Problems in Sampling for Time Series</td>
<td>23</td>
</tr>
<tr>
<td>III. SAMPLE DESIGN</td>
<td>98</td>
</tr>
<tr>
<td>A. Background Data</td>
<td>98</td>
</tr>
<tr>
<td>1. Location and area</td>
<td>99</td>
</tr>
<tr>
<td>2. Population</td>
<td>30</td>
</tr>
<tr>
<td>3. Political subdivision</td>
<td>30</td>
</tr>
<tr>
<td>B. Sample Design</td>
<td>31</td>
</tr>
<tr>
<td>1. Regional stratification</td>
<td>31</td>
</tr>
<tr>
<td>2. Stratification within region</td>
<td>39</td>
</tr>
<tr>
<td>3. Methods of sample selection</td>
<td>33</td>
</tr>
<tr>
<td>4. Multi-stage sampling with complete replacement of primaries</td>
<td>39</td>
</tr>
<tr>
<td>5. Details of methods of estimation</td>
<td>48</td>
</tr>
<tr>
<td>C. Comparison of 'Survey' and 'Stratified'</td>
<td>63</td>
</tr>
<tr>
<td>Estimates of Variance</td>
<td></td>
</tr>
<tr>
<td>IV. ROTATION OF SAMPLE</td>
<td>65</td>
</tr>
<tr>
<td>A. Development of a Rotation Scheme</td>
<td>65</td>
</tr>
<tr>
<td>1. Purpose</td>
<td>65</td>
</tr>
<tr>
<td>2. Description of the new rotation scheme</td>
<td>68</td>
</tr>
<tr>
<td>B. Sampling Rate within Segments</td>
<td>70</td>
</tr>
<tr>
<td>1. Equal sampling rate</td>
<td>72</td>
</tr>
<tr>
<td>2. Unequal sampling rate</td>
<td>84</td>
</tr>
</tbody>
</table>
V. RATIO ESTIMATION

A. Control of the Variation in the Size of Barrio

B. Ratio Estimation in the Urban Area

1. Biased ratio estimation
2. Unbiased ratio estimation

C. Empirical Results

VI. COMPOSITE ESTIMATION

A. Estimate of Stratum Total, \( \alpha_{nth} \), from the \( i^{th} \) Primary

1. Structure
2. Variance

B. Estimate of Visit-to-Visit Change, \( \alpha_{nth} - \alpha_{(n-1)th} \), from the \( i^{th} \) Primary

1. Structure
2. Variance

C. Estimate of Visit-to-Visit a Year Ago Change, \( \alpha_{nth} - \alpha_{(n-2)th} \), from the \( i^{th} \) Primary

1. Structure
2. Variance

D. Sector Estimate

E. Choice of an Appropriate \( Q \)

F. Finite Population Approach

1. Structure of estimate
2. Variance of composite estimate

VII. SUMMARY AND RECOMMENDATIONS

VIII. LITERATURE CITED

IX. ACKNOWLEDGMENT
I. INTRODUCTION

From the time of her discovery by Spain in 1521, to the year of her independence in 1946, the Philippines had experienced three important historical periods. She was under the Spanish colonial regime from 1521 to 1900; the American administration from 1900 to 1941; and under the Japanese military occupation from 1941 to 1945. Shortly after World War II, she became an independent republic on July 4, 1946.

No major statistical activity was conducted prior to 1900. Censuses of population and selected characteristics of socio-economic life in the Philippines were conducted under American administration in 1903, 1918, 1939, but without the help and support of a general purpose statistical organization.

The present Bureau of the Census and Statistics was created in 1940, under the office of the President, by Commonwealth Act No. 591 which transferred powers, functions, duties, personnel, appropriations, property and records from various offices of the Commonwealth Government to the new Bureau, effective January 1, 1941. Little could be done by the new Bureau until after the Japanese occupation. A national census of population and agriculture was taken in 1948 under the auspices of the new Bureau. On February, 1951, the Bureau was transferred to the Department of Commerce and Industry.
The idea of centralization of government statistics which was reflected in the Bureau's set-up, has not been workable on a government-wide basis, and statistical activities soon reappeared in the other offices and departments of the government. The pattern thus became one of decentralization, with each office or department active in the statistical operations for its own administrative purposes, while the Bureau of the Census and Statistics serves as the specialized multipurpose statistical office.

During the past few years, the importance of statistics for government and private use, in the Philippines, has received growing recognition. This is reflected in a widespread growth of statistical activities in all echelons of the governmental system and also in the production of basic information on the population and socio-economic life of the country necessary for economic development programs. This recognition is basically reflected in the legal provisions which have set a framework for the administration of a coordinated statistical program and for the improvement of statistical standards.

The Government Survey and Reorganization Commission created by Republic Act No. 997, 1952, recommended, among others, that the present system of decentralized statistical operations should be continued and that a central office be created to coordinate this decentralized system. To implement these recommendations, the Office of Statistical Coordination
and Standards (OSCAS) was created under the revitalized National Economic Council (NEC) by Executive Order No. 119, July 1, 1955 and the OSCAS was formally organized January 1, 1956. This office is charged with the responsibility to: (a) promote an orderly and efficient program to meet the statistical requirements of the government, especially the NEC; (b) develop and prescribe adequate standards and methods for the use of other statistical agencies; and (c) maintain effective coordination of decentralized statistical activities. Within this new framework of statistical organization, several inter-departmental committees have been set up to study the best allocation of responsibilities for collecting data in fields where serious duplications existed and the improvements of standards and methods for various types of data.

The OSCAS has realized that with a fast growing and dynamic population even a complete census at various intervals, say five or ten years, is of limited use. Data from the census of 1948, for example, will not be very useful in the preparation of the NEC five-year economic and social development programs which require a thorough knowledge of the socio-economic characteristics of the population in 1956. Data obtained from a series of sample surveys conducted every year

---

*The author is Assistant Director of this new office.*
or every six months will be more useful.

One of the most important results of these considerations was the creation and organization of a national sample survey, The Philippine Statistical Survey of Households (PSSH), which was started in May, 1956 under the direct administrative and technical supervision of the OSCAS and with the cooperation of the Bureau of the Census and Statistics as the operating agency. The PSSH was organized for the purpose of (a) obtaining on a sample basis up-to-date statistical data on the size, composition and distribution of the population, labor force, economic activities of the people, income and expenditures, toilet and water facilities and other related information; and (b) maintaining these statistics on a current basis.

The PSSH utilized the sub-sampling technique, commonly known as multi-stage sampling. This technique is especially valuable in a country like the Philippines where no adequate frames exist. As mentioned earlier, the last census was conducted in 1948. Multi-stage sampling enables the complete utilization of existing political or administrative divisions such as provinces, towns or municipalities, and barrios in the rural sector and precincts in the urban sector of the country as sampling units at various stages. This technique permits also the permanent concentration of field work at various primary or secondary sampling units scattered throughout the country. The time of the supervisor assigned in each region
will be spent efficiently for training and supervision instead of travelling and locating the interviewers. Another advantage is the fact that the construction of the frame for succeeding stages need only be carried out for those previous stage units which are actually included in the sample. One of the most important requirements of the present design of the PSSH is the complete listing of households in certain sampling units.

The structure of the present design of the PSSH, and the operations involved in each stage will be discussed in Chapter III. It will also be shown how some of the design details have been obtained as 'optimum solutions' subject to the limitations imposed by local conditions and resources. Formulas for the estimation of population totals and sector totals are given and formulas for estimating their variances are discussed.

The present PSSH is a 'panel survey' as far as certain sampling units (areas) are concerned - and once put into operation commits the estimation procedures to the permanent design. The choice of technical devices is limited to a redefinition of the sampling units, to different methods of sampling within the selected primary or secondary sampling units, to the choice of the estimation procedure, or to the use of techniques which will result in the reduction of the cost of the survey. Such improvements in the design of the PSSH through the use of these various techniques will be presented in Chapters IV, V and VI.
The design and estimation procedure of the PSSH require the complete listing of the ultimate clusters or sampling areas. This first phase of the field operation is done about one week prior to actual interviewing of the sample households and consists of a complete enumeration of all households, population in households and occupation of head in each of the designated sampling areas. About 0.163 million households and 0.969 million population are listed during this first phase of each survey or visit. These figures give an idea of the extent and possible cost of the listing operation alone. More than 15 per cent of the field cost is devoted to this operation. Since the PSSH was devised to sample the same population repeatedly, another important point which has to be considered is the frequency of visit to the panel households and the pattern in which the sample will have to be changed from visit to visit. It has been observed that heads of households or the respondents have become uncooperative during the third or fourth visit. This brings up the question of 'response resistance'. Even with complete cooperation, the respondents may also be influenced by the information which they give and receive during the interviews, and this constant exposure may make them progressively less representative in succeeding visits. A rotation scheme will be developed in Chapter IV which will provide some answers to the problem of cost of listing and to the question of 'response resistance' of panel...
An attempt was made to equalize the size of the barrio sector in the barrio stage. This development will be described and some empirical results will be given to indicate the effect of equalization. In addition, ratio estimation with rotation will be introduced into the estimation procedure. These techniques are described in Chapter V.

The rotation scheme for the 'matched sample' design developed in Chapter IV consists of changing a portion of the sample of households from visit to visit. This situation affords an excellent opportunity to use a new estimation procedure, the so-called composite estimate (Chapter VI), a similar version of which is used by the U. S. Bureau of the Census in their current population survey. For the preparation of socio-economic development plans, the interest is on the current total, particularly for data or characteristics of the population which are likely to change considerably with time. A composite estimate will be developed which will estimate the current total or level.

The NEC development programs are revised from year to year consistent with the performance attained and with whatever modifications changing conditions may warrant. In order to evaluate the effects of these programs on the various sectors of the economy, estimate of change is of prime importance. For this situation, the composite estimate for each succeeding
visit will be utilized to give an estimate of the visit-to-visit change and an estimate of the visit-to-visit a year ago change. The gain in reliability of these estimates will also be discussed.
II. REVIEW OF LITERATURE

The review of literature will consist of two main headings, namely: (a) related specifications of the sampling system; and (b) problems arising when sampling time series with panel households. The first will have special bearing on the developments given in Chapters III, IV and V; the second on Chapters IV and VI.

A. Related Specifications of the Sampling System

The requirements of an efficient sampling system are many and varied. These include, among others, the definition of the sampling unit, the method of selecting the elements or sampling unit which will constitute the sample and the method of estimation to be used. Techniques have been devised for the utilization of auxiliary information, quantitative or qualitative, to improve these requirements of the sampling system.

1. Sampling unit and hierarchy of sampling units

The definition of the sampling unit (su) in any sampling system is of great importance. It is a concept subject to considerable choice. In practice, it is often possible to subdivide the population into sampling units (sus) in many ways. An urban area may be regarded as composed of a number of pre-
cincts (voting areas), or of a number of blocks, or of a number of households, or of a number of persons. In multi-stage designs, the precincts may be used as the first stage units, the blocks as second stage units, the households as third stage units and the persons as the fourth stage units. In this way, one can create a hierarchy of sampling units.

Jessen [15] investigated the efficiencies of four sizes of units in sampling for farm facts. He compared a quarter section, a half section, a section (area of 1 square mile), and a block of two contiguous sections. In most of the items studied the quarter section was found to be optimum. In India, Sukhatme and Panse [29] considered the village as the optimum unit of sampling in agricultural surveys. The village is also a very convenient administrative subdivision.

With auxiliary information, one can define units which are as alike as possible with regard to one or more characteristics under study. This technique will result in an increase of efficiency of a sampling system with equal probability due primarily in a reduction of the variance of the individual units. Another method is to use clusters, aggregates, or groups of natural units as sus. The use of these sus is of some appeal to countries where no accurate frame exists. Under this method a sample of households may be obtained by associating every household with one and only one areal segment and then a sample of these areal segments is selected by some
random scheme. This is commonly known as area sampling. Jessen [14] introduced the use of auxiliary information in the form of maps and aerial photographs in defining areal sus which he found to be less variable in respect to the characteristics under measurement than units of equal area.

Several attempts have been made to work out a relation between the variance from a sample of sus of any size and the variance from a sample of units of a specified size. The effect on the variance of the estimate from using clusters made up of varying numbers of elements as su's may be expressed in terms of a measure $\delta$ of homogeneity within su's [9, Vol. 2, p. 164]. Sukhatme [28, p. 247] discussed the efficiency of cluster sampling in terms of the intra-class correlation.

For many items, $\delta$ tends to decrease with an increase of the number of elements, but the rate of decrease is small relative to the rate of increase in size. This relationship may be expressed as

$$\delta = \alpha \bar{K}^\beta$$

where $\bar{K}$ is the average number of elements per cluster, $\alpha$ is a positive constant and $\beta$ is a negative constant. If clusters as sus are used rather than individual elements, then there is usually an increase in the variance over that of simple random sampling of the same number of elements by a factor

$$\lambda = 1 + \delta(\bar{K} - 1).$$

In general, we expect the variance to increase with the size of
the cluster. In some cases, however, clusters may be superior to elements as sampling units. Sukhatme [28, p. 248] found that households, as a cluster, will be approximately twice as efficient as a single person for estimating the sex ratio.

In sub-sampling or multi-stage sampling plans where there exists a hierarchy of sampling units, the different methods of utilizing prior information are applicable at each stage of sampling.

In order to increase the efficiency in a two-stage sampling, Hansen and Hurwitz [8] recommended that strata be as homogenous as possible and that strata be composed of large primary internally heterogenous units. By increasing the heterogeneity within primary sampling units (psus) the variability between psus was reduced, and on items for which this component was large, considerable reduction in the total variance of the estimates was attained. Moreover, Dalenius [3, p. 14] pointed out that for a constant number of secondary sus in the sample, the total variance of the estimate will increase on decreasing the number of psus in the sample. In general multi-stage sampling, the type of psus usually has a much greater impact on reduction of total variance of the estimate than has the type of secondary sus. As a rule, the psus should be made as internally heterogenous as possible, when the primary sampling fraction is small.

There are ways other than changing the size of the psu in
which heterogeneity within ssus can be introduced and hence variability between sus can be reduced. If certain administrative or political units are being combined to form a psu, then pairing of political units in which one is predominantly agricultural and one which is largely urban, will introduce heterogeneity. More often, these units will depend upon the administrative or political subdivisions into which a given country is segmented. In the United States, the county has been used as a unit of sampling [13, 20]. The commune, smallest administrative unit, is used in the construction of the psu in the Swedish national sample surveys [3, p. 105]. Gray and Corlett [6] reported the use of districts as psus in most national sample surveys in Britain. In the Philippines, the municipality, the second largest administrative unit in the rural area, is used as psu while in the urban area the precinct or voting area is used as psu. The reason for this choice is predominantly of a practical administrative nature. It is possible that the variation of certain items within a municipality is smaller than the variation between municipalities in the same stratum so that this unit as a psu may not be ideal from a theoretical standpoint.

2. Some procedures of selection of sample

The availability of prior information on the elements of the population will determine to a great extent the procedure to follow in the selection of the units which will constitute
the sample. Procedures applicable to one-stage sampling may be utilized at any stage of a hierarchy of sampling units.

In simple random sampling $n$ units are drawn out of $N$ units such that every one of the $C_N^N$ samples has an equal chance of being chosen. As is well known, the variance of a sample mean is inversely proportional to the size of the sample. The precision of the sample mean is increased by increasing the size of sample $n$.

If prior information is available about the $N$ units, then it may be possible to divide a heterogeneous population of $N$ units into sub-populations called strata of $N_1, N_2, \ldots, N_h$ units, each of which is internally homogenous. Note that strata should be homogenous while psus should be heterogeneous. The sample is then allocated to each stratum ($n_h \leq N_h$) and precise estimates are obtained from a smaller sample in each stratum. These estimates can then be combined into a precise estimate for the whole $N$ population.

When the stratum sizes ($N_h$'s) are known, the simplest allocation is to make $n_h$ proportional to $N_h$. For optimum allocation of the sample in the respective strata with the sample size

$$n = \sum_h n_h$$

make

$$n_h \propto \frac{N_h \sigma_h}{\sum_h N_h \sigma_h} n$$
where $c_n$ has different forms for different estimators. This relation is often called 'Neyman' allocation [22], although this result was previously obtained by Tschuprow [30]. There are, however, some practical difficulties associated in the use of Neyman's method of allocation [2, p. 84].

When strata sizes, $N_n$'s, are not known, as often happens when stratum sizes are based on old census data which are out of date for use in current surveys, the procedure given above may still be utilized. A 'large' sample is first selected and stratified into $L$ strata and estimates of $N_n$ are made; then each stratum is considered as a sub-population. This procedure is a special case of double sampling for stratification or two-phase sampling. Double sampling may be very convenient when the information about $x_i$ is on file cards that have not been completely tabulated. Cochran [2, p. 268] states that this technique is profitable only if the gain in precision from ratio or regression estimates or from stratification more than offsets the loss in precision due to the reduction in the size of the large sample.

Mahalanobis [21] states that an optimum or nearly optimum solution to minimum variance stratification would be obtained when the expected contribution of each stratum to the total aggregate value of $x_i$ is made equal for all strata. Hansen et al. [9, Vol. 1, p. 416] give a rule of thumb which will provide a rough guide to the optimum sizes of strata when a
constant number of psus is selected from each stratum. The rule is to make the strata equal in terms of \( \sum_{i} x_{ih} \) total of the \( h \)th stratum, when the psu rel-variances are about the same and remain about the same on adjusting the sizes of strata.

Other methods of restriction, in addition to stratification, may be introduced in the selection procedure. The procedure of first selecting psu and then selecting within each psu a specified number of elements is known as sub-sampling or multi-stage. As mentioned before, this procedure will group the stage units into a hierarchical classification. The principal advantage of multi-stage sampling is its greater flexibility as compared to one-stage sampling. A balance between administrative considerations, costs and precision dictate that most sample surveys be of this type.

The selection of psu in a multi-stage design with complete replacement of psu and with probability proportional to size (pps) will be discussed in detail in Chapter III.

Systematic sampling consists of selecting a sample of size \( n \), by taking a unit at random from the first \( k \) units and then every \( k \)th subsequent unit. Systematic samples are convenient to draw and to execute \([1]\). When a list is used on which the units are recorded in an essentially random way, 'systematic' samples are equivalent to random samples. When such an assumption cannot be made at least the random start should be made. In such situations 'systematic' samples with
a random start compare often favorably with stratified sampling. The procedure is convenient especially in field operations. In this case, there is no gain in precision. This type of sampling was utilized in the PSSH sample areas.

Available prior information can be used in any of the operations which make up the sampling plan. This is done in order to increase the precision of the overall results. This information may be qualitative or quantitative. Some decision will have to be made as to which of these information items will be utilized for various operations of the survey such as stratification, construction of psu's, choice of the selection procedure, choice of the estimation procedure and others.

3. Some methods of estimation

Another general use of prior information, usually quantitative, arises in the estimation procedure. Its use will in general give rise to an estimate of higher precision than can be attained with estimators not utilizing this information.

In the succeeding discussions, the nomenclature usually given by Hartley [11] for \( \bar{x} \), \( \tilde{x} \) and \( \hat{x} \) will be used.

The simplest estimation procedure may be called 'x only' estimation. As estimate of the population mean, \( \bar{x} \), we define

\[
\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}
\]

as the 'x only' estimator of \( \bar{x} \). Similarly, we may define \( \bar{y} \) as the 'y only' estimator of \( \bar{y} \). If the quantitative informa-
tion, $\bar{Y}$, is known, then various ratio estimators may be used.

The 'ratio of means' estimate \([2, p. 112; 11]\)
\[ \tilde{X} = \frac{\bar{X}}{\bar{Y}} \]
is used to estimate $\bar{X}$, where $\bar{x}$ and $\bar{y}$ are the 'x only' and 'y only' estimators of $\bar{X}$ and $\bar{Y}$, respectively.

The 'mean of ratios' \([11]\) is another estimate of $\bar{X}$ and is given as
\[ \hat{X} = \bar{r} \bar{y} \]
where
\[ \bar{r} = \frac{1}{n} \sum_{i=1}^{n} r_i \]
and
\[ r_i = \frac{x_i}{y_i} . \]

The familiar approximate formula for the variance of $\tilde{X}$ and $\hat{X}$ \([2, p. 116; 11]\) is
\[ \frac{1}{n} \left\{ \sigma_x^2 + \sigma_y^2 - 2Q \text{Cov}(x, y) \right\} \]
where
\[ Q = \frac{\bar{X}}{\bar{Y}} . \]

$\tilde{X}$ and $\hat{X}$ are biased \([2, p. 130; 11]\). In most surveys, the bias may be a small portion of the standard deviation and may be ignored or neglected. In stratified sampling where the $n_h$ in each stratum is small and $L$, the number of strata, is large,
the bias in the stratified total estimate may not be negligible relative to its standard error. If the bias of the estimate has the same sign in each stratum, then the bias in the estimate of population total will be approximately \( L \) times that for an individual stratum total, but the standard error is only \( \sqrt{L} \) times that for individual stratum total. Hartley and Ross [12] derived exact expressions for the bias in \( \tilde{x} \) and \( \hat{x} \) and from these results an unbiased ratio estimator, \( x' \), was worked out. This is given as

\[
x' = \tilde{x} - \frac{N - 1}{N} \frac{n}{n - 1} (\tilde{x} - \overline{xy}) .
\]

Goodman and Hartley [5] compared the precision of \( x' \) with that of \( \tilde{x} \) and \( \hat{x} \). Results given in [2, 5, 11, 12] will have direct bearing in the introduction of ratio estimators into the design of the PSSH.

In the estimation of population total in multi-stage designs, one of the most stringent requirements for an efficient estimation procedure is an accurate knowledge of quantitative characteristics of the various stage frames. The unbiased estimate of the population total, \( X \), in a three-stage design without replacement and equal probability at each stage is

\[
\hat{X} = \frac{M}{m} \sum_{i} \frac{B_{i}}{b_{i}} \sum_{j} \frac{N_{i}^{1}}{n_{i}^{1}} \sum_{k} x_{ijk}
\]

where \( M \) is the total number of psus in the population,

\( m \) is the number of sample psus,
$B_i$ is the total number of secondary ssus (ssus) in the $i^{th}$ sample psu,

$B_i^s$ is the number of sample ssus in the $i^{th}$ sample psu,

$N_{ij}$ is the total number of listed households (hhs) in the $(ij)^{th}$ sample ssu,

$n_{ij}$ is the number of sample hhs in the $(ij)^{th}$ sample ssu,

and $x_{ijk}$ is the characteristic in the $k^{th}$ hh in the $j^{th}$ ssu of the $i^{th}$ psu.

If the sampling scheme is completely specified for each stage of sampling, then for $\hat{X}$ to be efficient, accurate knowledge of $M$, $B_i$'s, and the $N_{ij}$'s must be available. This points to the importance of the accuracy of the different frames. Similar requirements are prescribed in the estimation of the variance of $\hat{X}$.

The variance of $\hat{X}$ consists of three parts, namely: the between psu component, the between ssu within psu components, and the between terriaries within ssu's within psu's component. The between psu component depends on psu totals; the second component on ssu totals; and the last component on the tertiary observations and their means. Thus, to reduce the between primary component, which is usually the predominant one, we may either change the design by drawing the psu with probabilities proportional to size (pps) or without change of design, we use a different estimator, say, a ratio estimator making use of size of psu as the auxiliary variable. These
techniques are discussed in detail by Hartley [11].

The theory of estimation in multi-stage design with complete replacement of psu is given in Chapter III. Of interest are some fundamental formulas for estimation.

In a self-weighting two-stage sampling scheme with equal probability and without replacement at each stage, an unbiased estimate of the stratum total, \( X_h \) is given by

\[
\hat{X}_h = \frac{1}{f_h} \sum_{i} \sum_{j} x_{hij}
\]

where \( x_{hij} \) is the characteristic on the \( j \)th secondary in the \( i \)th primary of the \( h \)th stratum,

\( n_{hi} \) is the number of sample secondary in the \( i \)th sample primary of the \( h \)th stratum,

\( m_h \) is the number of sample primary in the \( h \)th stratum, and

\( f_h \) is the stratum constant obtained by the product of \( f_1 \), the primary sampling fraction and \( f_2 \), the secondary sampling fraction.

The relative variance of \( \hat{X}_h \) is

\[
V^2(\hat{X}_h) = \frac{\sigma^2(\hat{X}_h)}{x_h^2}
\]

which may be written as the sum of three components [9, Vol. 2, p. 204].
\[ V^2(\hat{X}_h) = V^2(\hat{X}_h) + V^2(n_h) + 2\int_{X_h, n_h} V(\hat{X}_h) V(n_h) \]

where \( \hat{X} = \frac{X}{n_h} \),

\[ n_h = \sum_{i=1}^{m_h} n_{hi} \],

\( \rho_{\hat{X}_h, n_h} \) is the correlation between \( \hat{X}_h \) and \( n_h \),

and \( V \) is the relative variance of corresponding variable in the parenthesis.

The first component, \( V^2(\hat{X}_h) \), may be written as the sum of two other sub-components: the variation between primaries and the variation within primaries. \( V^2(n_h) \) is the contribution from the variation in the size, \( N_{hi} \), of the different primaries in the \( h^{th} \) stratum. The last component contains the correlation coefficient, \( \rho_{\hat{X}_h, n_h} \), which in most practical applications is often found to be near to zero. This leaves the first two components

\[ V^2(\hat{X}_h) + V^2(n_h) \]

as an approximation to \( V^2(\hat{X}_h) \). Experience of these authors shows that \( V^2(n_h) \) is often relatively great and this feature is also apparent in the between primary component of the exact variance of \( \hat{X}_h \) which depends on the variation of the primary totals. These results will have some bearing on the development of the barrio size equalization which is presented in Chapter V.
B. Problems in Sampling for Time Series

When the same population is sampled repeatedly, opportunities for a flexible sampling design are increased. On visit $\lambda$, we may have parts of the sample that are matched with visit $(\lambda - 1)$, parts that are matched with both visits $(\lambda - 1)$ and $(\lambda - 2)$ and so on. To improve the current estimate, a multiple regression involving all matchings to previous visits may be utilized. To utilize this information, sampling must be done in such a way that the two samples drawn at successive times $(\lambda - 1)$ and $\lambda$ have some elements in common.

This method of sampling where old elements of the sample are eliminated and new members are added is called sampling on successive occasions with partial replacements of units \([24, 31]\). Hansen et al. \([9, \text{Vol. 2, p. 272}]\) refer to this method as sampling for a time series and Eckler \([4]\) named the method rotation sampling. Jessen \([15]\) has studied the method of replacement for current estimates from samples of the same size on both visits. Double sampling may be regarded as rotation sampling involving a current sample and one overlapping earlier sample \([2, \text{p. 268}]\).

In one-level rotation sampling, one can add to the same pattern only sample values that have been drawn from the population at the current time. The sample pattern has $n$ sample values at each time $\lambda$, $\lambda = 1, 2, \ldots, A$. $(1 - \lambda)n$ of the
elements in the sample at time $(\alpha - 1)$ are retained in the sample at time $\alpha$, and the remaining $\lambda$ elements are replaced with the same number of new ones. At time $\alpha$ the $\lambda$th row of sample values is added to the $(\alpha - 1)$st rows of earlier values. Since each enlargement of the pattern consists of a set of sample values associated with a single time, we call this one-level rotation sampling on the above pattern. This general problem of replacement has been studied by Patterson [24] and he derived a necessary and sufficient condition for a linear unbiased estimate to be minimum variance.

In two level and higher level rotation sampling, one can add earlier sample values as well as current ones to the pattern. Eckler [4] extended the method of Patterson [24] for finding minimum variance unbiased estimates to two level and three level rotation schemes. Rotation sampling is of value when the correlation between the two consecutive measures of a characteristic for a unit is high. If it costs no more to carry out rotation sampling than independent random sampling, then even a modest reduction of five to ten percent in variance will be worthwhile [4].

Kailin [17] gave a description of the sample for the U.S. monthly retail trade report which includes a discussion of the composite estimate used and its variance. Hansen et al. [9, Vol. 2, p. 272] presented a composite estimate and its variance in sampling for a time series. This composite
estimate is very similar to the one developed by Hansen et al. \cite{Hansen1954} in the redesign of the U.S. census survey of February 1954.

In February 1954, the U.S. Bureau of the Census introduced a redesign of the Current Population Survey (CPS), from which, information on employment, unemployment, and other related data are compiled each month. One feature of the subsampling which has an important bearing on the estimation procedure introduced in the new sample involves changing a part of the sample each month. To accomplish this rotation, eight systematic sub-samples or rotation groups of segments are identified for each sample. A given rotation group is interviewed for a total of eight months, divided into two equal periods. A rotation group (see Figure 1) is in the sample for four consecutive months one year, leaves the sample during the following eight months, then returns for the same four calendar months of the next year. It is then dropped from the sample. In any one month one-eighth of the sample segments are in their first month of enumeration, another eighth are in their second month, etc., with the last eighth in for the eighth time (see Month 4 in 2nd year of Figure 1). Under this scheme, 75 per cent of the sample segments are common from month to month (see Month 4 to Month 5 in 2nd year) and 50 per cent are common from year to year (see Month 4 in 1st and 2nd years).
<table>
<thead>
<tr>
<th>Year number</th>
<th>Month number</th>
<th>Rotation group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>** In first time</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>** In second time</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>**</td>
</tr>
</tbody>
</table>

Figure 1. Rotation of sample in the U.S. Bureau of the Census Current Population Survey
The composite estimate developed is

\[ X_u'' = K(X_u'' - 1 + X_u,u-1 - X_{u-1,u}) + (1 - K)X_u' \]  

(2.1)

where \( 0 \leq K \leq 1 \)

\( X_u'' \) is the composite estimate for month \( u \),

\( X_u' \) is the regular ratio estimate based on the entire sample for month \( u \),

\( X_{u,u-1}' \) is the regular ratio estimate for month \( u \) but made from the returns from the segments that are included in the sample for both months \( u \) and \( u-1 \), and

\( X_{u-1,u}' \) is the regular ratio estimate for the previous month \( (u-1) \) but made from the returns from the segments that are included in the sample in both months \( u \) and \( u-1 \).

The variance of this estimate under some simplifying assumptions is,

\[ \sigma^2_{X_u''} = \frac{\sigma^2_{Y_u'}}{1 - K^2} + \frac{2}{1 - K^2} \sum_{j} K^j \sigma_{Y_{u,j},Y_{u,j}} \]  

(2.2)

where \( Y_u' = K(X_u,u-1 - X_{u-1,u}) + (1 - K)X_u' \),

\[ \sigma^2_{Y_{u-1}'} = \sigma^2_{Y_u'} \]  

for all \( i \),

\[ \sigma_{Y_{u-1},Y_{u-1-j}'} = \sigma_{Y_u',Y_{u-j}'} \]  

for all \( i \) and \( j \),

and \( u \) is assumed to be large.

This composite estimate takes advantage of accumulated
information from earlier samples as well as the information from the current one, and results in smaller variance of estimates of both level and change for most items, but the larger gains are achieved, for the most part, in the estimates of change. The form of the composite estimate, $X_u^*$, will be used in the developments given in Chapter VI.
III. SAMPLE DESIGN

The sample design of the Philippine Statistical Survey of Households (PSSH) reflects, among others, the history of the Philippines, its political, social and economic institutions, resources available and various conditions which were imposed on the design. At each succeeding stages in the development of the design, the relevant theory will be pointed out and discussed.

A. Background Data

1. Location and area

The Philippines is located at 21° 21' N and 4° 30' N Latitude, which corresponds to a distance of about 1,900 kilometers (1 km = 0.621 miles) and 116° 55' E and 126° E Longitude, or a distance of about 1,100 km. The total land area is 229,681 sq. km., of which 227,410 is land and 2,271 sq. km. water.

There are about 7,100 islands and islets, 2,700 of which are named. It has a tremendous coastline of more than 17,000 km. This situation of extreme segmentation will have some bearing on the manner of regional subdivision into which the country is to be divided for purposes of sampling.
2. Population

The population in households as of May, 1956, was estimated at 21,527 million and the density of population per sq. km. as 72. About 80 per cent of this population live in rural areas and 20 per cent in urban areas. Metropolitan Manila, the capital city, has almost 2 million population.

The family household or household (hh) consists of the members of a family forming the nucleus of the household, and also includes resident domestic servants, and other persons who may be living with the family. One or more households (hhs) may occupy the same dwelling unit. The population estimate refers to the non-institutional population (only persons found in hhs), and excludes the population found in diplomatic and consular residences, ships, asylums, hospitals, penitentiaries, army barracks, hotels and other similar institutions.

3. Political subdivision

The largest political or administrative subdivision is the province. Each province is a compact geographic area consisting of an urban sector and a rural sector. The urban sector is usually a provincial capital and/or chartered city or cities, and the rural sector is made up of a number of municipalities.

Each municipality is divided into two parts: the
poblacion, the seat of the municipal government and run by an elective mayor, is usually referred to as the rural-urban sector; and the purely rural sector which consists of a number of barrios. The barrio is the smallest administrative unit of local government. The residents of the barrio elect their own barrio lieutenant.

In June 1955, there were 53 provinces, 1,300 municipalities, 16,640 barrios and 28 chartered cities. Metropolitan Manila is considered a separate urban sector.

B. Sample Design

1. Regional stratification

The country was first divided into ten regional divisions. The subdivision was realized with the help of quantitative information such as various factors of agriculture, climate, dialects spoken, ethnic origin, economic activities, natural boundaries and others. Metropolitan Manila was considered as a separate region. The ten regions are:

Region I - Metropolitan Manila
Region II - Ilocos-Mountain Province
Region III - Cagayan Valley-Batanes
Region IV - Central Luzon
Region V - Southern Luzon and Islands
Region VI - Bicol
Region VII - Western Visayas
Region VIII - Eastern Visayas
Region IX - Northeastern Mindanao
Region X - Western and Southern Mindanao and Sulu

Each region, except Region I, consists of provincial capitals, chartered cities and municipalities. The municipality is made up of one poblacion and a number of barrios. The province was ignored as a unit in the sampling procedure followed. Thus estimates of a given characteristic of the population are made separately for each of the four sectors, namely: (1) the barrio, (2) the poblacion, (3) provincial capitals and/or chartered cities, and (4) Metropolitan Manila. The sum of (1) and (2) will give a rural area estimate and the sum of (3) and (4) will give an urban area estimate. The rural and urban estimates will produce a country estimate.

2. Stratification within region

Stratification technique for the rural area will first be described. In each region, the municipalities were arranged in ascending order of densities of population prior to division into paper strata. Each stratum need not be a compact geographic area. The strata were formed by counting off its specified population starting from the municipality with the lowest density. The population in each stratum was made approximately equal (1/2 million) and in no case was a given municipality dissected. A region had at least one
stratum.

The PSSH is a multi-purpose survey and most of the items to be estimated are highly correlated with population count. The most convenient stratification may be obtained for most items by making strata of approximately equal size \[9, \text{ Vol. 1, p. 416; 2l}\]. A total of 30 strata was formed using this technique.

The method of stratification for the urban areas was as follows. In each region, the provincial capitals (capitals) and chartered cities (cities) were stratified according to the degree of urbanization, i.e., percentage of urban precincts in the descending order. Precincts are voting areas. Paper strata were formed by counting off their approximate specified number of precincts, starting from the highly urbanized capital or city. Similar stratification techniques were followed for Metropolitan Manila (Manila). Strata equalization as to number of precincts was imposed. In no case was a capital or city dissected. The number of precincts per stratum was about 130 in the capitals and cities and about 80 in Manila.

With this scheme, 30 strata were formed in the capitals and cities and 32 in Manila.

3. Methods of sample selection

The PSSH was originally designed to reach a target sample of 6,500 hhs scattered over 300 barrios, 150 poblaciones and
58 capitals and cities, including Manila. The method used in the selection of the sample is known as multi-stage sampling with complete replacement of primaries [11].

A three-stage design was used in the rural areas. From each stratum, one municipality or primary sampling unit (psu) was selected with probability proportional to size (pps) of population. From each sample psu, two barrios or secondary sampling units (ssus) were selected with equal probability and without replacement. The poblanos were selected with certainty since there was only one poblanon in each psu. All hhs or tertiary sampling units (tsus) in selected sample barrios or poblanos were completely listed to obtain the number of listed hhs. The number of sample hhs to be interviewed was determined by the product of the sampling rate (sr) within barrios (or poblanos), the psu constant, with the total number of listed hhs. The sample hhs were selected systematically using a random start between 1 and \( \frac{1}{sr} \). This number is drawn in the central office and given to the supervisor. The psu constant (sr) is known through a restriction on the estimation procedure. This will be discussed in the next section.

After listing, this random number is given to the enumerator for use as a starting point for systematic selection of sample hhs within barrio or poblanon. Sample hhs located by this scheme were interviewed. Listing or the first phase was
distinctly a separate operation to that of actual interview or the second phase. In the listing operation, the population in listed hhs and the occupation of the head of hh were also obtained.

The psu, ssus and tsus were replaced and the above selection of the sample was repeated five times. The drawing of the psus was numbered within each stratum. Thus we have in each stratum five numbered psus, ten sample barrios, and five sample poblacions. The number of target sample hhs in each sampling area (barrio or poblacion) was ten.

With 30 strata in the rural areas, this scheme gave a total of 150 psus, 300 barrios, 150 poblacions, 3,000 target sample hhs in the barrio sector and 1,500 hhs in the poblacion sector.

A simple method of selecting psus with probability proportional to size (pps) which eliminates the computation of cumulative totals was utilized in drawing the psus in the rural areas.

It is required to select from the hth stratum a psu from a list of $M_h$ units with pps of the psu. Let the measure of size of the ith psu be $p_{hi}$. The method is as follows:

Choose a pair of random numbers, $r$ and $s$, the first from the range 1 to $M_h$ and the second in the range 1 to $p_{hi}$, where $p_{hi}$ is the largest of the $p_{hi}$s. If $p_{hi} < s$, reject the unit and if $p_{hi} \geq s$, accept the unit. If rejected repeat the
operation until a sample selection is made.

Since the sampling is done with complete replacement of the psus, this operation is repeated until \( m \) psus are selected.

This method was devised by Lahiri [18, p. 202] and extended by Grundy [7] to \( m \) psus drawn with pps but without replacement. Jessen et al. [16] used a systematic scheme to draw \( m_n (m_n > 1) \) psus from a stratum of \( N_n \) units.

A two-stage design was followed in the urban areas. From each stratum five precincts or psus were drawn with equal probability and with complete replacement. The five psus were numbered as in the rural areas. All hhs in selected sample psus were listed to produce the number of listed hhs. The number of sample hhs was determined by the product of the sampling rate within psu, the so-called stratum constant, and the number of listed hhs in the psu. The stratum constant was obtained through a restriction on the estimation procedure. Selection of sample hhs followed the scheme used in the rural areas.

A total of 150 precincts and 1,200 target sample hhs were drawn from the capital and cities sector while 160 precincts and 800 hhs were selected for Manila.

A summary of the actual sample of the PSSH for May, 1956, is shown in Table 1.

In national sample surveys, the field operation must be made as simple as possible but made within the theoretical
<table>
<thead>
<tr>
<th>Region</th>
<th>Number of sample municipalities</th>
<th>Number of sample poblaciones</th>
<th>Number of sample barrios</th>
<th>Number of sample households</th>
<th>Number of sample strata precincts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>30</td>
<td>150</td>
<td>150</td>
<td>299</td>
<td>1,086</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>2</td>
<td>10</td>
<td>20</td>
<td>89</td>
<td>256</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>1</td>
<td>5</td>
<td>10</td>
<td>53</td>
<td>135</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>5</td>
<td>25</td>
<td>50</td>
<td>218</td>
<td>513</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>4</td>
<td>20</td>
<td>39</td>
<td>213</td>
<td>409</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VI</td>
<td>3</td>
<td>15</td>
<td>30</td>
<td>96</td>
<td>304</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VII</td>
<td>5</td>
<td>25d</td>
<td>50</td>
<td>147</td>
<td>477</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VIII</td>
<td>6</td>
<td>30</td>
<td>60</td>
<td>174</td>
<td>569</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IX</td>
<td>2</td>
<td>10</td>
<td>20</td>
<td>43</td>
<td>168</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>2</td>
<td>10</td>
<td>20</td>
<td>53</td>
<td>161</td>
</tr>
</tbody>
</table>

* One precinct selected twice.

* One municipality selected twice.

* One of the sample municipalities had only one barrio.

* Three municipalities each selected twice.
requirements of the design. For example, the use of systematic sampling of hhs within sampling areas (barrio, poblacion or precinct) was mainly for simplicity and convenience with little expectation of a gain in precision [2, p. 185].

One important advantage of equal take in number of psus in the design of the PSSH is in the simplicity of computation of estimates and their estimated variances for the stratum, the region, the sector and the whole country. It has also the advantage of simplifying the organization of the field work, in equalizing the work-loads per supervisor or per interviewer and other aspects of the design. In addition, one may compare the efficiency of equal number of psus per stratum against proportional allocation, when sampling is done with replacement. Stevens [27] has shown that if

\[ \bar{M}_{th} = \frac{M_t}{L_t} \]

is the average size of the stratum in the \( t \)th sector and

\[ \sigma^2_{M_{th}} \]

is the second moment about zero of the frequency distribution of the stratum sizes, \( M_{th} \), then the efficiency of equal number as compared to proportional allocation is

\[ E = \frac{\bar{M}_{th}^2}{\sigma^2_{M_{th}}} \times 100. \]

When \( M_{th} = M_{th'} \) (\( h \neq h' \)) for all strata, then \( E \) is equal to 100 per cent. As the ratio of the largest to the smallest stratum
sizes increases, the efficiency of equal number falls off. When the ratio is about 2, the value of $E$ is about 95 per cent. It is assumed that $\sigma^2_x$, the variance of a characteristic $x$ within the stratum, is constant.

In the PSSH, the range of the size of the stratum, $M_{th}$, for the capitals and cities is 101 to 187. For Manila, the range is 66 to 100. The efficiency of equal take in each of these two cases is still close to 100 per cent. In the rural area, barrio and poblacion sectors, the sizes of the stratum were made equal to the size of population, $P_{th}$.

4. Multi-stage sampling with complete replacement of primaries

From the $h^{th}$ stratum ($h = 1, 2, \ldots, L_t$), draw a psu ($i = 1, 2, \ldots, M_h$) with pps $W_{hi}$ or with equal probability as in the urban areas. Sample from this $i^{th}$ psu by a subsequent subsampling scheme involving ssus, tsus, etc., which permits the computation of an unbiased estimate, $x_{hi}'$, of the primary total, $X_{hi}$. Thus,

$$\hat{X}_{hi} = \frac{x_{hi}'}{W_{hi}} \quad (3.1)$$

is an unbiased estimate of the stratum total, $X_h$, from the $i^{th}$ psu.

The selection within psu is completely independent and is specified for all psus in case they are drawn. Replace all ssus, tsus, etc., and repeat the above sampling procedure
\( m_h \) times \((i = 1, 2, \ldots, m_h)\), yielding unbiased estimates, 
\[ \hat{X}_{h1}, \hat{X}_{h2}, \ldots, \hat{X}_{hm_h} \]
of the stratum total, \( X_h \). Thus, 
\[ \hat{X}_h = \frac{1}{m_h} \sum_{i=1}^{m_h} \hat{X}_{hi} \]  
(3.2)
is an unbiased estimate of \( X_h \), since it is a sample mean of size \( m_h \) from our infinite population of possible outcomes

\[ \hat{X}_{hi} = \frac{X_{hi}}{w_{hi}}. \]

If the variance of \( \hat{X}_{hi} \) is \( \sigma_{\hat{X}_{hi}}^2 \), then the variance of \( \hat{X}_h \) is

\[ \frac{\sigma_{\hat{X}_{hi}}^2}{m_h}. \]  
(3.3)

In general, to estimate unbiasedly the variance of any design in which psus are sampled with complete replacement, we use

\[ s_{\hat{X}_{hi}}^2 = \frac{\sum_{i=1}^{m_h} (\hat{X}_{hi} - \hat{X}_h)^2}{m_h - 1} \]  
(3.4)
as estimate of \( \sigma_{\hat{X}_{hi}}^2 \). Thus to estimate \( \frac{\sigma_{\hat{X}_{hi}}^2}{m_h} \), we use

\[ \frac{s_{\hat{X}_{hi}}^2}{m_h}. \]  
(3.5)

The above discussion is covered in Hartley's notes \[11\].

It will now be shown that the expectation of \( \hat{X}_h \) is equal to \( X_h \).
Given that
\[ \hat{X}_{hi} = \frac{x'_{hi}}{w_{hi}} \]

where
\[ x'_{hi} \] is an unbiased estimate of \( X_{hi} \), the \( i \)th psu total, and
\[ w_{hi} \] is the weight or probability,
then with
\[ E( | i ) \] as the average over a specific sub-population of the \( i \)th psu,
\[ E( ) \] as the average overall sub-population of \( i \)th psu (\( i = 1, 2, \ldots, M_{th} \)),
we have
\[ E(\hat{X}_{hi}) = E \left( \frac{x'_{hi}}{w_{hi}} \right) \]
\[ = E \left( E \left( \frac{x'_{hi}}{w_{hi}} \big| i \right) \right) \]
\[ = E \left( \frac{x_{hi}}{w_{hi}} \right) \]
\[ = \sum_{i}^{M_{h}} \frac{w_{hi} x_{hi}}{w_{hi}} \]
\[ = \sum_{i}^{M_{h}} x_{hi} \]
\[ = X_{h} \]

where
\( M_h \) is the total number of psu in the \( h^{th} \) stratum and

\( X_h \) is the \( h^{th} \) stratum total.

Finally, we have

\[
E(\hat{X}_h) = E \frac{1}{m_h} \sum_{i} \hat{X}_{hi} = \frac{1}{m_h} \sum_{i} E(\hat{X}_{hi}) = X_h.
\]

Thus, \( \hat{X}_h \) is unbiased.

5. **Details of methods of estimation**

The general characteristic will be denoted by \( x \). In the rural area,

\( x_{thijk} \) is the characteristic of the \( thijk \)th \( hh \) \((t = 1, 2)\)

and

\( x_{thij} \) is the characteristic of the \( thij \)th \( hh \) \((t = 3, 4)\)

for the urban area.

Capital letters will be used to designate parameters of the population. In general, a summation is replaced by dropping the corresponding subscript and the small \( x \) is replaced by the capital \( X \). Thus,
\[ X_{th} = \sum_{i}^{M_{th}} \sum_{j}^{B_{thi}} \sum_{k}^{N_{thij}} x_{thijk} \]

\[ = \sum_{ijk} x_{thijk} \]

\[ = \sum_{i}^{X_{thi}} \]

(3.8)

where

- \( X_{th} \) is the \( h^{th} \) stratum total of the \( t^{th} \) sector \((h = 1, 2, \ldots, L_t; t = 1, 2)\),
- \( M_{th} \) is the number of psus in the \( h^{th} \) stratum of the \( t^{th} \) sector \((i = 1, 2, \ldots, M_{th})\),
- \( B_{thi} \) is the number of ssus in the \( i^{th} \) psu; in the \( h^{th} \) stratum of the \( t^{th} \) sector \((j = 1, 2, \ldots, B_{thi})\), and
- \( N_{thij} \) is the number of tsus in the \( j^{th} \) ssu; in the \( i^{th} \) psu; in the \( h^{th} \) stratum of the \( t^{th} \) sector \((k = 1, 2, \ldots, N_{thij})\).

Also,

\[ X_t = \sum_{h}^{L_t} X_{th} \]

\[ = \sum_{h} X_{th} \]

(3.9)

is the sector total.

Similar notation is used for the urban area \((t = 3, 4)\).
Thus the country total is,

\[ X = \sum_{t=1}^{4} X_t \]

\[ = \sum_t X_t. \]  \hspace{1cm} (3.10)

The variance of an estimator, \( \hat{X}_{thi} \), will be described as

\[ \text{Var}(\hat{X}_{thi}) = \sigma^2_{\hat{X}_{thi}} \]  \hspace{1cm} (3.11)

and the covariance of \( \hat{X}_{thi} \) and \( \hat{X}_{thi'} \) \( (i \neq i') \) as

\[ \text{Cov}(\hat{X}_{thi}, \hat{X}_{thi'}) = \sigma_{\hat{X}_{thi}, \hat{X}_{thi'}} \]  \hspace{1cm} (3.12)

Small letters will be used to designate sample totals with corresponding subscripts as the population totals. A summation for the sample is designated by \( \sum' \) and the corresponding subscript. This summation is replaced by dropping or deleting the corresponding subscript from the small \( x \).

Thus, the \( i^{th} \) psu sample total from the \( th^{th} \) stratum in the rural sector is,

\[ x_{thi} = \sum_j \sum_k x_{thijk} \]

\[ = \sum' \sum' x_{thijk} \]

\[ = \sum' x_{thijk} \]  \hspace{1cm} (3.13)
where
\[ b_{thi} \] is the number of sample SSUs in the \( i \)th sample PSU of the \( th \) stratum,

and
\[ n_{thij} \] is the number of sample TSUs in the \( j \)th sample PSU; in the \( ith \) sample PSU of the \( th \) stratum.

Similarly,
\[ x_{th} = \sum_i x_{thi} \] (3.14)

is the \( th \) stratum sample total and
\[ x_t = \sum_h x_{th} \]
\[ = \sum_h x_{th} \] (3.15)

is the \( th \) sector sample total. Similar notations are used for the urban areas. Notation for estimators and estimates of variance will be described as they arise in the text.

From each stratum, we have five numbered PSUs selected with complete replacement and with PPS in the rural sectors but with equal probability in the urban sectors. Estimates of population totals and variances of estimates will be presented for the stratum, the sector, and the country as a whole.

The unbiased estimate of the \( th \) stratum total, \( x_{th} \), from the \( ith \) sample municipality is given by
\[ x^*_\text{thi} = \frac{P_{thi}}{p_{thi}} \frac{B_{thi}}{b_{thi}} \sum_j N_{n\text{thi}j} \sum_k x_{\text{thijk}} \]  

(3.16)

where

\( x_{\text{thijk}} \) is the characteristic of the \( k \)th sample hh; in the \( j \)th sample barrio (or poblacion); in the \( i \)th numbered psu \((i = 1, 2, 3, 4, 5)\); in the \( h \)th stratum \((h = 1, 2, \ldots, L_t)\) of the \( t \)th sector \((t = 1, 2)\),

\( n_{\text{thi}j} \) is the number of sample hh in the \( j \)th sample barrio (or poblacion); in the \( i \)th numbered psu; in the \( h \)th stratum of the \( t \)th sector,

\( N_{\text{n\text{thi}j}} \) is the number of hhs listed in the \( j \)th sample barrio; in the \( i \)th numbered psu; in the \( h \)th stratum of the \( t \)th sector,

\( b_{\text{thi}} \) is the number of sample barrios (or poblacion) in the \( i \)th numbered psu; in the \( h \)th stratum of the \( t \)th sector. This is equal to two for the barrio and to one for the poblacion,

\( B_{\text{thi}} \) is the number of barrios (one for poblacion) in the \( i \)th numbered psu; in the \( h \)th stratum of the \( t \)th sector,

\( P_{\text{thi}} \) is the population of the \( i \)th numbered psu; in the \( h \)th stratum of the \( t \)th sector,

and
\(P_{th}\) is the population in the \(h^{th}\) stratum of the \(t^{th}\) sector.

Population is defined as the 1948 census population projected to 1955. Note that

\[
P_{1hi} = P_{2hi} = P_{hi}
\]

and

\[
P_{1h} = P_{2h} = P_{h}.
\]

Equation (3.16) may be written as

\[
x_{thi}^* = \frac{x_{thi}}{w_{hi}}
\]

where

\[
x_{thi}' = \frac{b_{thi}}{b_{thi}} \sum' \sum' n_{thi} \sum' x_{thijk}
\]

is an unbiased estimate of the corresponding \(i^{th}\) psu total, \(x_{thi}\), and

\[
w_{hi} = \frac{P_{hi}}{P_{h}}.
\]

By virtue of the results in Chapter III (Section B4),

\[
E(x_{thi}^*) = x_{th}
\]

and the best linear unbiased estimate of \(x_{th}\) is the mean of the five independent estimates which is

\[
x_{th}^* = \frac{1}{5} \sum' x_{thi}^*,
\]

with variance equal to
where \( \sigma^2_{th} \) is the variance of \( X_{thi} \).

For processing mass data, (3.16) and (3.18) are very inconvenient and laborious to use. Consider the following two relations:

\[
(a) \quad R_t = \frac{1}{m} \frac{\bar{E}}{p} \cdot \frac{\bar{B}_t}{b_t} \cdot \frac{\bar{N}_t}{n_t}
\]

where

- \( R_t \) is the overall raising factor for the \( t \)th sector \((t = 1, 2)\),
- \( \bar{n}_t \) is the expected average number of sample hhs per barrio (or poblacion),
- \( \bar{N}_t = \sum_{hi} \frac{N_{thij}}{} = \frac{N_t}{\bar{B}_t} \) is the average number of hhs listed per barrio (or poblacion),
- \( \bar{b}_t \) is the expected average number of sample barrio (or poblacion). This is equal to two for the barrio and to one for the poblacion,
- \( \bar{B}_t = \sum_{hi} \frac{B_{thi}}{} = \frac{B_t}{M_t} \) is the average number of barrios (one for the poblacion) per psu,
\[ \bar{p} = \frac{\sum_{hi} P_{thi}}{\sum_{h} M_{th}} = \frac{P_t}{M_t} \] is the average population per psu,
\[ \bar{P} = \frac{\sum_{h} P_{th}}{L_t} = \frac{P_t}{L_t} \] is the average population per stratum

and

\[ m = 5 \] is the number of sample psu in the th stratum and is constant for all th;

(b) \[ f_{thi} = \frac{n_{thij}}{N_{thij}} \]
\[ = \frac{1}{m} \frac{P_{thi}}{B_{thi}} \frac{B_{thi}}{R_t} \]

where

\( f_{thi} \) is the th psu constant and determines \( n_{thij} \) from \( f_{thi} N_{thij} \) for all \( j = 1, 2, \ldots \), \( B_{thi} \) in the th psu. The other symbols are as defined in (3.16) and (3.20). The overall raising factor for the barrio is \( R_1 = 850 \) and for the poblacion, \( R_2 = 450 \).

With the use of (3.20) and (3.21), a self-weighting estimator of \( X_{th} \) from the i th psu was utilized and is given by
\[ \hat{X}_{thi} = \frac{\bar{P} \bar{B}_{t}}{p_b} \frac{N_t}{n_t} \sum' \sum'' x_{thijk} \]
\[ = m R_t x_{thi} \]

where the symbols are as defined in (3.20) and (3.21). The
mean of the five \( (m) \) independent estimates, \( \hat{x}_{thi}'s \), will give

\[
\hat{x}_{th} = \frac{1}{m} \sum_{i}^{'} \hat{x}_{thi}
\]

(3.23)

\[= R_t x_{th}.\]

This estimation procedure is very simple to compute.

Since we are using population averages, \( \bar{P}, \bar{p}, \bar{E}_t, \bar{E}_t, \)
\( \bar{n}_t \) and \( \bar{N}_t \) in (3.23) for \( P_h, \phi_i, B_{thi}, b_{thi}, n_{thi} \) and \( N_{thij} \) in
(3.16), respectively, a bias may have been introduced in \( \hat{x}_{th} \).

Dropping the subscript \( t \) for simplicity, the bias is given

by the difference

\[
D_h = E[\hat{x}_h^* - \hat{x}_h].
\]

Substitute (3.18) for \( \hat{x}_h^* \) and (3.22) for \( \hat{x}_h \), we get

\[
D_h = E\left[\frac{1}{m} \sum_{i}^{'} \frac{P_h}{\phi_i} \frac{B_{hi}}{b_{hi}} \sum_{j}^{'} \frac{N_{hi}}{n_{hi}} \sum_{k}^{'} x_{hijk} - \frac{1}{m} \sum_{i}^{'} \frac{\bar{P}}{\bar{p}} \cdot \frac{\bar{E}_t}{\bar{b}} \sum_{j}^{'} \sum_{k}^{'} x_{hijk}\right].
\]

(3.24)

This simplifies into

\[
D_h = E\left[\frac{1}{mb} \sum_{i}^{'} \left\{ \frac{P_h}{\phi_i} \frac{B_{hi}}{b_{hi}} \frac{1}{r_{hi}} - \bar{P} \bar{E}_t \bar{N} \right\} x_{hi}\right]
\]

(3.25)

\[
= E\left[\frac{1}{mb} \sum_{i}^{'} (u_{hi} - \bar{V}) x_{hi}\right]
\]

\[
= E\left[\frac{1}{mb} \sum_{i}^{'} d_{hi} x_{hi}\right]
\]
where

\[ u_{h1} = p_{h1} \frac{B_{h1}}{p_{h1}} \frac{1}{f_{h1}} , \]

\[ V = \frac{\bar{P}}{h} \frac{\bar{N}}{n} , \]

\[ d_{h1} = u_{h1} - V , \]

\[ f_{h1} = \frac{n_{h1}^{2}}{N_{h1}} , \]

\[ b = b_{h1} . \]

Because of the independence of the psus,

\[ D_h = \frac{1}{mb} \sum_{i} E(\sigma_{d_{h1},x_{h1}}) \]

\[ = \frac{1}{mb} \sum_{i} [ \rho \sigma_{d_{h1},x_{h1}} + E(d_{h1}) E(x_{h1}) ] \]

\[ = \frac{1}{b} \left[ \rho \sigma_{d_{h1},x_{h1}} + E(d_{h1}) E(x_{h1}) \right] \]

We can make

\[ D_h = 0 \]

by letting

\[ d_{h1} = 0 . \]

Thus

\[ d_{h1} = 0 \]

implies

\[ u_{h1} = V . \]

A condition for unbiasedness is given by the relation
Let us now consider the relations imposed on $X_{th}$ by examining (3.20) and (3.21). Solving for $f_{hi}$ in terms of $R$, we get

$$f_{hi} = \frac{1}{b_{hi}} \left( \frac{P}{P_{hi}} B_{hi} \right) = \frac{P_{hi}}{P} \frac{B_{hi}}{B} \frac{1}{N}$$

or

$$P_{hi} \frac{B_{hi}}{P_{hi}} \frac{1}{f_{hi}} = \frac{P}{P} \frac{B}{B} \frac{1}{N}$$

where

$$\overline{b} = b_{hi}.$$

This is the same as the condition for unbiasedness given in (3.27). In general, since $f_{thi}$ is a fraction also defined by

$$\frac{n_{thi}}{N_{thi}} (j = 1, 2, \ldots, B_{thi}),$$

it will not be possible to find barrio or poblacion sample sizes $n_{thi}$ satisfying

$$f_{thi} = \frac{n_{thi}}{N_{thi}}$$

exactly. Apart from these rounding errors, however, our sampling scheme is self-weighting and hence unbiased. Thus $\hat{X}_{th}$ is unbiased and its variance is given by (3.19).

Since $\hat{X}_{th}$ is unbiased, then the estimate of the sector
total, $X_t$, is

$$
\begin{align*}
\hat{X}_t &= \sum_h \hat{X}_{th} \\
&= \sum_h R_t x_{th} \\
&= R_t x_t
\end{align*}
$$

and its variance is

$$
\sigma^2_{\hat{X}_t} = \sum_h \frac{\sigma^2_{x_{th}}}{S_t^2} .
$$

We now turn to the urban areas. As in the rural areas, the estimate of the $h^{th}$ stratum total in the $t^{th}$ sector, $X_{th}$, as given by the $i^{th}$ precinct (psu) is

$$
X_{thi} = M_{th} \sum J x_{thij}
$$

where

- $x_{thij}$ is the characteristic of the $j^{th}$ sample hh; in the $i^{th}$ numbered sample psu ($i = 1, 2, 3, 4, 5$); in the $h^{th}$ stratum ($h = 1, 2, \ldots, L_t$) of the $t^{th}$ sector ($t = 3, 4$),

- $n_{thi}$ is the number of sample hhs; in the $i^{th}$ numbered sample psu; in the $h^{th}$ stratum of the $t^{th}$ sector,

- $N_{thi}$ is the number of hhls listed in the $i^{th}$ numbered sample psu; in the $h^{th}$ stratum of the $t^{th}$ sector,
$M_{th}$ is the number of psus in the $h^{th}$ stratum of the $t^{th}$ sector.

Note that psus were drawn with equal probability and complete replacement.

From the results of Chapter III B4,

\[ E(X_{th}^*) = E \left[ \frac{X_{th1}'}{W_{th1}} \right] \]
\[ = X_{th} \]

(3.33)

where

\[ X_{th1}' = \frac{N_{th1}}{n_{th1}} \sum_{j} x_{th1j} \]

is an unbiased estimate of the $i^{th}$ psu total, $X_{th1}$ and

\[ W_{th1} = \frac{1}{M_{th}} . \]

Thus $X_{th1}'$ is unbiased and the best linear unbiased estimate of $X_{th}$ is

\[ X^*_{th} = \frac{1}{5} \sum_{1} x^*_{th1} \]

(3.34)

with variance

\[ \frac{\sigma^2_{th}}{5} \]

(3.35)

where $\sigma^2_{th}$ is the variance of $X^*_{th1}$. Equations (3.34) and (3.35) are of the same form as (3.18) and (3.19), respectively.

In estimating the parameters of many characteristics,
Equation (3.34) becomes rather unwieldy. The actual estimator used to estimate $X_{th}$ is

$$
\hat{X}_{th} = \frac{1}{m} \bar{M}_t \frac{N_t}{n_t} \sum'_i \sum'_j x_{thij}
$$

$$
= R_t x_{th}
$$

where

$$
R_t = \frac{1}{m} \bar{M}_t \frac{N_t}{n_t}
$$

is the overall raising factor for the $t$th sector ($t = 3, 4$),

$\bar{n}_t$ is the expected number of sample hhs per psu,

$$
\bar{N}_t = \frac{\sum N_{thi}}{\sum M_{th}} = \frac{N_t}{M_t}
$$

is the average number of hhs listed per psu,

$$
\bar{M}_t = \frac{\bar{h}}{L_t} = \frac{M_t}{L_t}
$$

is the average number of psu per stratum,

and the stratum constant, $f_{th} = \frac{n_{thi}}{N_{thi}}$, is given by

$$
f_{th} = \frac{1}{m} \frac{M_{th}}{R_t}.
$$

The overall raising factor for the capitals and cities was $R_3 = 450$, while for Manila, $R_4$ was 400. Substituting (3.26) in (3.27), we get

$$
f_{th} = \frac{M_{th}}{M_t} \frac{\bar{n}_t}{N_t}.
$$
or
\[
\frac{M_{th}}{f_{th}} = \frac{\bar{M}_t}{\bar{n}_t} \quad \text{(3.39)}
\]

As in the estimator for the rural areas, the bias is
\[
D_{th} = E\left[\hat{x}_{th}^* - \hat{x}_{th}\right] \\
= E\left[\frac{1}{m} \sum_{i=1}^{m'} \left\{ \frac{M_{th}}{f_{th}} \left( \frac{1}{f_{th}} \frac{\bar{N}_t}{\bar{n}_t} \right) x_{thi} \right\} \right] \quad \text{(3.40)}
\]

A condition for unbiasedness is derived, if we make
\[
D_{th} = 0
\]
or
\[
M_{th} \frac{1}{f_{th}} - \bar{M}_t \frac{\bar{N}_t}{\bar{n}_t} = 0.
\]

This will give
\[
\frac{M_{th}}{f_{th}} = \frac{\bar{M}_t}{\bar{n}_t} \quad \text{(3.41)}
\]
or
\[
f_{th} = \frac{M_{th}}{m\bar{n}_t} \quad \text{(3.42)}
\]

which is similar to the relation given in (3.39). It should be pointed out that since \( f_{th} \) is a proportion defined also by
\[
\frac{n_{thi}}{N_{thi}} \quad (i = 1, 2, \ldots, M_{th})
\]
it is not possible to find precinct sample sizes \( n_{thi} \) which will exactly satisfy
\[
f_{th} = \frac{n_{thi}}{N_{thi}}.
\]

Apart from these rounding errors which in general are small,
our estimation procedure is self-weighting and therefore unbiased. This result implies that \( \hat{X}_{th} \) is unbiased and its variance is similar to (3.19).

Utilizing the results given in (3.35) and (3.36), the unbiased estimate of the \( t \)th sector total, \( X_t \), is

\[
X_t = \sum_h \hat{X}_{th} = \sum_h R_t x_{th} = R_t x_t
\]

(3.43)

and its variance is

\[
\sigma^2_{\hat{X}_t} = \sum_h \sigma^2_{\hat{X}_{th}}
\]

(3.44)

which is of the same form as (3.30) and (3.31).

An unbiased estimate of the country total, \( X \), is given by summing the unbiased estimates for the four sectors \( (t = 1, 2, 3, 4) \). This is

\[
\hat{X} = \sum_t \hat{X}_t
\]

(3.45)

which is the sum of (3.30) and (3.43) and its variance is

\[
\sum_t \sum_h \sigma^2_{\hat{X}_{th}}
\]

(3.46)

The variance of \( \hat{X}_{th} \), as given in (3.19), irrespective of sector is

\[
\frac{\sigma^2_{\hat{X}_{th}}}{5}
\]
From Chapter III B4, an unbiased estimate of $s_{th}^2$ is

$$s_{th}^2 = \frac{1}{m-1} \sum' (\hat{x}_{th1} - \bar{x}_{th})^2. \quad (3.47)$$

Using sample totals, (3.47) becomes

$$s_{th}^2 = \frac{(m R_t)^2}{m-1} \left[ \sum' x_{th1}^2 - \frac{x_{th}^2}{m} \right] \quad (3.48)$$

where

- $R_t$ is the overall raising factor for the $t^{th}$ sector ($t = 1, 2, 3, 4$),
- $m = 5$,
- $x_{thi}$ is the corresponding $i$th PSU sample total,
- and
- $x_{th}$ is the corresponding $h^{th}$ stratum sample total.

It should be noted that (3.48) is very simple to compute.

Since $s_{th}^2$ is an unbiased estimate of $\sigma_{th}^2$ ($t = 1, 2, 3, 4$) and independent sampling was considered separately for each stratum, then to estimate the stratum, sector and country variances, we may use

$$s_{x_{th}}^2 = \frac{s_{th}^2}{m} \quad (3.49)$$

to estimate $\frac{s_{th}^2}{m}$;

$$s_{x_{th}}^2 = \sum h \frac{s_{th}^2}{m} \quad (3.50)$$

to estimate $\sum h \frac{s_{th}^2}{m}$; and
\[ s^2_{\hat{x}} = \sum_t \sum_h \frac{s^2_{th}}{m} \]  

(3.51)

to estimate \[ \sum_t \sum_h \frac{s^2_{th}}{m} \].

We will refer to (3.49), (3.50) and (3.51) as the standard or 'stratified' estimates of the variances of \( \hat{x}_{th} \), \( \hat{x}_t \), and \( \hat{x} \), respectively.

The estimation of the sector variance is now described. The use of (3.50) as estimate of the variance of \( \hat{x}_t \) becomes rather unwieldy since it requires separate computation of \( s^2_{th} \) for each \( h \) of the sector. A short cut estimator which may be used to estimate

\[ \sum_h \sigma^2_{th} \]

is given by

\[ \hat{\sigma}^2_{\hat{x}_t} = \frac{1}{m} \sum_{i=1}^{m} \frac{(\hat{x}_{ti} - \hat{x}_t)^2}{m-1} \]  

(3.52)

where

\[ \hat{x}_{ti} = \sum_h x_{thi} \]

\[ = m R_t x_{ti} \]

and

\( \hat{x}_t \) is as defined in (3.30) and (3.43).

It should be noted that the numbering \( i = 1, 2, 3, 4, 5 \) of the psus in each stratum should be independent and at
random so that each of the \( \hat{X}_{t1} \) can be regarded as an independent estimate of \( X_t \). With certain items the estimate (3.52) based on 4 df may not be sufficiently accurate, and the standard estimator

\[
\sum_h \frac{s_{th}^2}{m}
\]

should be used. The form of the estimator (3.52) of \( \sigma^2_{x_t} \) is similar for both rural and urban sectors. The difference is in the values of \( R_t \) and \( x_{t1} \).

Using sample totals, (3.52) becomes

\[
\hat{\sigma}^2_{x_t} = \frac{m R_t^2}{m - 1} \left[ \sum_i x_{t1}^2 - \frac{x_{t}^2}{m} \right].
\]  

(3.53)

This estimate is very simple to compute. Let us call \( \hat{\sigma}^2_{x_t} \) as the 'survey' variance of \( \hat{X}_t \).

To compare \( \hat{\sigma}^2_{x_t} \) and \( \sigma^2_{x_t} \), i.e., the 'survey' estimate with the 'stratified' estimate, the analysis of variance shown in Table 2 is utilized. The between primary numbers \( i = 1, 2, 3, 4, 5 \) with 4 df is used for obtaining the 'survey' estimate of variance of \( \hat{X}_t \). For the standard or 'stratified' estimator, we used the between primary numbers \( i = 1, 2, 3, 4, 5 \) with 4 df plus the residual with \( 4(L - 1) \) df or a total \( 4L \) df for primaries within strata. Thus, the short cut method of estimating the variance of the sector total estimate, \( \hat{X}_t \), is
Table 2. Analysis of variance

<table>
<thead>
<tr>
<th>Sources of variation</th>
<th>Degrees of freedom</th>
<th>Sum of squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total primary variation</td>
<td>5L - 1</td>
<td>[ \sum \sum' \hat{x}_{th1}^2 - CT ]</td>
</tr>
<tr>
<td>Between strata</td>
<td>L - 1</td>
<td>[ \frac{1}{m} \sum_h \left( \sum_i \hat{x}_{th1} \right)^2 - CT ]</td>
</tr>
<tr>
<td>Between primary numbers</td>
<td>4</td>
<td>[ \frac{1}{L} \sum_i \left( \sum_h \hat{x}_{th1} \right)^2 - CT ]</td>
</tr>
<tr>
<td>Residual</td>
<td>4(L - 1)</td>
<td>[ \sum \sum' \hat{x}<em>{th1}^2 - \frac{1}{5} \sum_h \left( \sum_i \hat{x}</em>{th1} \right)^2 - \frac{1}{L} \sum_i \left( \sum_h \hat{x}_{th1} \right)^2 + CT ]</td>
</tr>
</tbody>
</table>

where

\[ CT = \left( \sum_h \sum_i \hat{x}_{th1} \right)^2 / 5L \]

\[ = \frac{5R_t^2}{L} \left[ \sum_h \sum_i x_{th1} \right]^2 \]
\[
\frac{1}{5} \left\{ \sum' \left[ \sum \hat{x}_{th1} \right]^2 - \left[ \sum' \sum \hat{x}_{th1} \right]^2 / 5 \right\}
\]

which is equal to

\[
\hat{\sigma}_{\hat{x}_t}^2 = \frac{1}{5} \frac{1}{4} \left\{ \sum' \hat{x}_{th1}^2 - \left[ \sum' \hat{x}_{th1} \right]^2 / 5 \right\}.
\]

This is similar to (3.52). The standard method is given by

\[
s^2_{\hat{x}_t} = \frac{1}{5} \left\{ \sum h \sum' \hat{x}_{th1}^2 - \sum h \left[ \sum' \hat{x}_{th1} \right]^2 / 5 \right\}
\]

\[
= \frac{1}{5} \sum h \left\{ \sum' \hat{x}_{th1}^2 - \left[ \sum' \hat{x}_{th1} \right]^2 / 5 \right\}
\]

\[
= \frac{1}{5} \sum h s^2_{th}
\]

which is similar to (3.50).

If the numbering \( i = 1, 2, 3, 4, 5 \) is independent and is done at random, then the 'survey' estimate, \( \hat{\sigma}_{\hat{x}_t}^2 \) is unbiased. This estimate is preferred because of computational considerations. It is, of course, of low precision. Moreover, if there is a systematic correlation between the numbers \( i = 1, i = 2, \ldots, i = 5 \) psu in the various strata, \( \hat{\sigma}_{\hat{x}_t}^2 \) will be an overestimate.
C. Comparison of 'Survey' and 'Stratified' Estimates of Variance

Sample-wise, a difference will exist between $\hat{\sigma}^2_{X_t}$ and $s^2_{X_t}$ for a given characteristic. This difference is apparent from the analysis of variance.

In Table 3 are shown values of $\hat{X}_t$, $\hat{X}$, $\hat{\sigma}^2_{X_t}$, $\hat{\sigma}^2_{X}$, $s^2_{X_t}$ and $s^2_{X}$ for the May, 1956 and October, 1956 PSSH samples, where the characteristic, $x$, is the population in interviewed hhs. Note that $\hat{\sigma}^2_{X_t}$ is greater than $s^2_{X_t}$ in the rural areas ($t = 1, 2$) but is lesser in the urban areas ($t = 3, 4$). There is apparently a cancellation effect and the value of $\hat{\sigma}^2_{X}$ is now slightly larger than $s^2_{X}$ for the country estimate, $\hat{X}$. These results indicate that some systematic biases had been introduced into the numbering of the psus in the various strata. To correct these suspected biases, an independent and random numbering of the psus within each stratum should be initiated.
Table 3. Comparison of survey and stratified standard deviations and coefficients of variation of corresponding sector and country estimates of population in interviewed households for May and October, 1956 PSSH samples

<table>
<thead>
<tr>
<th>Sector</th>
<th>( t )</th>
<th>( \hat{X} )</th>
<th>May, 1956</th>
<th>( \sigma )</th>
<th>October, 1956</th>
<th>( \sigma )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td>Barrio</td>
<td>1</td>
<td>X_1</td>
<td>13.954</td>
<td>0.943 (6.76)</td>
<td>0.622 (4.46)</td>
<td>0.954 (6.77)</td>
</tr>
<tr>
<td>Poblacion</td>
<td>2</td>
<td>X_2</td>
<td>2.848</td>
<td>0.252 (8.84)</td>
<td>0.190 (6.67)</td>
<td>0.248 (8.55)</td>
</tr>
<tr>
<td>Capitals</td>
<td>3</td>
<td>X_3</td>
<td>2.902</td>
<td>0.086 (2.98)</td>
<td>0.129 (4.45)</td>
<td>0.084 (2.85)</td>
</tr>
<tr>
<td>and cities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manila</td>
<td>4</td>
<td>X_4</td>
<td>1.833</td>
<td>0.056 (3.06)</td>
<td>0.085 (4.67)</td>
<td>0.032 (1.79)</td>
</tr>
<tr>
<td>Philippines</td>
<td></td>
<td>X</td>
<td>21.527</td>
<td>0.784 (3.64)</td>
<td>0.669 (3.11)</td>
<td>0.708 (3.27)</td>
</tr>
</tbody>
</table>


\( ^b \)Number in parenthesis ( ) is the coefficient of variation (CV) in per cent where \( CV = \frac{\text{Standard deviation}}{\text{Estimate}} \times 100. \)
IV. ROTATION OF SAMPLE

A. Development of a Rotation Scheme

1. Purpose

One of the most important requirements of the design of the PSSH is an accurate knowledge of the number of hhs in the sampling area (barrio, poblacion or precinct). This is needed in order to determine the number of sample hhs to be interviewed from each of these sampling areas (sas). The number of sample hhs is obtained from the relation

\[ n_{thij} = f_{thi} N_{thij} \]  \hspace{1cm} (4.1)

where

- \( f_{thi} \) is the \( i^{th} \) psu constant of the \( h^{th} \) stratum in the \( t^{th} \) sector, \( (t = 1, 2) \),

and

- \( N_{thij} \) is the number of listed hhs in the \( thij \)th barrio (or poblacion) of the rural area;

or

\[ n_{thi} = f_{th} N_{thi} \]  \hspace{1cm} (4.2)

where

- \( f_{th} \) is the \( h^{th} \) stratum constant in the \( t^{th} \) sector, \( (t = 3, 4) \),

and

- \( N_{thi} \) is the number of listed hhs in the \( th \)th precinct
of the urban area.

To obtain $N_{thj}$ or $N_{thi}$, for each visit a listing operation was instituted into the field operation. Listing (the first phase) is done a week or so prior to actual interviewing of the hh (the second phase) and consists of canvassing and travelling over the sa in order to obtain actual counts of hhs, population in hhs, and occupation of head of hh in each of the sas. About 0.163 million hhs and 0.969 million population are listed for each visit. These figures give an idea of the extent of the listing operation. It is estimated that about 15 per cent of the overall field cost is devoted to the actual listing operation alone. The extent of the listing operation may also be indicated by the ratio of the average number of hhs interviewed to the average number of hhs listed in each sa for each of the four sectors. These ratios are shown in column 6, Table 4. All ratios are small with extremely low ratios in the poblacion sector (1:62) and in Metropolitan Manila (1:27). Increase of these ratios will be taken into consideration in the development of a new rotation scheme.

In addition, it was reported that during the third or fourth visit, the respondents have shown some signs of being uncooperative. It becomes highly desirable to institute a rotation scheme which will minimize this response resistance and at the same time reduce the cost due to listing.
Table 4. The number of sampling areas, average number of interviewed (sample) households per sampling area, average number of listed households per sampling area, and the ratio of average number of interviewed households to average number of listed households per sampling area for each of the four sectors, May, 1958 PSSH sample

<table>
<thead>
<tr>
<th>Sector</th>
<th>Sector number</th>
<th>Number of sampling areas</th>
<th>Average number of households per sampling area</th>
<th>Ratio (4):(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Interviewed</td>
<td>Listed</td>
</tr>
<tr>
<td>Barrio</td>
<td>1</td>
<td>300</td>
<td>9.3</td>
<td>169.7</td>
</tr>
<tr>
<td>Poblacion</td>
<td>2</td>
<td>150</td>
<td>7.8</td>
<td>484.3</td>
</tr>
<tr>
<td>Capitals and cities (precincts)</td>
<td>3</td>
<td>150</td>
<td>8.2</td>
<td>132.7</td>
</tr>
<tr>
<td>Manila (precincts)</td>
<td>4</td>
<td>160</td>
<td>4.4</td>
<td>119.6</td>
</tr>
</tbody>
</table>

Some conditions which have to be met in the development of the new rotation scheme are:

a) The survey will be conducted twice a year, possibly with visits at six-month intervals.

b) Feasibility and simplicity of operation commensurate with existing resources and conditions.

The basic idea is to split the barrio or poblacion (ssu) and the precinct (psu) into segments and to carry out the listing operation only for a sample of three of the segments to be rotated from visit to visit.
2. Description of the new rotation scheme

The segmentation procedure (referred to above) is done in the central office prior to the start of the field work in accordance with the rotation pattern. A new subscript, \( \kappa \), is added, whenever necessary, to indicate the time of visit. Thus, \( \kappa = 0 \) will indicate the time for the initial visit.

The steps in the segmentation procedures are as follows:

a) With the aid of up-to-date maps, each sa (barrio, poblacion or precinct) is divided into \( S_{thij} \) \((t = 1, 2)\) or \( S_{thi} \) \((t = 3, 4)\) segments, each segment with distinct 'natural' boundaries. Each hh in the sa must be associated with one and only one segment.

b) The number of segments, \( S_{thij}(S_{thi}) \), in each sa must be

\[
3 \leq S_{thij}(S_{thi}) \leq \text{convenient size.} \quad (4.3)
\]

The convenient size will be dictated by the availability of natural boundaries such as roads, streams or creeks, barrio or town boundaries, etc., and by the restriction that for any visit \( \kappa \), the size of sample within segment, \( \alpha_n^{thijk}(\alpha_n^{thij}) \) must be less than or equal to the actual number of listed hhs in the segment, \( \alpha_{Nthijk}(\alpha_{Nthij}), \text{i.e.} \)

*The 'up-dating of maps' by interviewers is referred to in Chapter IV.
\[ a_{n_{thijk}} \leq a_{N_{thijk}} \]

or

\[ a_{n_{thij}} \leq a_{N_{thij}} . \]

c) Starting from some designated point in the sample and following a random pattern, the segments are numbered from 1 to \( S_{thij}(S_{thi}) \).

d) Arrange the \( S_{thij}(S_{thi}) \) segments into permutation or rotation groups (rg), each group consisting of three segments. One rg will be interviewed for each visit. A segment will be in the sample for three consecutive visits and then dropped from the sample. It may come in again in subsequent visits, the period will depend on the number of segments and also on the rotation pattern. Under this system of rotation 2/3 or 67 per cent of the segments are common from visit to visit and 1/3 or 33 per cent from year to year.

e) For two surveys a year, the rotation pattern for \( S_{thij}(S_{thi}) = 9 \) is given in Table 5. The three numbers in brackets refer to the segment numbers. They need not appear consecutive in the actual rotation group. For simplicity, the numbering is made consecutive. Thus, segments 1, 2 and 3 are visited at the \( \alpha = 1 \) visit; segments 2, 3, and 4 at the \( \alpha = 2 \) visit and so on. Segments 3 and 4 are common for \( \alpha = 2 \) and \( \alpha = 3 \) visits and segment 3 is common for \( \alpha = 1 \) and
Table 5. Rotation scheme for nine segments in sampling area

<table>
<thead>
<tr>
<th>Year</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visit number</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
</tbody>
</table>

1 2 2
3 3 3
4 4 4
5 5 5
6 6 6
7 7 7
8 8 8
9 9 9
1 1 1
2 2 2
3 3

\( \alpha = 3 \) visits.

At the \( \alpha \) visit, the interviewer will be instructed, through the regional supervisor, as to the composition of the rg, i.e., which two segments are matched and which is new.

B. Sampling Rate within Segments

The reduction in cost due to actual listing will depend upon the number of segments, \( S_{thi} \) (or \( S_{thi} \)) into which the sampling area (barrio, poblacion or precinct) is segmented. A simplified relation of the number of segments, \( S_{thi} \), and the proportional reduction in cost due to listing is shown
below, assuming that the listing cost is proportional to the number of segments in the sample at a particular visit:

<table>
<thead>
<tr>
<th>S_{thij}</th>
<th>Proportional reduction in cost due to listing:</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>6/7</td>
</tr>
<tr>
<td>15</td>
<td>4/5</td>
</tr>
<tr>
<td>9</td>
<td>2/3</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

In general, the larger the number of segments, the greater is the proportional reduction in cost due to listing.

With the use of this rotation scheme, rough estimates of the ratio of average number of hhs interviewed to the average number of hhs listed per sampling area are obtained for each sector. For the barrio, capitals and cities the ratio is increased to about 1:10 and for the poblacion and Manila, the ratio is about 1:15. These ratios may be compared with column 6, Table 4. This indicates a reduction of about 60 per cent of the cost due to listing or an estimated saving of approximately 14,000 a year at two surveys a year.

With segmentation and with the use of the rotation scheme described in Section A2, the problem of sampling rate within segments (srws) arises. Three aspects of this problem will be discussed. The first will deal on equal or unequal srws; the second on the field procedure to be followed at the visit; and the third on the relation of srws and the estima-
tion procedure.

1. **Equal sampling rate**

   The segment in each sa will be considered as an additional stage-unit. The derivation of equal srws is given for both rural and urban areas, and a simplified field procedure is described for the urban area only. With equal srws, the forms of the estimation procedures remain unchanged.

   In the development of the design for the rural area of the PSSH in Chapter III, two basic relations were utilized to make the estimation procedures self-weighting and unbiased. For easy reference, these relations are reproduced here.

\[
R_t = \frac{1}{5} \frac{P_t}{p_t} \frac{B_t}{b_t} \frac{N_t}{n_t}
\]

(4.5)

and

\[
f_{thi} = \frac{1}{5} \frac{P_{thi}}{p_{thi}} \frac{B_{thi}}{b_{thi}} \frac{1}{R_t}
\]

(4.6)

where the symbols are as defined in (3.27) and (3.28).

After segmentation, consider the segments as the third-stage units or tertisries and the hhs as fourth-stage units. Let us now apply a restriction,

\[
\frac{\bar{N}_t}{N_t} = \frac{\bar{s}_t}{s_t} \cdot \frac{\bar{n}_t}{n_t}
\]

(4.7)

where

\(\bar{N}_t\) and \(\bar{n}_t\) are as defined in (3.27),
and

\[ \bar{n}_t \] is the expected average number of sample hhs per segment and is considered fixed,

\[ \bar{N}_t = \sum \frac{N_{thijk}}{\sum \bar{s}_{thij}} \]

is the average number of hhs per segment,

\[ \bar{s}_t \] is the expected average number of sample segments per barrio (or poblacion) and is considered fixed (= 3),

and

\[ \bar{s}_t = \sum \frac{s_{thij}}{\sum B_{thi}} \]

is the average number of segments per barrio (or poblacion).

With (4.7) in (4.5), (4.5) becomes

\[ R_t = \frac{1}{5} \frac{P_t}{p_t} \frac{B_t}{b_t} \cdot \frac{\bar{s}_t}{\bar{n}_t} \cdot \bar{N}_t \]

(4.8)

and the srws is

\[ g_{thij} = \frac{1}{5} \frac{P_t}{p_{thi}} \frac{B_{thi}}{B_{thi}} \cdot \frac{s_{thij}}{\bar{s}_{thij}} \frac{1}{R_t} \]

(4.9)

which simplifies into

\[ g_{thij} = f_{thi} \frac{s_{thij}}{\bar{s}_{thij}} \]

(4.10)

where \( f_{thi} \) is given by (4.6) and \( \bar{s}_{thij} = 3 \).

Thus the srws or the so-called sa (barrio or poblacion) constant, \( g_{thij} \), does not depend on \( \alpha \). The restriction on
$s_{thij}$ is

$$f_{thi} \frac{s_{thi j}}{3} \leq 1. \quad (4.11)$$

Since $f_{thi}$ is a psu constant, equation (4.11) can be written as

$$S_{thij} \leq \frac{3}{f_{thi}}. \quad (4.12)$$

This restriction on $S_{thij}$ will depend primarily on $f_{thi}$. The range of values of $f_{thi}$ is between 0.007 to 0.185 for the barrio; and, 0.005 to 0.074 for the poblacion. For the urban area the relations are

$$R_t = \frac{1}{5} \overline{M}_t \overline{N}_t \frac{\overline{n}_t}{n_t} \quad (4.13)$$

and

$$f_{th} = \frac{1}{5} M_{th} \frac{1}{R_t} \quad (4.14)$$

where the symbols $R_t$, $\overline{M}_t$, $\overline{N}_t$, $\overline{n}_t$, $f_{th}$, and $M_{th}$ are as defined in (3.38) and (3.42) for $t = 3, 4$.

With segmentation, the segments are considered as second-stage units or secondaries and hhs as tertiaries. With the same restriction as (4.7) imposed on (4.13), we get

$$R_t = \frac{1}{5} \overline{M}_t \frac{s_t \overline{N}_t}{S_t \overline{n}_t} \quad (4.15)$$

and the srws for the urban area is

$$s_{thi} = \frac{1}{5} M_{th} \frac{s_{thi} 1}{R_t} \quad (4.16)$$

With (4.14) in (4.16), we get
where

\[
\varepsilon_{thi} = f_{th} \frac{S_{thi}}{s_{thi}} \quad (4.17)
\]

\(f_{th}\) is the stratum constant,

\(S_{thi}\) is the number of segments in the \(th\) th PSU,

and

\(s_{thi}\) is equal to three.

The restriction on \(\varepsilon_{thi}\) is

\[
S_{thi} \leq \frac{3}{f_{th}} \quad (4.18)
\]

and the range of values of \(f_{th}\) is between 0.046 to 0.088 for the capitals and cities and between 0.031 to 0.047 for Manila.

The field procedure at the \(\text{\#} \) visit with equal srws is described for the urban area only. There are two types of segments, namely: new segment and matched segment.

The field procedure in a new segment is as follows:

a) A random start is drawn in the interval 1 to \(\varepsilon_{thi}\).

Let this number be \(q\). This \(q\) th hh and every \(\frac{1}{\varepsilon_{thi}}\) th hh thereafter is interviewed. The srws, \(\varepsilon_{thi}\), is obtained from,

\[
\varepsilon_{thi} = f_{th} \frac{S_{thi}}{3} \quad (4.19)
\]

where

\(f_{th}\) is the \(th\) th stratum constant

and

\(S_{thi}\) is the number of segments in the sampling area.
These operations are done in the central office.

b) The interviewer is instructed to list the segment and simultaneously interview this $q^{th}$ hh and every $\frac{1}{s^{th}}$ hh thereafter. To facilitate this field instruction, the interviewer is provided with a provisional form containing a column of serial numbers of hhs in which is indicated the particular hhs to enumerate. An illustration of the use of the provisional form for $f_{th} = 0.10$ and various values of $S_{thi}$ is shown in Table 6. Thus, for $S_{thi} = 3$ and $q = 8$, the numbers of hhs to be interviewed are 8, 18, 28, 38. For $S_{thi} = 7$ and $q = 3$, the interviewed hhs are 3, 8, 13, 18, 23, 28, 33, 38, 43. In case of non-interview of indicated sample hhs, the next hh to the one indicated is interviewed.

The following steps are followed in the matched segments:

a) To maintain the required sampling interval, adjustments are made in the central office on the list containing old hhs, i.e., hhs in for the ($\alpha-1$) and/or ($\alpha-2$) visits. A provisional list is then prepared. The list contains the sample hhs to be interviewed at the $\alpha$ visit.

b) Instruct the interviewer to list sample segment and interview indicated hhs in provisional list of old hhs. In case of non-interview of indicated sample
Table 6. An illustration of the use of a provisional form which contains the indicated sample households to be interviewed for $f_{th} = 0.10$ and various values of $S_{thi}$

<table>
<thead>
<tr>
<th>$f_{th}$</th>
<th>0.10</th>
<th>0.10</th>
<th>0.10</th>
<th>0.10</th>
<th>0.10</th>
<th>0.10</th>
<th>0.10</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_{thi}$</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>$S_{th}$</td>
<td>0.10</td>
<td>0.133</td>
<td>0.167</td>
<td>0.200</td>
<td>0.233</td>
<td>0.267</td>
<td>0.300</td>
<td>0.333</td>
</tr>
<tr>
<td>$1/g_{th}$</td>
<td>10</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$q$</td>
<td>8</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

**Serial no. of hh in segment**

1. X
2. X
3. X
4. X
5. X
6. X
7. X
8. X
9. X
10. X
11. X
12. X
13. X
14. X
15. X
16. X
17. X
18. X
19. X
20. X

a$X$ indicates hh to be interviewed.
Table 6. (Continued)

<table>
<thead>
<tr>
<th>f^th</th>
<th>0.10</th>
<th>0.10</th>
<th>0.10</th>
<th>0.10</th>
<th>0.10</th>
<th>0.10</th>
<th>0.10</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>S^thi</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>g^thi</td>
<td>0.100</td>
<td>0.133</td>
<td>0.167</td>
<td>0.200</td>
<td>0.233</td>
<td>0.267</td>
<td>0.300</td>
<td>0.333</td>
</tr>
<tr>
<td>1/g^thi</td>
<td>10</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>q</td>
<td>8</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

| 21    | X  |
| 22    |    | X  |
| 23    |    | X  | X  | X  |
| 24    |    |    |    |    | X  |
| 25    |    |    |    |    |    |    |    | X  |
| 26    |    |    |    |    |    |    |    |    | X  |
| 27    |    |    |    |    |    |    |    |    |    | X  |
| 28    |    |    |    |    |    |    |    |    |    |    | X  |
| 29    |    |    |    |    |    |    |    |    |    |    |    | X  |
| 30    |    |    |    |    |    |    |    |    |    |    |    |    | X  |
| 31    |    |    |    |    |    |    |    |    |    |    |    |    |    | X  |
| 32    |    |    |    |    |    |    |    |    |    |    |    |    |    |    | X  |
| 33    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    | X  |
| 34    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    | X  |
| 35    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    | X  |

hhs, the next hh to the one indicated is interviewed. By and large, the indicated sample hhs for the \( \alpha \) visit are the same sample hhs at the \( (\alpha-1) \) visit. For the new hhs (in first time), a provisional form similar to that for a new segment is used.

For new and matched segments, the interviewer is asked to prepare up-to-date maps of sample segments, indicating
therein the approximate location of listed and sample hhs.

A new aspect of this field operation is that listing and interviewing are done in one canvassing operation of the sampling area or segment. This is an improvement over the original procedure where two visits are made in the sampling area; one for the listing operation and one for the actual interviewing. In addition, the time gap between listing and interviewing has been shortened. This new feature in the field operation should result in a reduction in cost not only of the actual listing operation but in the cost of other operations as well.

If a sampling area is not dissected, then the list of hhs may be used to divide the area into three rotation 'segments' at the initial visit \(\alpha=0\). Rotation of segment is accomplished by rotation of hhs within 'segments'. This is accomplished by taking a new random start in the unmatched or new 'segment' at the \(\alpha\) visit. Steps given in a) and b) above are followed for the new 'segment' and for the matched 'segments', respectively. For the undissected sampling areas, the main purpose of rotation is to minimize response resistance which may arise.

The same field procedure is followed in the rural area. For the rural area, \(g_{thij}, i_{thi}\) and \(S_{thij}\) are substituted for \(s_{thi}, f_{th}\) and \(S_{thi}\), respectively.

Essentially, we have extended the theory of Chapter III
to a four-stage sampling design* in the rural area and to a three-stage design* in the urban area, utilizing the important relations $R_t (4.15, 4.13), f_{thi} (4.6)$ and $f_{th} (4.14)$.

With segmentation a general characteristic, observed in the rural area $hh$, is symbolized by $\alpha_{xthijk} (k = 1, 2, \ldots, S_{thij}; l = 1, 2, \ldots, N_{thijk})$ for the $\alpha$ visit. The $i$th PSU unbiased estimate of the stratum total, $\alpha_{Xth}$, at the $\alpha$ visit is

$$\alpha_{Xthi} = \frac{P_{thi}}{p_{thi}} \sum_j \sum_k \sum_l \alpha_{Nthijk} \sum_l \alpha_{xthijk}^{*}$$

(4.20)

and the mean of $\alpha_{Xthi}$ for $i = 1, 2, 3, 4, 5$ is

$$\alpha_{Xth} = \frac{1}{5} \sum_i \frac{P_{thi}}{p_{thi}} \sum_j \sum_k \sum_l \alpha_{Nthijk} \sum_l \alpha_{xthijk}^{*}$$

(4.21)

where the symbols are as defined in (4.20) and (4.21). Equations (4.20) and (4.21) are unbiased estimates of $\alpha_{Xth}$.

Since the srws is equal for all $\alpha$, then

$$\frac{\alpha_{Nthijk}}{\alpha_{nthijk}} = \frac{1}{S_{thij}}$$

(4.22)

Substitution of (4.22) in (4.21) causes (4.21) to become

*For a fixed visit, say the $\alpha$th visit, the set of $S$ segments in the sample are a random selection of all possible segments since the original numbering was random. There is of course a systematic rotation listing the sampled segments from visit to visit.
\[ \hat{X}_{th} = \frac{1}{5} \sum \frac{P_{th}}{P_{thi}} \frac{B_{thi}}{b_{thi}} \sum' \frac{S_{thij}}{s_{thij}} \sum' \frac{1}{g_{thij}} \sum' \sum' x_{thijkl} \] 

\[ = \frac{1}{5} \sum' \frac{P_{th}}{P_{thi}} \frac{B_{thi}}{b_{thi}} \sum' \frac{S_{thij}}{s_{thij}} \sum' \frac{1}{g_{thij}} \sum' \sum' x_{thijkl} \] 

(4.23)

But from (4.10),

\[ \frac{1}{f_{thi}} = \frac{S_{thij}}{s_{thij}} g_{thij} \] 

(4.24)

and (4.23) becomes,

\[ \hat{X}_{th} = \frac{1}{5} \sum' \frac{P_{th}}{P_{thi}} \frac{B_{thi}}{b_{thi}} \frac{1}{f_{thi}} \sum' \sum' \sum' x_{thijkl} \] 

(4.25)

which is the same form as (3.25) and is equal to

\[ \hat{X}_{th} = R_t \hat{X}_{th} \] 

(4.26)

where

\[ R_t \] is as defined in (3.27),

and

\[ \hat{X}_{th} = \sum' \sum' \sum' \sum' x_{thijkl} \] is the sample stratum total at the \( \alpha \) visit.

The estimate of the variance of \( \hat{X}_{th} \) is given by

\[ s^2_{th} = \frac{m}{m-1} \frac{R_t^2}{\sum' \left( X_{thi} - \frac{\hat{X}_{th}}{m} \right)^2} \] 

(4.27)

which is similar in form to (3.52).
Following the development of the estimation procedure in Chapter III, we estimate, $\hat{X}_t$, the sector total by

$$\hat{X}_t = R_t \cdot \alpha_t$$

(4.28)

where

$$\alpha_t = \sum \sum' \sum' \sum' \sum' x_{ti}$$

is the sample sector total at the $i$th visit.

The estimated 'survey' variance of $\hat{X}_t$ is

$$\hat{\sigma}^2_t = \frac{m \cdot \hat{R}^2_t}{(m - 1)} \left\{ \sum' \alpha^2_{ti} - \frac{\alpha^2_t}{m} \right\}$$

(4.29)

where

$$\alpha_{ti} = \sum \sum' \sum' \sum' \sum' x_{thijk}$$

is the sample sector total from the $i$th psu at the $i$th visit.

For the urban area ($t = 3, 4$), the estimate of the stratum total, $\alpha_{th}$, from the $i$th psu at the $i$th visit is,

$$\alpha_{thi} = M_{th} \cdot \sum' s_{thi} \cdot \sum' n_{thi} \cdot \alpha_{thijk}$$

(4.30)

and the estimate from the five numbered psus is,

$$\alpha_{th} = \frac{1}{m} \sum_{i} M_{th} \cdot \sum' s_{thi} \cdot \sum' n_{thi} \cdot \alpha_{thijk}$$

(4.31)

With (4.17) in (4.31), we get

$$\alpha_{th} = \frac{1}{m} \sum_{i} M_{th} \cdot \frac{1}{f_{th}} \sum' \sum' \alpha_{thijk}$$

(4.32)
As in the rural area development, the imposition of restrictions given in (4.13) and (4.14), (4.32) becomes

\[ \hat{X}_{th} = R_t \times_{th} \]  

(4.33)

where

\[ R_t \] is the raising factor for the urban area \( t = 3, 4 \), and

\[ \times_{th} = \sum' \sum' \sum' \times_{thijk} \] is the sample stratum total

for the urban sectors at the \( t \) visit.

The estimated variance is similar in form to (4.27).

To estimate \( \hat{X}_t \), the sector total, we use

\[ \hat{X}_t = R_t \times_t \]  

(4.34)

where

\[ \times_t = \sum \sum' \sum' \sum' \times_{thijk} \] is the sample sector total for the urban sectors at the \( t \) visit,

and its estimated variance is of the same form as (4.29) where

\[ \times_{ti} = \sum \sum' \sum' \times_{thijk} \] is the sample sector total

from the \( i \)th ordered psu of the urban sectors at the \( t \) visit.

Combining the estimates given in (4.28) and (4.34), we get an estimate of the country total, \( \hat{X} \),

\[ \hat{X} = \sum_{t} \hat{X}_t \] for \( t = 1, 2, 3, 4 \)  

(4.35)
and its estimated variance is given by

\[ \hat{\sigma}^2 = \frac{1}{m(m - 1)} \sum_{i}^{'} (\hat{x}_i - \hat{x})^2 \]

where

\[ \hat{x}_i = \sum_{t}^{h} \sum_{h}^{'} \hat{x}_{thi} \]

and

\[ \hat{x}_{thi} = m R_t \sum_{j}^{'} \sum_{k}^{'} x_{thijk} \text{ for } t = 1, 2: \]

or

\[ \hat{x}_{thi} = m R_t \sum_{j}^{'} \sum_{k}^{'} x_{thijk} \text{ for } t = 3, 4. \]

Estimates given in equations (4.26), (4.27), (4.28), (4.29), (4.33), (4.35) and (4.36) have been shown to be unbiased, since they are the same estimators given in Chapter III. The difference is merely in the addition of a new stage unit due to the segmentation procedure inherent to the new rotation scheme. A comparison of the 'survey' variance and the 'stratified' or standard variance is given in Chapter III.

2. Unequal sampling rate

A serious attempt will be made to equalize the size of segments in the sa. This is found to be feasible in the urban areas and in the poblacion sector but not in the barrio sector. With approximately equal sized segments and with
equal srws, the sample take within each segment will be approximately equal. The rotation of segments will have little effect on the variation of estimates from visit to visit.

With unequal sized segments and with equal srws, the self-weighting, unbiased estimator may become very variable from visit to visit. This may be so, since in the new rotation pattern, small sized segments may replace large sized segments and vice versa. The barrio sample take or the overall psu take will therefore vary considerably from visit to visit. With varying take within psu,

\[ \alpha_{thi}^* = m^R_t \alpha_{thi} \]

will also vary from visit to visit. Also, under this situation, there would exist considerable variation in work loads within segments of the sample barrio.

One may attempt to use unequal srws but equal sample take within segments in the sample barrio. From (4.9), we replace \( S_{thi} \) by \( \frac{N_{thi}}{N_{thijk}} \) and we have

\[ \alpha_{thijk} = \frac{1}{5} \frac{p_t}{p_{thi}} \frac{b_{thi}}{2} \cdot \frac{\alpha_{N_{thi}}}{3} \frac{1}{\alpha_{N_{thijk}} R_t} \]

(4.37)

\[ = \frac{f_{thi}}{3} \frac{\alpha_{N_{thi}}}{\alpha_{N_{thijk}}} \]

The srws

\[ \alpha_{thijk} \]
will now depend on \( \alpha \), and is inversely proportional to the size of the segment. Equation (4.20) may be written as follows:

\[
\alpha^*_{thi} = \frac{p_{thi}}{p_{thi}} \frac{b_{thi}}{b_{thi}} \sum_j \sum' k \frac{s_{thi}}{3} \frac{N_{thijk}}{n_{thijk}} \sum'_\ell x_{thijk}\ell.
\]

(4.38a)

Substitute \( \frac{S_{thijk}}{N_{thijk}} \) for \( S_{thijk} \) in (4.38a), we get

\[
\alpha^*_{thi} = \frac{p_{thi}}{p_{thi}} \frac{b_{thi}}{b_{thi}} \sum_j \sum' k \frac{N_{thijk}}{3} \frac{1}{n_{thijk}} \sum'_\ell x_{thijk}\ell.
\]

(4.38b)

From (4.37), we have

\[
N_{thijk} = \frac{f_{thi}}{3} \frac{n_{thijk}}{N_{thij}}
\]

(4.39)

which implies that \( N_{thijk} \) in equation (4.38b) is now a barrio constant and does not depend on \( k \). With (4.39) in (4.38b), we get

\[
\alpha^*_{thi} = \frac{p_{thi}}{p_{thi}} \frac{b_{thi}}{b_{thi}} \sum_j \sum' k \frac{N_{thijk}}{3} \frac{f_{thi}}{n_{thijk}} \frac{3}{N_{thij}} \sum'_\ell x_{thijk}\ell
\]

\[
= \frac{p_{thi}}{p_{thi}} \frac{b_{thi}}{b_{thi}} \frac{1}{f_{thi}} \sum_j \sum' k \sum'_\ell x_{thijk}\ell
\]

(4.40)

\[
= m R_t \alpha^*_{thi}
\]

which is self-weighting but biased.
The within barrio bias is

\[
B = E \left\{ \frac{S_{thij}}{3} \sum_k \frac{N_{thijk}}{n_{thijk}} \sum' x_{thijk} \right\} - \frac{N_{thij}}{3} \sum_k \frac{1}{N_{thijk}} \frac{N_{thijk}}{n_{thijk}} \sum' x_{thijk} \right\} \tag{4.41}
\]

\[
= x_{thij} - \frac{N_{thij}}{S_{thij}} \sum_k \frac{x_{thilk}}{n_{thijk}}
\]

where

\[
E \left\{ \frac{S_{thij}}{3} \sum_k \frac{N_{thijk}}{n_{thijk}} \sum' x_{thijk} \right\} = x_{thij} \tag{4.42a}
\]

is the thijth barrio total at visit \( \alpha \),

and

\[
E \left\{ \frac{N_{thijk}}{3} \sum_k \frac{1}{n_{thijk}} \sum' x_{thijk} \right\} = \frac{N_{thij}}{S_{thij}} \sum_k \frac{x_{thijk}}{n_{thijk}}. \tag{4.42b}
\]

Equation (4.42b) is, in general, not equal to \( x_{thij} \), unless the \( N_{thijk} \)'s are all equal such that

\[
\frac{N_{thijk}}{n_{thijk}} = S_{thij},
\]

or the ratios
are equal for all $k$.

If the segments are drawn with pps, then the bias in (4.41) disappears. With the new rotation scheme, however, the drawing of segments with pps is not theoretically feasible. For a fixed visit $a$, the set of three segments in the sample is assumed to be a random selection of all possible segments in a given barrio.

Growth in the counts of hhs in the sa or barrio may be dispersed or concentrated [19]. Dispersed growth is one in which the construction or birth of new hhs and the disappearance or drop-out of old ones, occur at a relatively even rate in most of the segments in the barrio. This implies that the rate of growth of each segment, $r'$, is equal to the rate of growth of the entire barrio, $r$, from the ($a-1$) visit to the $a$ visit. Thus for

$$r' = r,$$

$$\frac{f_{thi} (1+r)}{3} \frac{\alpha^{-1} N_{thi}}{\alpha^{-1} N_{thijk}}$$

$$= \frac{f_{thi}}{3} \frac{\alpha^{-1} N_{thi}}{\alpha^{-1} N_{thijk}}$$

(4.40)

In concentrated growth, the rate of growth in a relatively
small number of segments is greater than the rate of growth of the entire sa. Thus

\[ r' > r \]

for some segments and

\[ \alpha_{\text{thijk}} < \alpha_{-1}\text{thijk} \].

If

\[ r' < r \],

then

\[ \alpha_{\text{thijk}} > \alpha_{-1}\text{thijk} \].

In any case, adjustments are made in the central office and the field procedure is still simplified since the relationships which have to be maintained are,

\[ \alpha_{n\text{thij}} = f_{\text{thi}} \alpha_{N\text{thij}} \]

is the sample take in the \( \text{thij} \)th barrio at the \( \alpha \) visit,

and

\[ \alpha_{n\text{thijk}} = \frac{\alpha_{n\text{thij}}}{3} \]

\[ = \alpha_{\text{thijk}} \alpha_{N\text{thijk}} \]

is the sample take in \( k \)th segment of the \( \text{thij} \)th barrio at the \( \alpha \) visit,

where

\[ \alpha_{\text{thijk}} \]

is as defined in (4.37).

For the new segment the field procedure at the \( \alpha \) visit
with unequal srws is as follows:

At the central office, a random start, \( q \), is drawn in the interval \( 1 \) to \( \frac{1}{\alpha_{\text{thijk}}} \). This \( q^{\text{th}} \) hh and every \( \frac{1}{\alpha_{\text{thijk}}} \)-th hh thereafter is to be interviewed. The value of \( \alpha_{\text{thijk}} \) is obtained from (4.37). A provisional form similar to the one described in Table 4 is given to the interviewer who lists the \( k^{\text{th}} \) sample segment and simultaneously interviews the indicated hhs in the prescribed form. Additional instructions are given in case of non-interview. If an indicated sample hh is not interviewed, the cause or causes will be given, and the next hh to the one indicated is interviewed instead. Refusals will be considered as a different category. This field procedure is similar to the one instituted for a new segment with equal srws.

Two aspects are considered in the field procedure for the matched segments with unequal srws. These are:

a) dispersed growth in sampling area. In this case,

\[ \alpha_{\text{thijk}} = \alpha_{-1} \alpha_{\text{thijk}} \]

and the same srws is used for the \( \alpha \) visit as in the previous visit (\( \alpha_{-1} \)). The interviewer is given a list of hhs from the (\( \alpha_{-1} \)) visit wherein are indicated the sample hhs to interview. This list has been adjusted in the central office for sampling interval; to take care of drop-outs, non-interviews
and other related causes. In addition, the interviewer is provided with a provisional form (Table 6) which will be used for new hhs (in for first time). The procedure for new hhs will be similar to hhs in a new segment. Listing and interviewing are done in only one operation. Any adjustments are made in the central office.

b) concentrated growth in sampling area. In this case,

\[ g_{thijk} < g_{thijk} - l \]

which implies that the srws at the \((< - l)\) visit is greater than srws at the \(\alpha\) visit. The interviewer is provided with a list of hhs from the \((< - l)\) visit wherein are indicated the list of sample hhs to interview. By and large, these sample hhs are the same as in the \((\alpha - l)\) visit but the list has been adjusted for sampling interval. An adjusted sampling rate is applied to the list of new hhs such that

\[
\text{adjusted rate} \times \text{number of new hhs at } \alpha \text{ visit} = \alpha^d
\]

is the size of sample accruing from the list of new hhs,

and

\[
g_{thijk} = n_{thijk} = n_{thijk} - l = n_{thijk} - l + n_d
\]
where

\( \alpha^{-1} \text{thijk} \) is obtained from the \((\alpha - 1)\) visit,

\( \alpha^{n} \text{thijk} \) and \( \alpha^{n} \text{nd} \) are obtained in advance from (4.37).

In case,

\( \alpha^{E} \text{thijk} > \alpha^{-1} \text{thijk} \),

then similar adjustment procedures as outlined above for

\( \alpha^{E} \text{thijk} < \alpha^{-1} \text{thijk} \)

may be used. Note that adjustments are made only in the second and third visit for any given segment, since the segment is dropped from the sample after the third visit.

Listing of the new segment and the matched segments at the \( \alpha \) visit to record births and deaths of hhs will adequately up-date the sample for the dispersed and concentrated growths. In general, only new hhs which are included by application of the appropriate or adjusted sampling interval within segments will be added to a particular subsample of hhs from the previous visit \((\alpha - 1)\). Note that \( \alpha^{N} \text{thij} \) and \( \alpha^{N} \text{thijk} \) are available in advance and an accurate knowledge of the srws

\( \alpha^{E} \text{thijk} \)

is available prior to the actual field operation. Under this
situation any adjustments are easily made in the central office. Also from the relation

\[ n_{thij} = f_{thi} \cdot N_{thij}, \]

we note that

\[ f_{thi} \] is a psu constant

and

\[ N_{thij} \] is the listed number of hhs in the thijth barrio.

In general, \[ N_{thij} \] increases with \[ \alpha \] and therefore, \[ n_{thij} \] and \[ n_{thijk} \] will also increase with \[ \alpha \], since

\[ n_{thijk} = \frac{n_{thij}}{3}. \]

In summary, the estimation procedures with equal srws are self-weighting and unbiased and the field procedure is simple to implement. On the other hand, the estimation procedures with unequal srws and equal sample take within segments are biased. The field operation although still simple is more difficult to implement than with equal srws. There is however some control on the variation of estimates from visit to visit because of equal sample take within segments. Also with equal take, the work load is equalized within segments. The extended use of unequal srws will be dictated by the actual experience gained in the field.
V. RATIO ESTIMATION

This chapter will consider an attempt on the control of the variation in the size of the barrio and its effect on the precision of estimates of regional and sector totals for the barrio and on the precision of estimates of the rural area total and country total.

In addition, ratio estimation will be introduced with segmentation into the estimation procedures in the urban area. Since sampling is multi-stage with complete replacement of the psu sampled at each stage with equal probability, there is a loss in precision represented by not applying the finite population correction (fpc),

\(1 - \frac{m}{M_{th}}\),

which is between 0.93 to 0.95 for Manila, if sampling is without replacement of psu. For the capitals and cities, \(\frac{m}{M_{th}} < 0.05\), and so the fpc is negligible. In this case, a larger loss in precision arises through sampling the psu with equal probability rather than with probability proportional to some measure of size. The reason for the latter procedure is that at the initiation of the design for the FSSH urban area, no measure of size of the precinct (psu) was available.

In view of the above mentioned reasons for unavoidable losses in precision it appeared desirable to examine methods of estimation which would improve the precision. To increase
the precision of estimates for the urban area, ratio estimation will be introduced into the design. These techniques will be extended to the rural area, if found feasible and applicable.

A. Control of the Variation in the Size of Barrio

This section will consider an initial aspect in the development of the sample design of the PSSH, prior to the rotation scheme developed in Chapter IV. Specifically, this development deals with an attempt to redefine the barrio stage unit in the barrio sector through control of the variation in size, $N_{thij}$, the number of listed hhs of the $thij^{th}$ barrio in the $i^{th}$ municipality of the $h^{th}$ stratum in the barrio sector ($t = 1$). The equalization of size of barrio was realized through the utilization of an ancillary variable, $Z_{thij}$, the number of population of the $thij^{th}$ barrio.

It should be recalled at this point that the sizes of the strata were made equal through equalization of size of population of stratum. In addition, the psus were drawn with pps, the measure of size is again size of population of psu.

One of the most important interests in the design of the PSSH is the estimation of totals such as the stratum total, $X_{th}$; the sector total, $X_t$; and the country total, $X$, of many characteristics from the sample. Because of this interest in the estimation of totals or population aggregates, the control
of variation in the size of the ultimate cluster [9, Vol. 2, p. 204], for example, the barrio, assumes a very important role in the efficiency of the design. The barrio is the ultimate cluster in the barrio sector of the rural area.

Let us consider the role of the size of the barrio in the precision of estimates from the barrio sector (t = 1) prior to segmentation. From (3.29), the estimate of the stratum total, $X_{th}$, from the $i^{th}$ PSU is

$$\hat{X}_{lhi} = m R_l x_{lhi}$$

where

$m = 5$

$R_l = 850$

and

$x_{lhi}$ is the sample total of the $i^{th}$ PSU.

Note that $x_{lhi}$ is the only variable in the estimation procedure. Any control on the variation of the $x_{lhi}$ will be reflected on the reduction of the variance of $\hat{X}_{lhi}$ or $\sigma^2_{lhi}$.

This sample total depends on the size of sample from each of the barrios. In fact, from (3.28)

$$n_{lhi} = f_{lhi} \sum_j N_{lhij}$$

is the sample size from the $i^{th}$ PSU and depends on the PSU constant, $f_{lhi}$, and on the relative sizes, $N_{lhij}$, of the two sample barrios. One aspect of control in the size of barrio, $N_{lhij}$, is to equalize their sizes. This equalization will have a corresponding equalization effect on the $n_{lhi}$'s and
consequently on the $x_{ih}$'s.

Extreme variations existed in the $n_{ih}$'s (column 7, Table 7), for some selected strata. The cv of the estimate of population in hhs in the barrio sector was estimated at 6.76 per cent in May, 1956 (column 5, Table 3) and at 6.77 per cent for October, 1956 (column 8, Table 3). Most of the estimates for the region for these visits have cvs greater than 10 per cent.

Control in the variation of $n_{ih}$'s was attempted through control in size of population in barrio, $Z_{ih}$, which in turn is highly correlated with $N_{ih}$, the number of listed hhs in barrio. Barrios with small population contained very few sample hhs while barrios with large population contained a large number of sample hhs. An attempt to equalize $n_{ih}$'s was made through equalization in size of population, $Z_{ih}$'s.

To improve the barrio sample and to bring the barrio sector estimates, particularly those for the region to a higher level of precision, control in the variation in size of barrio, $Z_{ih}$, was carried out prior to the third round of the PSSH, March, 1957. Steps taken toward this development were:

a) The size of population, $Z_{ih}$, of each of the $B_{ih}$ barrios in the $i$th sample psu prior to March, 1957 was obtained from the municipal mayors and the barrio lieutenants. Within the $i$th sample municipality
(psu), the barrios (ssus) were redefined and made into units of approximately equal size of population. The first process consisted of combining of small barrios, with the requirement that the new unit should have

\[ Z_{lhij} \geq 450. \]  
(5.1)

The second process was to split extremely large barrios into two or more new units with the same restriction as (5.1).

The new unit thus formed is either a group of barrios, a whole barrio, or a part of a barrio. Rigorous equalization was attempted in regions with high cv's.

b) Original sample barrios which were untouched remained in the sample. If two original sample barrios were combined, another 'new' unit was selected at random; and in case of dissection of a large sample barrio into two or more 'new' units, then a 'new' unit is drawn at random to replace the original big barrio.

c) A new psu constant, \( f'_{lhi} \) (3.28), was computed for the \( i^{th} \) sample psu. This was brought about by the formation of a new barrio count, \( B'_{lhi} \), in the \( i^{th} \) sample psu. Thus

\[ f'_{lhi} = \frac{1}{m} \frac{P_{lhi}}{P_{lhi}} \frac{B'_{lhi}}{b'_{lhi}} \frac{1}{R_l} \]
where \( m, \beta_{lh}, \beta_{lhi}, b_{lhi} \) and \( R_l \) are as defined in (3.23).

d) Efforts were made to determine the true boundaries of the 'new' sample areas. Interviewers were required to draw sketch maps of the sample areas assigned to them, indicating areas actually covered, 'natural' boundaries, location of hhs, and other pertinent facts about the 'new' sample area. Each of these maps were carefully screened and checked for accuracy. Completely new listing of hhs was made in each 'new' sample area.

An illustration of the attempt on the equalization of population size of barrio through the combining of small barrios is given in Table 7.

The changes in the number of barrios, \( B_{lhi} \); the psu constant, \( f_{lhi} \); and the size of sample within psu, \( n_{lhi} \), before and after equalization of size of barrio, are illustrated in Table 8 for five selected strata.

There is a general indication of a decrease in variation, as measured by the range, \( R \), for \( f_{lhi} \) and \( n_{lhi} \). Some general results of this attempt on equalization are indicated by the cv's of various estimates (Table 9).

The cv of \( \hat{X}_1 \) = 14.286 millions of population for March, 1957 was reduced to 2.31 per cent as compared to a cv of 6.77 per cent for \( \hat{X}_1 = 14.089 \) millions for the previous round,
Table 7. An illustration of the attempt on the equalization of population size of barrio through the combining of small barrios for two selected sample municipalities

<table>
<thead>
<tr>
<th>Stratum V</th>
<th>Stratum X</th>
</tr>
</thead>
<tbody>
<tr>
<td>4th sample municipality</td>
<td>5th sample municipality</td>
</tr>
<tr>
<td><strong>Before equalization</strong></td>
<td><strong>Before equalization</strong></td>
</tr>
<tr>
<td>1</td>
<td>476</td>
</tr>
<tr>
<td>2</td>
<td>325</td>
</tr>
<tr>
<td>3</td>
<td>250</td>
</tr>
<tr>
<td>4</td>
<td>600</td>
</tr>
<tr>
<td>5</td>
<td>800</td>
</tr>
<tr>
<td>6</td>
<td>1,080</td>
</tr>
<tr>
<td>7</td>
<td>185</td>
</tr>
<tr>
<td>8</td>
<td>900</td>
</tr>
<tr>
<td>9</td>
<td>680</td>
</tr>
<tr>
<td>10</td>
<td>727</td>
</tr>
<tr>
<td>11</td>
<td>1,230</td>
</tr>
<tr>
<td>12</td>
<td>650</td>
</tr>
<tr>
<td>13</td>
<td>540</td>
</tr>
<tr>
<td>14</td>
<td>1,350</td>
</tr>
<tr>
<td>15</td>
<td>680</td>
</tr>
<tr>
<td>16</td>
<td>520</td>
</tr>
<tr>
<td>17</td>
<td>976</td>
</tr>
<tr>
<td>18</td>
<td>760</td>
</tr>
<tr>
<td>19</td>
<td>3,050</td>
</tr>
</tbody>
</table>

**Source:** Office of Statistical Coordination and Standards, National Economic Council, Manila, Philippines, 1957.

October, 1956. Similarly, the cv of $\hat{X}$, the country estimate, was decreased from 3.27 per cent in October, 1956 to 2.16 per cent in March, 1957. Most significant were the results on the reduction of the cv of barrio population estimates for
Table 8. Number of barrios, primary sampling unit constant, and number of sample households within primary sampling unit before and after equalization of population size of barrios, for selected strata

<table>
<thead>
<tr>
<th>Stratum (psu)</th>
<th>Municipality (i)</th>
<th>Number of barrios Before</th>
<th>Number of sample households within primary sampling unit constant Before</th>
<th>Number of sample households within primary sampling unit constant After</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>11</td>
<td>9</td>
<td>.038</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>29</td>
<td>34</td>
<td>.073</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>17</td>
<td>6</td>
<td>.098</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>19</td>
<td>17</td>
<td>.066</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>11</td>
<td>11</td>
<td>.041</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R = .060 = .054</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>17</td>
<td>8</td>
<td>.177</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>48</td>
<td>33</td>
<td>.094</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>.028</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>20</td>
<td>18</td>
<td>.053</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>20</td>
<td>16</td>
<td>.045</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R = .149 = .053</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>24</td>
<td>28</td>
<td>.043</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>7</td>
<td>5</td>
<td>.061</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>21</td>
<td>29</td>
<td>.039</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>22</td>
<td>40</td>
<td>.036</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>46</td>
<td>36</td>
<td>.108</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R = .077 = .049</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>9</td>
<td>6</td>
<td>.049</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>57</td>
<td>30</td>
<td>.147</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>67</td>
<td>44</td>
<td>.123</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>38</td>
<td>21</td>
<td>.082</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>38</td>
<td>21</td>
<td>.082</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R = .098 = .048</td>
</tr>
<tr>
<td>25</td>
<td>1</td>
<td>34</td>
<td>27</td>
<td>.096</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>27</td>
<td>26</td>
<td>.070</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>13</td>
<td>13</td>
<td>.052</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>25</td>
<td>18</td>
<td>.074</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>10</td>
<td>9</td>
<td>.070</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R = .044 = .024</td>
</tr>
</tbody>
</table>

Source: Office of Statistical Coordination and Standards, Manila, Philippines, 1957.

bR = maximum - minimum.
Table 9. Estimates of barrio population and coefficient of variation of estimates in per cent by region before equalization, October, 1956, and after equalization of size of population of barrio, March, 1957\(^a\)

| Region | October, 1956 | | March, 1957 | |
|--------|---------------|----------------|----------------|
|        | Estimate      | Coefficient    | Estimate      | Coefficient    |
|        | (millions)    | of variation   | (millions)    | of variation   |
|        |               | (per cent)     |               | (per cent)     |
| II     | 1.076         | 21.92          | 0.934         | 10.25          |
| III    | .593          | 9.92           | .767          | 6.27           |
| IV     | 2.505         | 7.73           | 2.024         | 4.74           |
| V      | 1.871         | 18.96          | 1.617         | 10.43          |
| VI     | 1.509         | 17.72          | 1.431         | 4.18           |
| VII    | 2.275         | 8.07           | 2.399         | 4.77           |
| VIII   | 2.675         | 11.34          | 2.876         | 6.23           |
| IX     | .839          | 11.86          | 1.153         | 6.53           |
| X      | .746          | 19.79          | 1.085         | 11.01          |
| Philippines | 14.089   | 6.77          | 14.286       | 2.31          |

\(^a\)Source: Office of Statistical Coordination and Standards, National Economic Council, Manila, Philippines, 1957.

regions II to X (Table 9). Most of the cv's were reduced from more than 10 per cent to about 10 per cent or less. Similar reduction in the cv of the most important characteristics of the survey was observed.

The equalization of size of population of the barrio will be a permanent feature of the barrio sector at least for a
period of time. The reduction of the cv of estimates for the region and for the sector is considerable. There is also some reduction in the cv of the estimate for the rural area and for the country. This redefinition of the barrio stage unit did not in any way affect the form of the self-weighting estimation procedures developed in Chapters III and IV. The simplicity of the self-weighting scheme has been retained in the design.

B. Ratio Estimation in the Urban Area

As mentioned before, the sample of the PSSH has been constructed and it consists, apart from minor substitutions, of the same panel of primaries, secondaries and tertiaries. Thus, in the urban area, improvement of the efficiency of the design may be accomplished through a change to a ratio estimator. As auxiliary variable, we will use the list or count of registered voter ($Z_{th1}$) which is available by precinct from the Philippine Electoral Commission. National election is held every two years; the latest was held on November 10, 1959. This auxiliary variable is up-dated every two years. Biased and unbiased ratio estimators will be discussed for possible incorporation into the estimation procedures in the urban area.
1. Biased ratio estimation

We have from the $i^{th}$ PSU of the $th$ stratum an independent unbiased estimate, $\hat{X}_{thi}$, of the stratum total, $X_{th}$. There are five such PSUs drawn sequentially and thus there are five such independent estimates, and the mean of these estimates, $\hat{X}_{th}$, is also an unbiased estimate of $X_{th}$. Each $\hat{X}_{thi}$ is considered as an independent observation from an infinite population of $\hat{X}_{thi}$ at the $\alpha$ visit. With this consideration, standard techniques on ratio estimation for simple random sampling can be utilized. Specifically, the ratio of means estimate [11] will be introduced into the estimation procedures. The regular unbiased estimate, $\hat{X}_{thi}$ or $\hat{X}_{thi}$, of the stratum total from the $i^{th}$ PSU, will take different forms depending upon the nature of the design employed for the secondary and tertiary stages.

Consider first the estimation of the stratum total, $X_{th}$, will be for population in hhs and occupation of head of hh. These characteristics are obtained in the first phase or listing operation of the segments in a rotation group or of the whole precinct if there is no segmentation. Three forms of ratio estimation will be discussed; one of which is applicable to any given stratum.

a) complete ratio estimation. The segments in precincts are known, recognizable and identical for both electoral list and survey list. Street names or city
blocks may be used to identify segments in precincts with electoral list for the precinct. Bias should be eliminated in the listing within segments.

Consider as estimate of $X_{th}$,

$$X_{th}^{**} = Z_{th}$$

(5.2)

where

$$X_{th}^{**} = \frac{1}{m} \sum_{i}^{m} X_{th} S_{th} \frac{1}{3} \sum_{j}^{n} X_{thij}$$

$$Z_{th} = \frac{1}{m} \sum_{i}^{m} X_{th} S_{th} \frac{1}{3} \sum_{j}^{n} Z_{thij}$$

(5.3)

$X_{thij}$ is the listed segment total at the $th$ visit

$Z_{thij}$ is the corresponding electoral list total for the segment at the $th$ visit (closest electoral year)

and

$Z_{th}$ is the electoral list total for the $th$ stratum at the $th$ visit.

Equation (5.3) is too cumbersome to compute. With the use of self weighting scheme given in (4.31), (4.32) and (4.33), (5.2) becomes,

$$X_{th}^{**} = Z_{th}$$

(5.4)

where

$$X_{th}^{**} = \sum_{i}^{m} \sum_{j}^{n} X_{thij}^{**}$$

(5.5)
Thus, the numerator in (5.3) is equal to
\[
\alpha^{x**}_{thj} = \gamma_{thj} \alpha^{x}_{thj} \tag{5.6}
\]
and
\[
\gamma_{thj} = \frac{1}{m} M_{th} \frac{S_{thi}}{3} \frac{1}{R_t} . \tag{5.7}
\]

The same result is obtained for the denominator. The \( R_t \) cancels and we have the result given in equation (5.4).

The bias in \( \alpha^{x**}_{th} \) is obtained from the relation,
\[
E \{ \hat{\alpha}^{x**}_{th} \} = \alpha^{x}_{th} - \text{Cov}( \hat{\alpha}^{q}_{th}, \hat{\alpha}^{z**}_{th} ) \tag{5.9}
\]
and the bias is therefore equal to
\[
- \text{Cov}( \hat{\alpha}^{q}_{th}, \hat{\alpha}^{z**}_{th} ) \tag{5.10}
\]

where
\[
\hat{\alpha}^{q}_{th} = \frac{\alpha^{x**}_{th}}{\alpha^{z}_{th}} \tag{5.11}
\]
and
\[
\hat{\alpha}^{z**}_{th} \text{ or } \hat{\alpha}^{z**}_{th} \text{ is as defined in (5.8).}
\]

The upper limit of this bias is
where

\[ \text{CV}(\widehat{\beta}_{\text{th1}}) = \frac{\text{CV}(\widehat{\beta})}{\text{CV}(\beta)} \]

\( \beta_{\text{th1}} = m R \beta_{\text{th}} \)

\[ \sigma^2 X_{\text{th}} = \frac{1}{m} \left\{ \sigma^2 X_{\text{th}} + \sigma^2 \beta_{\text{th1}} \beta_{\text{th}} \right\} \]

or

\[ \sigma^2 X_{\text{th}} = \frac{1}{m} \sigma^2 X_{\text{th}} - \sigma^2 \beta_{\text{th1}} \beta_{\text{th}} \]

and

\[ \beta_{\text{th1}} = \frac{X_{\text{th}}}{\beta_{\text{th}}} \]

An approximate estimate of the approximate variance formula given in equation (5.14a) is

\[ \hat{s}^2_{\beta_{\text{th}}} = \frac{1}{m} \frac{1}{m-1} \left\{ \sum_{i=1}^{m} \hat{\beta}_{\text{th1}}^2 + \hat{\beta}_{\text{th}}^2 \sum_{i=1}^{m} \hat{\beta}_{\text{th1}}^2 \right\} - 2 \hat{\beta}_{\text{th1}} \sum_{i=1}^{m} \hat{\beta}_{\text{th1}} \hat{\beta}_{\text{th1}} \]

which simplifies for computational purposes into
\[
\hat{x}_{th} = \frac{m \hat{R}_t}{m - 1} \left\{ \sum_i \hat{x}_{thi}^2 + \frac{m}{m - 1} \sum_i \hat{x}_{thi}^2 \sum_i \hat{z}_{thi}^2 \right. \\
- \left. 2 \alpha_{thi} \sum_i \hat{x}_{thi} \hat{z}_{thi} \right\} 
\]

(5.16)

where

\( \hat{x}_{thi} \) and \( \hat{z}_{thi} \) are as defined in (5.13).

A gain in reliability of \( \hat{x}_{th} \) to that of \( \hat{x}_{th} \), the x-only estimate, is achieved if

\[
\sqrt{\frac{\hat{x}_{thi}}{\hat{x}_{thi}}} > \frac{1}{2} \frac{CV(\hat{z}_{thi})}{CV(\hat{x}_{thi})} 
\]

(5.17)

where \( \overline{\hat{x}}_{thi} \) and \( \overline{\hat{z}}_{thi} \) is the correlation between \( \hat{x}_{thi} \) and \( \hat{z}_{thi} \). \( CV(\hat{z}_{thi}) \) and \( CV(\hat{x}_{thi}) \) are as defined in (5.10). Relation (5.17) is derived by making the last two terms in (5.14a) less than zero, i.e.,

\[
\alpha_{thi}^2 \sigma_{thi}^2 \hat{x}_{thi}^2 - 2 \alpha_{thi} \sigma_{thi}^2 \hat{x}_{thi} \hat{z}_{thi} < 0
\]

where

\[
\alpha_{thi}^2 = \frac{\hat{x}_{thi}}{\hat{z}_{thi}}
\]

b) ratio estimation in the primary stage. In this case, the segments in precinct are known and recognizable for the survey only. Available are segment totals, \( \alpha_{x_{thij}} (j = 1, 2, 3) \) and the precinct (psu) totals, \( \alpha_{Z_{thi}} (i = 1, 2, \ldots, M_{th}) \). The ratio estimator has the form
\[ \hat{X}_{th} = \frac{\sum_{i=1}^{m} \sum_{j=1}^{3} \alpha_{thij} X_{th}}{m \sum_{i=1}^{3} \alpha_{thi}} \quad (5.18a) \]

which simplifies into

\[ \hat{X}_{th}^* = \frac{\sum_{i=1}^{3} \alpha_{thi} X_{th}^i}{\sum_{i=1}^{3} \alpha_{thi}} \quad (5.18b) \]

where

\[ \alpha_{thi} \text{ is an unbiased estimate of the } i \text{th PSU total, } \alpha_{thi}. \]

The ratio in (5.18a) or (5.18b) may be expressed in terms of 'sample' totals. From equation (5.8) the numerator is equal to

\[ \hat{X}_{th}^* = R_t \alpha_{th}^* \]

and the denominator is equal to

\[ \hat{Z}_{th} = R_t \alpha_{th} \quad (5.19) \]

where

\[ \alpha_{th}^* = \sum_{i=1}^{3} \alpha_{thi}^*, \]

\[ \alpha_{thi}^* = f_{th} \alpha_{thi} \]

and

\[ f_{th} = \frac{1}{m} M_{th} \frac{1}{R_t}. \]

Thus, we may write (5.18b) as
110

\[ \hat{X}^{**} = \hat{X}_{th} \hat{Z}_{th} \quad (5.20) \]

where

\[ \hat{X}_{th} = \frac{\hat{X}^{**}}{\hat{Z}_{th}} \]

The bias in \( \hat{X}_{th} \) is

\[ - \text{Cov}(\hat{X}_{th}, \hat{Z}_{th}) \quad (5.21) \]

and the upper limit of the bias in (5.21) is

\[ |\text{Bias}| \leq \frac{1}{\sqrt{m}} \sigma_{\hat{X}^{**}} \text{CV}(\hat{Z}_{th1}) \quad (5.22) \]

where

\[ \sigma_{\hat{X}^{**}}^2 = \frac{1}{m} \left\{ \sigma_{\hat{X}^{**}}^2 + \sigma_{\hat{Z}_{th1}}^2 \sigma_{\hat{Z}_{th1}}^2 \right\} \quad (5.23) \]

\[ \hat{Z}_{th1} = \frac{1}{\sqrt{m}} \text{Rt} \hat{Z}_{th1} \quad (5.24) \]

and

\[ \text{CV}(\hat{Z}_{th1}) = \frac{\sigma_{\hat{Z}_{th1}}}{\hat{Z}_{th1}} \]

The covariance will probably be smaller than in case a) and hence the variance of the ratio estimator larger. The approximate estimate of the variance in equation (5.23) is
\[ s_{\alpha_{th}}^{2} = \frac{m \cdot R_{th}^{2}}{m - 1} \left\{ \sum_{i}^{2} x_{thi}^{*2} + \alpha_{th}^{2} \sum_{i}^{2} z_{thi}^{*2} \right\} - \frac{2}{\alpha_{th}} \sum_{i}^{2} x_{thi}^{*} z_{thi}^{*} \]  

(5.25a)

where

\( \alpha_{th} \) is as defined in (5.20),

\( \alpha_{th}^{*} \) is as defined in (5.5)

and

\( \alpha_{th}^{*} \) is as defined in (5.19).

A gain in precision of the ratio estimator, \( \alpha_{th}^{*} \), to that of the x-only estimator, \( \alpha_{th}^{*} \), is obtained if

\[ \frac{\text{CV}(\hat{\alpha}_{th}^{*})}{\text{CV}(\hat{\alpha}_{th}^{*})} > \frac{1}{2} \frac{\text{CV}(\hat{\alpha}_{th}^{*})}{\text{CV}(\hat{\alpha}_{th}^{*})} . \]  

(5.25b)

c) no segmentation. The precinct is completely listed by the survey. We have available at the precinct or psu level, the psu total, \( x_{thi} \) \((i = 1, 2, 3, 4, 5)\) and the psu electoral count \( z_{thi} \) \((i = 1, 2, \ldots, m_{th})\). Let the estimator be

\[ \alpha_{th}^{*} = \alpha_{th} z_{th} \]  

(5.26)

where

\[ \alpha_{th} = \frac{M_{th}}{m} \sum_{i}^{2} x_{thi} \]

\[ = \frac{M_{th}}{m} \sum_{i}^{2} z_{thi} \]
or for computational purposes

\[ \alpha_{th} = \frac{1}{\sum_{1}^{\prime} \alpha_{th1}} \cdot \sum_{1}^{\prime} \alpha_{z1} \]

With the use of the relations in (5.19), we may write (5.26) as

\[ \hat{X}_{th}^* = \frac{X_{th}}{Z_{th}} \cdot Z_{th} . \]  

(5.27)

The bias in \( \hat{X}_{th}^* \) is

\[ - \text{Cov}(q_{th}, \hat{Z}_{th}^*) \]  

(5.28)

and the upper limit of this bias is

\[ |\text{Bias}| \leq \frac{1}{\sqrt{m}} \sigma_{X_{th}}^* \cdot \text{CV}(\hat{Z}_{th1}^*) \]  

(5.29)

where

\[ \text{CV}(\hat{Z}_{th1}^*) \]  

is as defined in (5.22),

\[ \sigma_{X_{th}}^* = \frac{1}{m} \left\{ \sigma_{X_{th1}}^* + \sigma_{Q_{th}}^2 \cdot \sigma_{Z_{th1}}^* \right\} \]

(5.30a)

and

\[ Q_{th} = \frac{X_{th}}{Z_{th}} . \]  

(5.30b)
To estimate the variance in equation (5.30a), we use

\[
\sigma_{\alpha_{th}}^2 = \frac{m}{m-1} \left\{ \sum_{i} \alpha_{th}^{x} + \alpha_{th}^{z} \sum_{i} \alpha_{th}^{z} \right\}
\]

(5.31)

where

- \(\alpha_{th}\) is as defined in (5.26)
- \(\alpha_{th}^{x}\) and \(\alpha_{th}^{z}\) are as defined in (5.19).

A gain in reliability of \(\hat{\alpha}_{th}\) to \(\hat{\alpha}_{th}\) is achieved if

\[
\int \hat{\alpha}_{th}^x, \hat{\alpha}_{th}^z > \frac{1}{2} \frac{CV(\hat{\alpha}_{th}^z)}{CV(\hat{\alpha}_{th}^x)}.
\]

(5.32)

The forms of the formulae for the estimator, bias, upper limit of the bias, approximate variance and its estimate are similar in the three forms of ratio estimator. The difference lies in the estimation procedure accruing from the \(i^{th}\) numbered psu and the sample mean or total of the five psu estimates. These estimates were then reduced to their corresponding self-weighting form for easy comparison with the regular unbiased self-weighting estimates developed in Chapters III and IV.

A 'separate ratio estimate' for the \(t^{th}\) sector total, \(\alpha_{X_t}\), is obtained by summing through \(h(=1,2,\ldots,L_t)\), the stratum estimates given in (5.2), (5.18a) and (5.26) and similarly for the variance of estimate and estimate of variance. Thus, the separate ratio of means estimate of \(\alpha_{X_t}\) at
the α visit assuming no segmentation in all precincts is obtained from (5.26) and the estimate is given by

\[ \hat{X}_{ts}^* = \sum_{h} \alpha_{th} \]

\[ = \sum_{h} \alpha_{th} \alpha_{zth} \quad (5.33) \]

If (5.2), (5.18a), and (5.26) are used with different number of strata, \( \alpha_{qth} \) in (5.33) is replaced by \( \hat{\alpha}_{qth} \) (5.2) or by \( \hat{\alpha}_{qth} \) (5.20) as the case may be. The auxiliary variable, \( \alpha_{zth} \), the count of registered voters in the electoral list in the th stratum at the α visit, is available at no cost.

The approximate variance formula of \( \hat{X}_{ts}^* \) is derived from (5.30a) and its estimate is obtained from (5.31). Thus

\[ \sigma_{\hat{X}_{ts}^*}^2 = \sum_{h} \sigma_{\hat{X}_{th}^*}^2 \quad (5.34) \]

and its estimate is

\[ \hat{\sigma}_{\hat{X}_{ts}^*}^2 = \sum_{h} \hat{\sigma}_{\hat{X}_{th}^*}^2 \quad (5.35) \]

With different forms of the ratio estimate in different strata, the approximate variances of the strata are added to obtain the variance of the separate ratio estimate. An estimate of each of the variances is computed in a similar manner.

A 'separate ratio estimate' for the urban area (t = 3,4)
is obtained in a similar manner. The sum on \( h \) will be through 
\((L_3 + L_4)\) or the total number of strata in the two sectors.

A 'combined ratio of means estimate' of the \( t^{th} \) sector
total, \( \alpha X_t \), will now be derived. An unbiased estimate of
\( \alpha X_t \) from the \( i^{th} \) psu is

\[
\tilde{\alpha}_t^i = \sum_{h=1}^{L_{t_e}} \tilde{\alpha}_{t^*h} + \sum_{h=1}^{L_{t_b}} \tilde{\alpha}_{t^*h} + \sum_{h=1}^{L_{t_c}} \tilde{\alpha}_{t^*h}
\]

\[= m R_t (\alpha_{t^*e}^i + \alpha_{t^*b}^i + \alpha_{t^*c}^i) \]

where

\[
\alpha_{t^*i} = \alpha_{t^*e} + \alpha_{t^*b} + \alpha_{t^*c}
\]

\( L_{t_e} \) is the number of strata under complete ratio
estimation (5.2),

\( L_{t_b} \) is the number of strata under ratio estimation in
the primary stage (5.18),

\( L_{t_c} \) is the number of strata under no segmentation (5.28),

\( L_{t_e} + L_{t_b} + L_{t_c} = L_t, (t = 3,4) \)

and the other symbols are as defined in (5.2), (5.18a) or
(5.18b), and (5.26). The components in \( \tilde{\alpha}_t^i \) are corre­
sponding numerators in the ratio in (5.5), (5.18a) or (5.18b)
and (5.26) but only from the \( i^{th} \) psu. Similarly, the de­
nominators will provide an unbiased estimate of the sector
count, \( Z_t \), from the \( i^{th} \) psu. Let the \( i^{th} \) psu estimate of
\( \alpha Z_t \) be
\[
\alpha Z_{t1}^{''} = \sum_{h=1}^{L_{t_e}} \alpha Z_{t_1 h}^{**} + \sum_{h=1}^{L_{t_b}} \alpha Z_{t_1 h}^{*} + \sum_{h=1}^{L_{t_c}} \alpha Z_{t_1 h}^{*} 
\]
\[
= m R_t \alpha Z_{t1}^{''} \tag{5.37}
\]
where
\[
\alpha Z_{t1}^{''} = \alpha Z_{t_e1}^{**} + \alpha Z_{t_b1}^{*} + \alpha Z_{t_c1}^{*}
\]
and the other symbols are as defined in (5.36).

The combined ratio of means estimate of \( \alpha X_t \) from the mean or total of the five psu is
\[
\alpha X_{tc} = c_t Z_t \
\]
where
\[
c_t = \frac{\alpha X_t^1}{\alpha Z_t^''}, \tag{5.38}
\]
\[
\alpha X_t^1 = \frac{1}{m} \sum_{1}^{'} \alpha X_{t1}^1 
\]
\[
= R_t \alpha X_{t1}^1 \tag{5.39}
\]
\[
\alpha Z_t^'' = \frac{1}{m} \sum_{1}^{'} \alpha Z_{t1}^'' 
\]
\[
= R_t \alpha Z_{t1}^'' \tag{5.40}
\]
and
\( \alpha Z_t \) is the count of registered voters in the electoral list of the \( t \)th sector.

Again, \( \alpha Z_t \) is available for all \( t \) (3.4) and at no cost to
the survey.

The bias in \( \widetilde{X}_{tc} \) is

\[
- \text{Cov}(\alpha q_t, \widetilde{Z}_t')
\]

and the upper limit of this bias is

\[
|\text{Bias}| \leq \frac{1}{\sqrt{m}} \frac{\sigma}{\alpha_{X_{tc}}} \text{CV}(\alpha_{Z_{t1}})
\]

where

\[
\text{CV}(\alpha_{Z_{t1}}) = \frac{\sigma_{Z_{t1}}^2}{\alpha_{Z_t}^2}
\]

\[
\sigma_{X_{tc}}^2 = \frac{1}{m} \left\{ \sum^m_{i=1} \alpha_{X_{t1}}^2 + \alpha_t^2 \sum^m_{i=1} \alpha_{Z_{t1}}^2 - 2 \alpha_t \sigma_{X_{t1}}, \alpha_{Z_{t1}} \right\}
\]

and

\[
\alpha_t = \frac{\alpha_{X_t}}{\alpha_{Z_t}}
\]

The estimate of the variance in (5.43) is

\[
\sigma_{X_{tc}}^2 = \frac{m R_t^2}{m-1} \left\{ \sum^m_{i=1} \alpha_{X_{t1}}^2 + \alpha_t^2 \sum^m_{i=1} \alpha_{Z_{t1}}^2 - 2 \alpha_t \sum^m_{i=1} \alpha_{X_{t1}}, \alpha_{Z_{t1}} \right\}
\]

where the symbols are as defined in (5.37), (5.38), (5.39) and (5.40).

A gain in precision is achieved with the use of \( \widetilde{X}_{tc} \) (5.38) over \( \widetilde{X}_t \) (5.39) if

\[
\int_0^{\alpha_{X_{t1}}, \alpha_{Z_{t1}}} > \frac{1}{2} \frac{\text{CV}(\alpha_{Z_{t1}})}{\text{CV}(\alpha_{X_{t1}})}.
\]
We can obtain an urban estimate \((t = 3\) and \(t = 4\)) by combining \(\tilde{X}_t^i\) for \(t = 3\) and \(t = 4\) to give an unbiased estimate of \(\alpha X_3 + \alpha X_4\) and so with \(\tilde{Z}_t^"\) for \(t = 3\) and \(t = 4\) to give an unbiased estimate of \(\alpha Z_3 + \alpha Z_4\). Thus our combined ratio estimator of \(\alpha X_3 + \alpha X_4\) is

\[
\tilde{X}_{(3+4)} = \frac{\tilde{X}_3^i + \tilde{X}_4^i}{\tilde{Z}_3^" + \tilde{Z}_4^"} (\alpha Z_3 + \alpha Z_4)
\]

\[
= \frac{\sum \tilde{X}_t^i}{\sum \tilde{Z}_t^"} \sum \alpha Z_t
\]

where

\[
\sum^t_3 = \sum^t_3
\]

For example, the \(i^{th}\) psu estimate of \(\alpha X_3 + \alpha X_4\) is

\[
\sum \tilde{X}_t^i = m R_t(\alpha X_3^i + \alpha X_4^i)
\]

and the \(i^{th}\) psu estimate of \(\alpha Z_3 + \alpha Z_4\) is

\[
\sum \tilde{Z}_t^" = m R_t(\alpha Z_3^" + \alpha Z_4^")
\]

where the symbols are as defined in (5.36) and (5.37). The bias, upper limit of bias, the approximate variance formula and its estimate and the relation to obtain a gain in precision are similar in form to (5.41), (5.45), (5.43), (5.44) and (5.45), respectively.
Three forms will be derived for the case in which a
geneneral characteristic, $x_{thijk}$, is observed in the sample
hh. The three forms of ratio estimator will depend on the
stage in which the $Z$ variable is observed or is made avail-
able. The three forms are as follows:

ea) estimation from the sample hh stage. The auxiliary
variable, $z_{thijk}$, is observed on a sample hh basis.
This may be obtained by adding a question in the
demographic block or section of the questionnaire
which is used in the second phase of the field work.
It should be pointed out that this technique is too
difficult in practice. In this case, the ratio of
means estimate will involve the ratio of two regular
unbiased estimates (4.96). The form of the estimate
for the stratum total, $x_{th}$, is

$$\hat{x}_{th} = \hat{z}_{th} z_{th}$$

(5.49)

where

$$\hat{z}_{th} = \frac{\hat{x}_{th}}{\hat{z}_{th}}$$

and

$$\hat{x}_{th}, \hat{z}_{th}$$

are as defined in (4.26).

b) estimation from the sample segment or secondary stage.
The count of registered voters, $z_{thij}$, is available
at the segment or secondary stage while $x_{thijk}$ is
observed at the hh level. In this case, the form of
the estimator is
\[ \bar{X}_{th}^{**} = \bar{X}_{th}^{*} Z_{th} \]  
(5.50)
where
\[ \bar{X}_{th}^{*} = \frac{\hat{X}_{th}}{\hat{Z}_{th}} \]
\( \hat{X}_{th} \) is as defined in (4.26)
and
\( \hat{Z}_{th} \) is as defined in (5.8).

c) estimation from the sample precinct or primary stage.
In this case, the count of registered voters, \( Z_{thi} \), is available at the precinct level and \( X_{thijk} \) is still observed at the hh level. The estimator is of the form
\[ \bar{X}_{th}^{**} = \bar{X}_{th}^{*} Z_{th} \]  
(5.51)
where
\[ \bar{X}_{th}^{*} = \frac{\hat{X}_{th}}{\hat{Z}_{th}} \]
\( \hat{X}_{th} \) is as defined in (4.26),
and
\( \hat{Z}_{th} \) is as defined in (5.19).

The bias, upper limit of the bias, approximate variance formula and its estimate and the gain in reliability of the ratio estimator to that of \( \hat{X}_{th} \), the x only estimator have the same form as described in the case of estimation for listed population and will not be given in detail. Similar forms as given in (5.33), (5.34) and (5.35) are derived for
the separate ratio estimates of a sector total or the rural area total. The combined ratio of means estimate is

\[ \alpha_{\text{gt}} \hat{X}_t = \alpha_{\text{gt}} \hat{Z}_t \]  

(5.59)

where

\[ \alpha_{\text{gt}} = \frac{\alpha_t X_t}{\alpha_t Z_t}, \]

\[ \alpha_t X_t = R_t L_t x_t, \]

\[ \alpha_t Z_t = \sum_h \alpha_{th} + \sum_h \alpha_{th}^* + \sum_h \alpha_{th}^{**}, \]

and

\[ L_{tg} + L_{tb} + L_{tc} = L_t. \]

The bias, variance and its estimate, and the relation to achieve a gain in precision to that of \( \hat{X}_t \) are easily derived and are of similar form to that obtained for a stratum estimate.

2. Unbiased ratio estimation

If the size of sample pairs, say, \( (\hat{X}_{th}, \hat{Z}_{th}) \) drawn from each of the \( L_t \) strata in the \( t^{th} \) sector is small \( (m = 5) \), then the bias in the ratio of means estimate \( (5.7, 5.18, 5.26, 5.49, 5.50, 5.51) \) of the stratum total, \( \alpha X_{th} \), may have the same sign, and the bias in the separate ratio estimate \( (5.33) \) of the sector total, \( \alpha X_t \), will be approximately \( L_t \) times that for a single stratum total while the standard deviation of the sector estimate is multiplied by \( \sqrt{L_t} \). [5, p. 130; 5]. Note
that the component estimates arising from each stratum may
have different forms. In this situation, the order of mag-
nitude of the mean square error of estimate of the sector
total is $L_t^2$; while, if the estimate of each stratum total is
unbiased, the order of magnitude is only $L_t \left[ 5 \right]$. The use of
unbiased ratio estimators is of some advantage under such
circumstances.

Hartley and Ross \[12\] gave exact expressions for the
bias in the ratio of means \[11\] and mean of ratios \[9, p.
124\]. From these expressions, an unbiased ratio estimator
is derived. In subsequent discussions on unbiased ratio
estimation, it will be assumed for simplicity and clarity,
that the precinct or psu is completely listed by the survey
(5.26) and that there is no segmentation. In addition,
empirical results given in the next section (Section C) will
utilize the forms which will be derived under the above
assumption.

The unbiased ratio estimate of the stratum total, $\bar{X}_{th}$,
is defined as

$$\bar{X}_{th} = \alpha^{r}_{th} \bar{Z}_{th} + \frac{m}{m - 1} (\hat{X}_{th}^{*} - \alpha^{r}_{th} \hat{Z}_{th}^{*}) \quad (5.53)$$

where

$$\alpha^{r}_{th} = \frac{1}{m} \sum_{i}^{'} \alpha^{r}_{thi}$$

$$\alpha^{r}_{thi} = \frac{\hat{X}_{thi}}{\hat{Z}_{thi}}$$
\[ \hat{\phi}_{\text{th}}, \hat{\psi}_{\text{th}}, \hat{\theta}_{\text{th}} \text{ and } \hat{\sigma}_{\text{th}} \text{ are as defined in (5.19), (5.21), and (5.26)}, \]
\[ \hat{\theta}_{\text{th}} \text{ is the } \text{th} \text{ th stratum count of the auxiliary variable,} \]
and
\[ m = 5. \]

Note that the first term in (5.53) is the mean of ratio estimate \[ \bar{r}_{\text{th}} \text{ of } \phi \]. Let
\[ \hat{\phi}_{\text{th}} = \phi_{\text{th}} \hat{\sigma}_{\text{th}}. \] (5.54)
The bias in (5.54) is
\[ - \text{Cov}(\phi_{\text{th}}, \hat{\sigma}_{\text{th}}) \] which is estimated unbiasedly by the expression
\[ \frac{1}{m-1} \sum (\phi_{\text{th}} - \bar{\phi}_{\text{th}})(\hat{\sigma}_{\text{th}} - \hat{\sigma}_{\text{th}}) \]
\[ = - \frac{m}{m-1} (\hat{\phi}_{\text{th}} - \phi_{\text{th}} \hat{\sigma}_{\text{th}} - \hat{\sigma}_{\text{th}}) \] (5.56)
where the symbols are as defined in (5.53). The expression in (5.56) is added to (5.54) with a positive sign to derive the unbiased ratio estimator, \[ \hat{\phi}_{\text{th}} (5.53). \]

The exact variance formula of \[ \hat{\phi}_{\text{th}} \] is derived [5] by essentially utilizing the moments of bivariate sample cumulants. A simpler form of this variance [5] is
\[ \sigma^2_{\phi_{\text{th}}} = \frac{1}{m} \sigma^2_{\phi_{\text{th}}^{*}} - R_{\text{th}} \bar{\phi}_{\text{th}} \bar{\sigma}_{\text{th}}^{*} \]
\[ + \frac{1}{m(m-1)} (1 + \sum \sigma^2_{\phi_{\text{th}}, \sigma_{\text{th}}^{*}} \sigma^2_{\phi_{\text{th}}, \sigma_{\text{th}}^{*}} \sigma^2_{\phi_{\text{th}}, \sigma_{\text{th}}^{*}}) \]
\[ (5.57) \]
which becomes for large $m$,

$$\sigma_{X_{th}}^2 = \frac{1}{m} \sigma_{X_{th}}^2 - R_{th} Z_{th}^2$$  \hspace{1cm} (5.58a)

or

$$\sigma_{X_{th}}^2 = \frac{1}{m} \left\{ \sigma_{X_{th}}^2 + R_{th} \sigma_{Z_{th}}^2 \right\} - 2 R_{th} \sigma_{X_{th}}^2 \left\{ \sigma_{R_{th}}^2 \right\}$$  \hspace{1cm} (5.58b)

where

$$R_{th} = E(\alpha_{R_{th}}^2).$$

The variance of $X_{th}$ in (5.58b) may be compared with the approximate variance formula (5.30) of the biased ratio of means estimate (5.26). Equation (5.30) is appropriate for large $m$ and is derivable by a Taylor expansion with exact remainder term $[11]$. This form is considered as standard in literature. Thus, the difference between (5.30) and (5.58b) is

$$\sigma_{X_{th}}^2 - \sigma_{X_{th}}^2 = \frac{1}{m} \left\{ \sigma_{Q_{th}}^2 - \sigma_{R_{th}}^2 \right\} \sigma_{Z_{th}}^2 - 2 \sigma_{X_{th}}^2 \left\{ \sigma_{Q_{th}}^2 - \sigma_{R_{th}}^2 \right\}$$  \hspace{1cm} (5.59)

$$= \frac{1}{m} \sigma_{Z_{th}}^2 \left\{ \sigma_{Q_{th}}^2 - \sigma_{B_{th}}^2 \right\} - \left( \sigma_{R_{th}}^2 - \sigma_{B_{th}}^2 \right)$$

where

$$Q_{th} = \sigma_{Q_{th}}^2$$

$$R_{th} = E(\alpha_{R_{th}}^2).$$
\[ \beta_{th} = \frac{\sigma_{\hat{x}_{th}}^2}{\sigma_{\hat{z}_{th}}^2} \]

is the regression coefficient of \( \hat{x}_{th} \) on \( \hat{z}_{th} \).

\( \alpha_{th} \) is more efficient than \( \hat{x}_{th} \) if

\[ (\alpha_{Q_{th}} - \beta_{th})^2 > (\alpha_{R_{th}} - \beta_{th})^2 \]

i.e., the regression coefficient, \( \beta_{th} \), is closer to \( \alpha_{Q_{th}} \) than to \( \alpha_{R_{th}} \). When \( \alpha_{R_{th}} = \alpha_{Q_{th}} \) or \( \beta_{th} \) is halfway between \( \alpha_{R_{th}} \) and \( \alpha_{Q_{th}} \), i.e., when \( \hat{x}_{th} \) is unbiased, then \( \alpha_{X_{th}} \) and \( \hat{x}_{th} \) are equally efficient. Experience of Goodman and Hartley [5] seems to indicate that for \( \alpha_{R_{th}} \neq \alpha_{Q_{th}} \), \( \beta_{th} \) is often found closer to \( \alpha_{Q_{th}} \) than to \( \alpha_{R_{th}} \) and in this case \( \hat{x}_{th} \) is more efficient than \( \alpha_{X_{th}} \). The extent of the bias in \( \hat{x}_{th} \) must be taken into consideration in making a decision on the estimator to use.

The appropriate estimate of variance given in (5.58) is

\[ s_{\alpha_{th}}^2 = \frac{1}{m} \left\{ s_{\hat{x}_{th}}^2 + \alpha_{R_{th}}^2 s_{\hat{z}_{th}}^2 - 2 \alpha_{R_{th}} s_{\hat{z}_{th}} s_{\hat{x}_{th}}^* \right\} \] (5.60)

where

\[ \alpha_{R_{th}} = \frac{1}{m} \sum_{i}^{'} \alpha_{r_{th}} \]

\[ s_{\alpha_{th}}^2 = \frac{1}{(m - 1)} \sum_{i}^{'} \left( \hat{x}_{th} - \alpha_{r_{th}} \right)^2 \]

and
Nieto [23] has shown that

\[
\sum_{i=1}^{m-1}\left(\hat{x}_{\text{th}i} - \hat{x}_{\text{th}}\right)\left(\hat{z}_{\text{th}i} - \hat{z}_{\text{th}}\right) = \frac{1}{m-1}\sum_{i=1}^{m-1}\left(\hat{x}_{\text{th}i} - \hat{x}_{\text{th}}\right)\left(\hat{z}_{\text{th}i} - \hat{z}_{\text{th}}\right).
\]

where

\[
\alpha^2_{\text{th}} = \left\{\frac{m-1}{m-1}\right\}^2 \left(\hat{x}_{\text{th}} - \alpha_{\text{th}} \hat{x}_{\text{th}}\right)^2,
\]

estimates with a slight overestimation

\[
(1 + \int_{\alpha_{\text{th}}}^{\hat{x}_{\text{th}}} \sigma_{\text{th}}^2 \sigma_{\text{th}}^2)\frac{\alpha_{\text{th}}}{\alpha_{\text{th}}^2} \left(\hat{z}_{\text{th}} - \alpha_{\text{th}} \hat{z}_{\text{th}}\right)^2
\]

which is the second term in the variance expression for \(\alpha_{\text{th}}\) (5.57). In fact, an upper limit estimate of this second term is obtained if in (5.61), \(\int_{\alpha_{\text{th}}}^{\hat{x}_{\text{th}}} \sigma_{\text{th}}^2 \sigma_{\text{th}}^2\) is made equal to one

which implies a maximum value for \(\alpha_{\text{th}}\) (5.57). Thus, to estimate completely (5.57), we use the sum of (5.60) and (5.61) with \(\int_{\alpha_{\text{th}}}^{\hat{x}_{\text{th}}} \sigma_{\text{th}}^2 \sigma_{\text{th}}^2 = 1,\)

\[
\hat{\sigma}_{\alpha_{\text{th}}}^2 = \frac{1}{m}\left\{\frac{s_{\alpha_{\text{th}}}^2}{2} + \alpha_{\text{th}} \sigma_{\text{th}}^2 \sigma_{\text{th}}^2 - \alpha_{\text{th}} \sigma_{\text{th}}^2 \hat{z}_{\text{th}} - \alpha_{\text{th}} \hat{z}_{\text{th}}\right\} + \left\{\frac{m-1}{m-1}\right\}^2 \frac{\sigma_{\text{th}}^2}{(m^2 - m + 1) + 1}\left(\hat{x}_{\text{th}} - \alpha_{\text{th}} \hat{x}_{\text{th}}\right)^2.
\]
To indicate the magnitude of the second term in (5.62) for \( m = 5 \),

\[
\left( \frac{m}{m-1} \right)^2 \frac{\sigma^2}{\left( m^2 - m + 1 \right) + 1} = 0.142 .
\]

A separate unbiased ratio estimator of the sector total or a regional total may be derived with the use of (5.53). With independent sampling in each stratum, the separate unbiased ratio estimate of the \( t \)th sector total, \( \alpha X_t \), is

\[
\alpha X^1_{ts} = \sum_{h} \alpha^1_{th}
\]

\[
= \sum_{h} \left\{ r_{th} \alpha Z_{th} + \frac{m}{m-1} \left( \alpha^*_h - r_{th} \alpha^{2*}_{th} \right) \right\}
\]

(5.63)

where

\( L_t \) is the total number of strata in the \( t \)th sector

\( (t = 3, 4) \), or the number of strata in a region if the estimate is for a regional total.

The variance of \( \alpha X^1_{ts} \) is obtained from (5.57) and is given by

\[
\sigma^2_{\alpha X^1_{ts}} = \sum_{h} \sigma^2_{\alpha X^1_{th}}
\]

\[
= \sum_{h} \left\{ \frac{1}{m} \sigma^2_{\alpha^{*}_h} - \alpha R_{th} \sigma^2_{\alpha^{2*}_th} + \frac{1}{m(m-1)} \right\}
\]

\[
(1 + \rho^2_{\alpha^{*}_{th}, \alpha^{2*}_{th}} \sigma^2_{\alpha R_{th}} \sigma^2_{\alpha^{2*}_{th}})
\]

(5.64)
Estimate of this variance is given by

\[ s_{X \alpha}^2 = \sum_h L_t s_{X \alpha}^2 \]  

(5.65)

where

\[ s_{X \alpha}^2 \] is defined in (5.60),

or by

\[ \hat{s}_{X \alpha}^2 = \sum_h \hat{s}_{X \alpha}^2 \]

where

\[ \hat{s}_{X \alpha}^2 \] is defined in (5.62).

The 'combined unbiased ratio estimator' of the \( t \)th sector total at the \( \alpha \) visit, \( \alpha X_t \) (\( t = 3,4 \)), is given by

\[ \alpha X_{tc} = \alpha r_t \alpha Z_t + \frac{m}{m-1} (\alpha \hat{X}_{ti} - \alpha r_t \hat{Z}_{ti}) \]  

(5.66)

where

\[ \alpha r_t = \frac{1}{m} \sum_i^r \alpha r_{ti} \]

\[ \alpha r_{ti} = \frac{\alpha \hat{X}_{ti}}{\alpha \hat{Z}_{ti}} \]

\[ \alpha \hat{X}_{ti} = \sum_h \alpha \hat{X}_{thi} \]

\[ \alpha \hat{X}_t = \frac{1}{m} \sum_i^r \alpha \hat{X}_{ti} \]
\( \hat{X}^*_{\text{thi}} \) is defined in (5.26),
\( Z^*_t \) is the \( t \)th sector count of the auxiliary variable, and
\( m = 5. \)
\( \hat{Z}^*_{\text{thi}}, \hat{Z}^*_{t1} \) and \( \hat{Z}^*_t \) are as defined for the \( X \)'s. The variance of \( \hat{X}^*_{\text{tc}} \) is similar in form to that given in (5.57) or (5.58). The components therein are however in terms of sector totals, sector ratio and sector correlation. The exact variance is
\[
\sigma^2_{X^*_{\text{tc}}} = \frac{1}{m} \left\{ \sigma^2_{\hat{X}^*_{t1}} + \lambda^2_{rt} \sigma^2_{\hat{Z}^*_{t1}} - \lambda_{rt} \right\}
\]
\[\text{where} \quad \lambda_{rt} = E(\hat{X}^*_{t1}). \]
To estimate the expression in (5.67), we have
\[
\frac{1}{m} \left\{ \sigma^2_{\hat{X}^*_{t1}} + \lambda^2_{rt} \sigma^2_{\hat{Z}^*_{t1}} - \lambda_{rt} \right\}
\]
\[ s_{x_{ti}}, \hat{Z}_{ti} = \frac{1}{m - 1} \sum_{i}^{n'} \left( \hat{x}_{ti} - \bar{x}_{i} \right) \left( \hat{Z}_{ti} - \bar{Z}_{i} \right), \]

and

\[ \alpha_{rt} = \frac{1}{m} \sum_{i}^{n'} \alpha_{rt}. \]

An estimate of (5.66) is

\[ \hat{s}_{x_{tc}} = \left\{ \frac{s_{x_{ti}}^2 + \alpha_{rt}^2 s_{Z_{ti}}^2 - 2 \alpha_{rt} s_{x_{ti}} s_{Z_{ti}}}{s_{x_{ti}}^2 - \alpha_{rt}^2 s_{Z_{ti}}^2} \right\} \]

\[ \text{where the symbols are as defined in (5.62).} \]

C. Empirical Results

Data on the number of registered voters, \( x_{z_{thi}} \), by sample precinct were obtained from the Philippine Electoral Commission for seven randomly selected strata \( h = 1, 7, 14, 20, 21, 30, 32 \) in urban Manila \( t = 4 \). The subscript \( \alpha \) in \( x_{z_{thi}} \) refers to the November 1957 election. Data on the number of listed population in sample precinct, \( x_{x_{thi}} \), and the number of interviewed population in sample hhs in sample precinct, \( x_{x_{thi}} \), for the same seven strata were obtained from OSCAS, NEC. The \( \alpha \) subscript in \( x_{x_{thi}} \) and \( x_{x_{thi}} \) refers to the October 1957 visit of the PSSH. Registration of voters for the November 1957 election was held at about the same time.
as the interview week of the October 1957 round.

$\alpha_{X_{th}},$ $\alpha_{X_{th}}$ and $\alpha_{Z_{th}}$ are given in columns 5, 6 and 7 of Table 10, respectively. An estimate of the $\text{th }$ stratum total, $X_{th}$ or $Z_{th},$ is computed from each of the five numbered psus (precincts) in a given stratum, using the appropriate $\alpha_{X_{th}},$ $\alpha_{X_{th}},$ or $\alpha_{Z_{th}}.$ These estimates are given in columns 8, 9 and 10 of Table 10, where

$$\hat{X}_{th}^* = m R_t \alpha_{X_{th}}^*,$$

and

$$\hat{Z}_{th}^* = m R_t \alpha_{Z_{th}}^*$$

are as defined in (5.26),

$$\hat{X}_{th} = m R_t \alpha_{X_{th}}$$

is as defined in (4.26),

$$m = 5,$$

and

$$R_t = 400 \ (t = 4).$$

The number of psus, $X_{th},$ in the $\text{th }$ stratum and the stratum constant, $f_{th},$ are given in columns 2 and 3, respectively, of Table 10.

An analysis of covariance of listed population estimate, $\hat{X}_{th}^*,$ on registered voters estimate, $\hat{Z}_{th}^*,$ is prepared and the results are shown in Table 11. A pooled estimate of the correlation between $\hat{X}_{th}^*$ and $\hat{Z}_{th}^*$ is computed with the help of the between primary within strata mean squares. The pooled sample correlation is
Table 10. Sample of strata in urban Manila for estimating gain in reliability of biased and unbiased ratio estimators

<table>
<thead>
<tr>
<th>Stratum (h)</th>
<th>$X_{th}$</th>
<th>$f_{th}$</th>
<th>Sample precinct (1)</th>
<th>$X_{th1}$</th>
<th>$X_{th2}$</th>
<th>$Z_{th1}$ (thousands)</th>
<th>$X_{th1}$ (thousands)</th>
<th>$Z_{th1}$ (thousands)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
<td>(7)</td>
<td>(8)</td>
<td>(9)</td>
</tr>
<tr>
<td>1</td>
<td>770</td>
<td>0.036</td>
<td>1</td>
<td>549</td>
<td>15</td>
<td>227</td>
<td>39.5</td>
<td>30.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>1127</td>
<td>25</td>
<td>180</td>
<td>81.1</td>
<td>104.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>937</td>
<td>35</td>
<td>151</td>
<td>67.5</td>
<td>66.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>759</td>
<td>42</td>
<td>176</td>
<td>54.6</td>
<td>94.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>899</td>
<td>19</td>
<td>180</td>
<td>59.7</td>
<td>38.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>302.4</td>
<td>322.0</td>
</tr>
<tr>
<td>7</td>
<td>68</td>
<td>0.034</td>
<td>1</td>
<td>543</td>
<td>26</td>
<td>143</td>
<td>36.9</td>
<td>55.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>699</td>
<td>24</td>
<td>275</td>
<td>47.5</td>
<td>57.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>422</td>
<td>25</td>
<td>142</td>
<td>33.5</td>
<td>35.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>531</td>
<td>26</td>
<td>251</td>
<td>38.1</td>
<td>50.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>1358</td>
<td>60</td>
<td>223</td>
<td>91.9</td>
<td>120.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>345.9</td>
<td>332.0</td>
</tr>
<tr>
<td>14</td>
<td>80</td>
<td>0.040</td>
<td>1</td>
<td>1865</td>
<td>69</td>
<td>248</td>
<td>149.7</td>
<td>138.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>1520</td>
<td>66</td>
<td>179</td>
<td>121.6</td>
<td>119.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>350</td>
<td>10</td>
<td>170</td>
<td>28.0</td>
<td>20.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>1138</td>
<td>35</td>
<td>309</td>
<td>91.0</td>
<td>70.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>2226</td>
<td>92</td>
<td>341</td>
<td>171.1</td>
<td>124.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>567.9</td>
<td>524.0</td>
</tr>
<tr>
<td>20</td>
<td>80</td>
<td>0.040</td>
<td>1</td>
<td>577</td>
<td>22</td>
<td>222</td>
<td>46.2</td>
<td>44.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>916</td>
<td>30</td>
<td>230</td>
<td>73.3</td>
<td>60.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>870</td>
<td>40</td>
<td>279</td>
<td>69.6</td>
<td>80.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>1405</td>
<td>57</td>
<td>281</td>
<td>112.4</td>
<td>114.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>1642</td>
<td>69</td>
<td>309</td>
<td>131.8</td>
<td>138.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>433.4</td>
<td>436.0</td>
</tr>
</tbody>
</table>
Table 10. (Continued)

<table>
<thead>
<tr>
<th>Stratum (n)</th>
<th>$X_{th}$</th>
<th>$f_{th}$</th>
<th>Sample precinct (i)</th>
<th>$X_{th1}$</th>
<th>$X_{th2}$</th>
<th>$Z_{th1}$ (thousands)</th>
<th>$X_{th2}$ (thousands)</th>
<th>$Z_{th2}$ (thousands)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) 21</td>
<td>74</td>
<td>0.037</td>
<td>1</td>
<td>1030</td>
<td>36</td>
<td>172</td>
<td>76.2</td>
<td>72.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>580</td>
<td>13</td>
<td>192</td>
<td>43.7</td>
<td>26.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>1064</td>
<td>37</td>
<td>336</td>
<td>78.7</td>
<td>74.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>469</td>
<td>10</td>
<td>118</td>
<td>34.7</td>
<td>20.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>704</td>
<td>37</td>
<td>195</td>
<td>53.2</td>
<td>74.0</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>286.9</td>
<td>266.0</td>
</tr>
<tr>
<td>(3) 30</td>
<td>100</td>
<td>0.050</td>
<td>1</td>
<td>1963</td>
<td>109</td>
<td>265</td>
<td>195.3</td>
<td>218.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>1046</td>
<td>69</td>
<td>181</td>
<td>104.6</td>
<td>138.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>1057</td>
<td>65</td>
<td>317</td>
<td>105.7</td>
<td>130.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>868</td>
<td>54</td>
<td>129</td>
<td>96.8</td>
<td>108.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>1319</td>
<td>72</td>
<td>281</td>
<td>131.2</td>
<td>144.0</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>624.3</td>
<td>738.0</td>
</tr>
<tr>
<td>(5) 32</td>
<td>76</td>
<td>0.038</td>
<td>1</td>
<td>515</td>
<td>27</td>
<td>210</td>
<td>39.1</td>
<td>54.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>498</td>
<td>15</td>
<td>173</td>
<td>30.5</td>
<td>35.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>1005</td>
<td>47</td>
<td>232</td>
<td>76.4</td>
<td>94.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>888</td>
<td>26</td>
<td>320</td>
<td>67.5</td>
<td>55.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>590</td>
<td>20</td>
<td>225</td>
<td>45.0</td>
<td>40.0</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>280.5</td>
<td>270.0</td>
</tr>
<tr>
<td>G. Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2781.3</td>
<td>2872.0</td>
</tr>
<tr>
<td>G. Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>77.75</td>
<td>17.77</td>
</tr>
</tbody>
</table>
Table 11. Analysis of covariance of listed population, $\alpha_{thi}$, on registered voters, $\alpha_{thi}$

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Degrees of freedom</th>
<th>Sum of squares (thousands)</th>
<th>Mean square (thousands)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$X^2$</td>
<td>$Z^2$</td>
</tr>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Between strata</td>
<td>6</td>
<td>28,851.14</td>
<td>470.37</td>
</tr>
<tr>
<td>Within strata</td>
<td>26</td>
<td>31,684.19</td>
<td>594.50</td>
</tr>
<tr>
<td>Total</td>
<td>34</td>
<td>60,535.33</td>
<td>1,064.87</td>
</tr>
</tbody>
</table>
An estimate of the coefficient of variation (CV) for an individual estimate, $\hat{\alpha}_{\text{thi}}^*$ or $\hat{\beta}_{\text{thi}}^*$, is obtained from Tables 10 and 11. The estimated CV's are

$$ CV(\hat{\alpha}_{\text{thi}}^*) = \frac{33.64}{77.75} = 0.4327 ,$$

and

$$ CV(\hat{\beta}_{\text{thi}}^*) = \frac{4.61}{17.77} = 0.2594 .$$

A gain in reliability is achieved in the use of $\hat{\alpha}_{\text{thi}}^*$ (5.26) to that of the $x$ only estimate, $\hat{\alpha}_{\text{thi}}$, if

$$ \int_{\hat{\alpha}_{\text{thi}}}^{\hat{\beta}_{\text{thi}}} \frac{1}{2} \frac{CV(\hat{\beta}_{\text{thi}}^*)}{CV(\hat{\alpha}_{\text{thi}}^*)} .$$

This relation is estimated by

$$ r_{\hat{\alpha}_{\text{thi}}, \hat{\beta}_{\text{thi}}} = 0.4929 $$

or

$$ \frac{1}{2} \frac{CV(\hat{\beta}_{\text{thi}}^*)}{CV(\hat{\alpha}_{\text{thi}}^*)} = 0.2998 $$
By and large, we expect a gain in precision in the use of $\hat{X}_{th}$, the ratio of means estimator as compared to $\alpha_{X_{th}}$, the x-only estimator.

Similarly, an analysis of covariance of the interviewed population estimate, $\hat{X}_{th}$, on the registered voters estimate, $\hat{Z}_{th}$, is computed and the results are summarized in Table 12.

From Table 12, we have

$$r_{\hat{X}_{th}, \hat{Z}_{th}} = 0.4567.$$  

From Tables 10 and 12, we obtain

$$\frac{1}{2} \frac{\text{CV}(\hat{Z}_{th})}{\text{CV}(\hat{X}_{th})} = 0.9774.$$  

Finally, we have the relation

$$r_{\hat{X}_{th}, \hat{Z}_{th}} > \frac{1}{2} \frac{\text{CV}(\hat{Z}_{th})}{\text{CV}(\hat{X}_{th})}.$$  

This result indicates that, on the average, there is a gain in reliability in the use of $\hat{X}_{th}$ (5.51), as compared to the regular unbiased estimate, $X_{th}$.

The mean of ratios ($r_{th}$), ratio of means ($\alpha_{X_{th}}$), and estimate of the variance of $\hat{X}_{th}$ and $\hat{Z}_{th}$ and their covariances are computed for each of the seven strata. The results
### Table 12. Analysis of covariance of interviewed population, $\hat{X}_{th1}$, on registered voters, $\hat{Z}_{th1}$

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Degrees of freedom</th>
<th>Sum of squares</th>
<th>XZ</th>
<th>Mean square</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$X^2$</td>
<td>$Z^2$</td>
<td>$XZ$</td>
</tr>
<tr>
<td>Between strata</td>
<td>6</td>
<td>35,414.18</td>
<td>470.37</td>
<td>3,466.93</td>
</tr>
<tr>
<td>Within strata</td>
<td>28</td>
<td>41,375.00</td>
<td>594.50</td>
<td>2,056.68</td>
</tr>
<tr>
<td>Total</td>
<td>34</td>
<td>76,789.18</td>
<td>1,064.87</td>
<td>5,523.61</td>
</tr>
</tbody>
</table>
are summarized in Table 13.

For \( h = 1 \) (Table 10),

\[
\alpha_{1}^{\text{th}} = \frac{1}{5} \left\{ 32.5 + 61.1 + 67.5 + 54.6 + 59.7 \right\} = 4.75 ,
\]

\[
\alpha_{1}^{\text{th}} = \frac{302.4}{65.9} = 4.59 ,
\]

\[
\frac{s_{1}^{2}}{\alpha_{1}^{\text{th}}} = \frac{1}{20} \left\{ 18.3^2 + 13.0^2 + 10.9^2 + 12.7^2 + 13.0^2 - \frac{65.9^2}{5} \right\} = 0.76
\]

and so on. From Table 13, the estimated variances of the

<table>
<thead>
<tr>
<th>Stratum (h)</th>
<th>( \alpha_{1}^{\text{th}} )</th>
<th>( \alpha_{1}^{\text{th}} )</th>
<th>( \frac{s_{1}^{2}}{\alpha_{1}^{\text{th}}} ) (thousands)</th>
<th>( \frac{s_{1}^{2}}{\alpha_{1}^{\text{th}}} ) (thousands)</th>
<th>( \frac{s_{1}^{2}}{\alpha_{1}^{\text{th}}} ) (thousands)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.75</td>
<td>4.59</td>
<td>47.49</td>
<td>0.76</td>
<td>-4.11</td>
</tr>
<tr>
<td>7</td>
<td>3.31</td>
<td>3.18</td>
<td>119.78</td>
<td>2.37</td>
<td>0.82</td>
</tr>
<tr>
<td>14</td>
<td>5.66</td>
<td>5.70</td>
<td>666.49</td>
<td>7.44</td>
<td>42.98</td>
</tr>
<tr>
<td>20</td>
<td>3.96</td>
<td>4.05</td>
<td>240.79</td>
<td>1.73</td>
<td>17.62</td>
</tr>
<tr>
<td>21</td>
<td>4.29</td>
<td>4.29</td>
<td>76.23</td>
<td>2.02</td>
<td>2.29</td>
</tr>
<tr>
<td>30</td>
<td>5.58</td>
<td>5.32</td>
<td>361.77</td>
<td>11.94</td>
<td>29.97</td>
</tr>
<tr>
<td>32</td>
<td>2.93</td>
<td>2.95</td>
<td>71.56</td>
<td>3.53</td>
<td>10.44</td>
</tr>
</tbody>
</table>
regular unbiased estimate (the x-only estimate), \( \hat{\alpha}^{\text{xth}} \) (5.12); the biased ratio of means estimate, \( \hat{\alpha}^{X^*}_{\text{xth}} \) (5.26); and the unbiased ratio estimate, \( \hat{\alpha}^{X^*}_{\text{xth}} \) (5.53) and the relative efficiencies of \( \hat{\alpha}^{X^*}_{\text{xth}} \) and \( \hat{\alpha}^{X^*}_{\text{xth}} \) to \( \hat{\alpha}^{X^*}_{\text{xth}} \) are obtained for each stratum. The results are summarized in Table 14. For clarity and easy reference, the estimated variances of \( \hat{\alpha}^{X^*}_{\text{xth}}, \hat{\alpha}^{X^*}_{\text{xth}}, \) and \( \hat{\alpha}^{X^*}_{\text{xth}} \) are given below. For \( \hat{\alpha}^{X^*}_{\text{xth}}, \)

\[ s_{\hat{\alpha}^{X^*}_{\text{xth}}}^2 \]

is as given in column 4, Table 13, and is computed similar to \( s_{\hat{\alpha}^{X^*}_{\text{xth}}}^2 \) above.

For \( \hat{\alpha}^{X^*}_{\text{xth}} \) (5.26, 5.31),

\[ s_{\hat{\alpha}^{X^*}_{\text{xth}}}^2 = s_{\hat{\alpha}^{X^*}_{\text{xth}}}^2 + s_{\hat{\alpha}^{X^*}_{\text{xth}}}^2 - s_{\hat{\alpha}^{X^*}_{\text{xth}}}^2 + s_{\hat{\alpha}^{X^*}_{\text{xth}}}^2 - s_{\hat{\alpha}^{X^*}_{\text{xth}}}^2 \]

For \( \hat{\alpha}^{X^*}_{\text{xth}} \) (5.53, 5.60),

\[ s_{\hat{\alpha}^{X^*}_{\text{xth}}}^2 = s_{\hat{\alpha}^{X^*}_{\text{xth}}}^2 + s_{\hat{\alpha}^{X^*}_{\text{xth}}}^2 - s_{\hat{\alpha}^{X^*}_{\text{xth}}}^2 + s_{\hat{\alpha}^{X^*}_{\text{xth}}}^2 - s_{\hat{\alpha}^{X^*}_{\text{xth}}}^2 \]

Note that the gain (\( h = 14, 20, 21, 30 \)) or loss (\( h = 1, 7, 30 \)) in estimated efficiency of \( \hat{\alpha}^{X^*}_{\text{xth}} \) and \( \hat{\alpha}^{X^*}_{\text{xth}} \) to \( \hat{\alpha}^{X^*}_{\text{xth}} \) are approximately equal for all \( h \). These results are to be expected since \( \hat{\alpha}^{X^*}_{\text{xth}} \) is approximately equal to \( \hat{\alpha}^{X^*}_{\text{xth}} \) for the seven strata. From these estimates, we say that \( \hat{\alpha}^{X^*}_{\text{xth}} \) and \( \hat{\alpha}^{X^*}_{\text{xth}} \) are equally efficient.

If the seven strata are considered as a sector, then five different estimates of the sector total, \( \hat{\alpha}^{X_{\text{t}}}_{\text{xth}} \), may be compared, namely:
Table 14. The estimated variances of $\hat{X}_{1h}$, $\hat{X}_{2h}$ and $\hat{X}_{3h}$ and the relative efficiencies of $\hat{X}_{1h}$ and $\hat{X}_{2h}$ to $\hat{X}_{3h}$ for each stratum

<table>
<thead>
<tr>
<th>Stratum (h)</th>
<th>Estimator</th>
<th>Estimated variance (thousands)</th>
<th>Relative efficiency (per cent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>1</td>
<td>$\hat{X}_{1h}$</td>
<td>47.49</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>$\hat{X}_{2h}$</td>
<td>101.93</td>
<td>48.91</td>
</tr>
<tr>
<td></td>
<td>$\hat{X}_{3h}$</td>
<td>103.68</td>
<td>45.80</td>
</tr>
<tr>
<td>7</td>
<td>$\hat{X}_{1h}$</td>
<td>119.78</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>$\hat{X}_{2h}$</td>
<td>138.52</td>
<td>86.47</td>
</tr>
<tr>
<td></td>
<td>$\hat{X}_{3h}$</td>
<td>140.33</td>
<td>85.36</td>
</tr>
<tr>
<td>14</td>
<td>$\hat{X}_{1h}$</td>
<td>666.49</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>$\hat{X}_{2h}$</td>
<td>418.25</td>
<td>159.35</td>
</tr>
<tr>
<td></td>
<td>$\hat{X}_{3h}$</td>
<td>417.00</td>
<td>159.33</td>
</tr>
<tr>
<td>20</td>
<td>$\hat{X}_{1h}$</td>
<td>240.79</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>$\hat{X}_{2h}$</td>
<td>126.44</td>
<td>190.44</td>
</tr>
<tr>
<td></td>
<td>$\hat{X}_{3h}$</td>
<td>122.37</td>
<td>187.57</td>
</tr>
<tr>
<td>21</td>
<td>$\hat{X}_{1h}$</td>
<td>76.23</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>$\hat{X}_{2h}$</td>
<td>42.27</td>
<td>180.34</td>
</tr>
<tr>
<td></td>
<td>$\hat{X}_{3h}$</td>
<td>42.27</td>
<td>180.34</td>
</tr>
</tbody>
</table>

*The estimated variance of $\hat{X}_{3h}$ is used as numerator.*
Table 14. (Continued)

<table>
<thead>
<tr>
<th>Stratum (h)</th>
<th>Estimator</th>
<th>Estimated variance (thousands)</th>
<th>Relative efficiency (per cent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>30</td>
<td>$\hat{x}_{th}$</td>
<td>361.77</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>$\hat{x}_{th}$</td>
<td>380.79</td>
<td>95.00</td>
</tr>
<tr>
<td></td>
<td>$x_{th}$</td>
<td>399.11</td>
<td>90.64</td>
</tr>
<tr>
<td>32</td>
<td>$\hat{x}_{ts}$</td>
<td>71.56</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>$x_{ts}$</td>
<td>40.67</td>
<td>175.95</td>
</tr>
<tr>
<td></td>
<td>$x_{tc}$</td>
<td>40.67</td>
<td>175.95</td>
</tr>
</tbody>
</table>

$\hat{x}_t$ (5.26, 5.66) is the regular unbiased estimate (the x-only estimate),

$\hat{x}_{ts}$ (5.33) is the separate biased ratio of means estimate,

$\hat{x}_{tc}$ (5.66) is the combined unbiased ratio estimate.

Data for the sector comparisons are given in Table 15. The estimated variances of the five estimators and the relative efficiency of $\hat{x}_{ts}$, $\hat{x}_{tc}$ to $\hat{x}_t$ are given in Table 16. Columns 2 and 3 of Table 15 are derived from
Table 15. Data for sector computations

<table>
<thead>
<tr>
<th>1&lt;sup&gt;st&lt;/sup&gt; ordered precinct</th>
<th>( \hat{X}_{t1} ) (thousands)</th>
<th>( \hat{Z}_{t1} ) (thousands)</th>
<th>( \alpha_{t1} )</th>
<th>( \alpha_t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>582.4</td>
<td>119.9</td>
<td>4.89</td>
<td>--</td>
</tr>
<tr>
<td>2</td>
<td>504.3</td>
<td>109.1</td>
<td>4.66</td>
<td>--</td>
</tr>
<tr>
<td>3</td>
<td>459.4</td>
<td>130.1</td>
<td>3.53</td>
<td>--</td>
</tr>
<tr>
<td>4</td>
<td>483.1</td>
<td>103.9</td>
<td>3.90</td>
<td>--</td>
</tr>
<tr>
<td>5</td>
<td>692.1</td>
<td>139.2</td>
<td>4.90</td>
<td>--</td>
</tr>
<tr>
<td>Total</td>
<td>2721.3</td>
<td>622.1</td>
<td>21.89</td>
<td>--</td>
</tr>
<tr>
<td>Mean</td>
<td>544.3</td>
<td>124.4</td>
<td>4.38</td>
<td>4.37</td>
</tr>
</tbody>
</table>

Columns 8 and 10 of Table 10, respectively. For example (i = 1),

\[
\hat{X}_{t1} = \left\{ 39.5 + 36.9 + 149.2 + 46.2 + 76.2 + 145.3 + 39.1 \right\} \\
= 582.4 \\
\]

\[
\alpha_{t1} = \frac{582.4}{119.2} \\
= 4.89 \\
\]

\[
\alpha_t = \frac{21.89}{5} \\
= 4.38 \\
\]

\[
\alpha_t = \frac{544.3}{124.4} \\
= 4.37 \\
\]
Table 16. The estimated variances of sector estimates, $\alpha_{xt}$, $\hat{\alpha}_{x_{t_s}}$, $\hat{\alpha}_{x_{t_c}}$, and $\hat{\alpha}_{x_{t}}$ and the relative efficiencies of $\hat{\alpha}_{x_{t_s}}$, $\alpha_{x_{t_s}}$, $\hat{\alpha}_{x_{t_c}}$, and $\hat{\alpha}_{x_{t}}$

to $\hat{\alpha}_{x_{t}}$

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Estimated variance (thousands)</th>
<th>Relative efficiency based on</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stratified variance estimate of $\hat{\alpha}<em>{x</em>{t}}$</td>
<td>Short cut variance estimate of $\hat{\alpha}<em>{x</em>{t}}$</td>
</tr>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>$\hat{\alpha}<em>{x</em>{t}}$</td>
<td>1,584.11 (stratified)</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>1,792.50 (short cut)</td>
<td>--</td>
</tr>
<tr>
<td>$\hat{\alpha}<em>{x</em>{t_s}}$</td>
<td>1,246.17</td>
<td>126.91</td>
</tr>
<tr>
<td>$\hat{\alpha}<em>{x</em>{t_c}}$</td>
<td>1,271.43</td>
<td>134.59</td>
</tr>
<tr>
<td>$\hat{\alpha}<em>{x</em>{t}}$</td>
<td>1,322.01</td>
<td>119.82</td>
</tr>
<tr>
<td>$\hat{\alpha}<em>{x</em>{t}}$</td>
<td>1,322.04</td>
<td>119.82</td>
</tr>
</tbody>
</table>

*Estimated variance of $\hat{\alpha}_{x_{t}}$ is used as numerator.

and so on. The estimated variance of $\hat{\alpha}_{x_{t}}$ is obtained in two ways, namely:

$$s_{\alpha_{xt}}^2 = \{47.49 + 119.78 + 66^2.49 + 240.72 + 76.23
+ 361.77 + 71.56\}$$

$$= 1,584 \text{ (column 3, Table 14); and}$$

$$s_{\alpha_{xt}}^2 \text{ (short cut)} = \frac{1}{20} \left\{ 582.4^2 + 504.3^2 + 459.4^2 + 483.1^2
+ 692.1^2 - \frac{2721.3^2}{5}\right\}$$
Both estimators, $s^2_{X^*}$ and $\hat{\sigma}^2_{X^*}$, are unbiased as shown in Section B5, Chapter III. The difference in the value of the estimates is also explained in that section.

The estimated variance of $\hat{X}^*_t$ is

$$s^2_{\hat{X}^*_t} = \left\{ 101.03 + 138.52 + 418.25 + 126.44 + 49.27 + 380.79 \right\} + 40.67$$

$$= 1,248.17 \text{ (Table 14)}.$$ 

and that of $\hat{X}^!_t$ is

$$s^2_{\hat{X}^!_t} = \left\{ 103.68 + 140.33 + 417.00 + 128.37 + 49.27 + 399.11 \right\} + 40.67$$

$$= 1,271.43 \text{ (Table 14)}.$$ 

From these estimates, $\hat{X}^*_t$ and $\hat{X}^!_t$, may be said to be equally efficient.

The variance of $\hat{X}^*_t$ and $\hat{X}^!_t$ are estimated with the use of data in Table 15. Thus,

$$\hat{\sigma}^2_{\hat{X}^*_t} = \hat{\sigma}^2_{\hat{X}^!_t} + \sigma^2_{t} \hat{\sigma}^2_{X^*} - \sigma^2_{t} \sigma^2_{X^!}$$

$$= \left\{ (1,792.50) + (4.37)^2 \cdot (111.84) \right\}$$

$$= 1322.04$$

where

$$\sigma^2_{t} = 4.37,$$

$$\hat{\sigma}^2_{\hat{X}^*_t} = \frac{1}{20} \left\{ 119.2^2 + 109.1^2 + 130.1^2 + 123.9^2 + 139.8^2 \right\} - \frac{522.1}{5}$$

$$= 26.55,$$
\[ \hat{\sigma}_t^2 = 1,799.50 \text{ (short cut),} \]

and

\[
\begin{align*}
\hat{\sigma}_{X_t}^2, \hat{\sigma}_t^2 &= \frac{1}{20} \left\{ (582.4)(119.2) + (504.3)(109.1) \\
&+ (459.4)(130.1) + (483.1)(123.9) \\
&+ (692.1)(139.8) - \frac{(2721.3)(692.1)}{5} \right\} \\
&= 111.84.
\end{align*}
\]

Similarly,

\[
\hat{\sigma}_{X_{tc}}^2 = \hat{\sigma}_t^2 + \alpha_t^2 \hat{\sigma}_{X_t}^2 - 2 \alpha_t \hat{\sigma}_{X_t, X_t}^2
\]

\[= 1322.01 \]

where

\[ \alpha_t^* = 4.38. \]

Again, we may note that these results indicate that \( \hat{X}_{tc}^* \) and \( \hat{X}_{tc}^1 \) are equally efficient since \( \alpha_{tc} = \alpha_t \).

For listed population in hhs at the precinct level, the estimated gain in efficiency with the use of the separate ratio estimator, \( \hat{X}_{ts}^* \) or \( \hat{X}_{ts}^1 \), as compared to the regular unbiased estimator, \( \hat{X}_t^* \), of the sector total \( X_t \), is about 25 per cent while with the use of the combined ratio estimators, \( \hat{X}_{tc}^* \) or \( \hat{X}_{tc}^1 \), the gain is close to 20 per cent. The comparison is based on the stratified variance estimate of \( \hat{X}_t^* \), \( s_{\hat{X}_t}^2 \), since this is the standard variance estimate.

Larger gains are obtained if the comparison is based on the short cut variance estimate of \( \hat{X}_t^* \).
Similar comparisons may be made for the interviewed population in sample hh estimate, \( \hat{X}_t \). The results derived from Table 1 seem to indicate that the reduction in variance with the use of ratio estimators for the sector will be of about the same magnitude as observed for listed population in hh at the precinct level. These empirical studies will be continued by the author and will include the most important characteristics in the survey for all four sectors \((t = 1, 9, 3, 4)\) and for the whole country.

An important point which has to be stressed is the fact that the stratum count, \( Z_{th} \), the sector or regional count, \( Z_t \), and even the country count, \( Z \), of the auxiliary variable are available at practically no cost to the survey. These counts are updated every two years and therefore they may be considered as current and up to date. The feasibility and applicability of these techniques to the rural area design will be studied.
VI. COMPOSITE ESTIMATION

One new feature introduced in the rotation scheme of Chapter IV involves changing a part of the sample each visit to minimize response resistance which may constantly increase when the same panel of sample hhs is interviewed indefinitely. This rotation afforded also a reduction in the costs due to listing the entire sampling area (barrio, poblacion or precinct). The reduction is accomplished primarily through the listing of only a small portion of the sample area.

This new feature of the rotation scheme will now be utilized in the development of an estimation procedure which has been called a composite estimate. The composite estimate is modelled after that developed by Hansen et al. [10]. However, we develop here a finite population theory for such estimators (see section F). In this section, a finite population model consisting of the \( S_{thij} \) rotations is developed where \( S_{thij} \) is the number of segments in the sampling area and this number is usually small as in the PSSH. It will be shown that the variance of the composite estimate at time \( \alpha \) of the barrio total or the poblacion total or the precinct total consists of two parts, namely: the within rotation component and the between rotation component.

The estimate of the stratum total for any characteristic, \( x \), at the \( \alpha \) visit is a composite or weighted average of two estimates. In a matched design, the correlation \( \rho \) from the
identical segment effectively increases the sample size. For the first two consecutive visits, the composite estimate for the second visit is obtained from four segments instead of only three for the regular self weighting unbiased estimator or an increase of 33 per cent in the sample size. This is another advantage derived from the use of composite estimator. With proper weights on each component estimate, the composite estimate may have a smaller variance than either of the component parts. Optimum weights may be determined for estimating totals of any particular characteristic. In processing mass data, however, as in the PSSH, one is left with the choice of using one appropriate weight or one set of weights which will satisfy the condition that for most of the important items in the survey, there will be some gain in precision of the composite estimate over those obtained using the regular unbiased estimate.

A. Estimate of Stratum Total, $\hat{X}_{th}$, from the $ith$ Primary

1. Structure

Consider as composite estimate of the stratum total, $\hat{X}_{th}$, from the $ith$ psu the following:

$$\hat{X}_{th} = Q(\alpha - 1)\hat{X}_{th} + \alpha, \alpha - 1\hat{X}_{th} - \alpha - 1, \alpha\hat{X}_{th} + (1 - Q)\hat{X}_{th}$$

(6.1)

where
\( \alpha - 1 \hat{X}_{\text{thi}} \) is the composite estimate of \( \alpha - 1 X_{\text{th}} \) from the \( i \)th psu at the \((\alpha - 1)\) visit,

\( \hat{X}_{\text{thi}} \) is the regular unbiased estimate of \( X_{\text{th}} \) from the \( i \)th psu based on the three segments at the \( \alpha \) visit (4.20),

\( \alpha, \alpha - 1 \hat{X}_{\text{thi}} \) is the regular unbiased estimate of \( X_{\text{th}} \) from the \( i \)th psu for the \( \alpha \) visit but made from returns of the two matched segments that are included in the sample in both \( \alpha \) and \((\alpha - 1)\) visits,

\( \alpha - 1, \alpha \hat{X}_{\text{thi}} \) is the regular unbiased estimate of \( \alpha - 1 X_{\text{th}} \) from the \( i \)th psu for the \((\alpha - 1)\) visit but made from returns of the two matched segments that are included in the sample in both \((\alpha - 1)\) and \( \alpha \) visits,

and

\( 0 \leq Q \leq 1. \)

The composite estimate in (6.1) consists of two parts. The first with a weight, \( Q \), consists of a composite estimate of \( \alpha - 1 X_{\text{th}} \) plus an estimate of the change from \((\alpha - 1)\) to the \( \alpha \) visit, and the second with a weight, \((1 - Q)\), is also an estimate of \( \alpha X_{\text{th}} \), the stratum total at the \( \alpha \) visit. Combining the two components with their proper weights give the composite estimate of \( \alpha X_{\text{th}} \). The composite estimate in (6.1) is of the same form as used by the U. S. Bureau of the Census (2.1).

From the pattern on the \( \alpha X_{\text{thi}}, (6.1) \) may be written as
\[ \hat{x}_{\text{thi}} = Q \hat{x}_{\text{thi}} + \alpha V_{\text{thi}} \]

or

\[ \hat{x}_{\text{thi}} = Q \alpha V_{\text{thi}} + Q \alpha - 1 V_{\text{thi}} + \ldots + Q \alpha - l V_{\text{thi}} + \alpha V_{\text{thi}} \]

\[ = \sum_{v=0}^{\infty} Q^v \alpha - v V_{\text{thi}} \tag{6.2} \]

where

\[ V_{\text{thi}} = Q((\alpha, \alpha - 1) \hat{x}_{\text{thi}} - (\alpha - l, \alpha) \hat{x}_{\text{thi}}) + (1 - Q) \hat{x}_{\text{thi}} \tag{6.2a} \]

Equation (6.11) is of the same form as the one given by Hansen et al. [14].

2. Variance

From (6.2), \( \text{Var}(\hat{x}_{\text{thi}}) \), may be expressed in terms of the variances and covariances of the \( \alpha - v V_{\text{thi}} \). The variance of \( \alpha V_{\text{thi}} \), is obtained from the relation given in (6.2a).

Thus,

\[ \text{Var}(\hat{x}_{\text{thi}}) = \sum_{v=0}^{\infty} \left\{ Q^v \text{Var}(\alpha - v V_{\text{thi}}) \right. \]

\[ + 2Q^{1+2v} \text{Cov}(\alpha - v V_{\text{thi}}, \alpha - v - 1 V_{\text{thi}}) \]

\[ + 2Q^{2+2v} \text{Cov}(\alpha - v V_{\text{thi}}, \alpha - v - 2 V_{\text{thi}}) \]

\[ + 2Q^{3+2v} \text{Cov}(\alpha - v V_{\text{thi}}, \alpha - v - 3 V_{\text{thi}}) + \ldots \right\}, \tag{6.3} \]

or
\[ \text{Var}(\hat{X}_{\text{thi}}) = \sum_{v=0}^{\infty} Q^v \text{Var}(\alpha - v V_{\text{thi}}) + 2 \sum_{v=0}^{\infty} Q^v \sum_{r=1}^{\infty} Q^r \text{Cov}(\alpha - v V_{\text{thi}}, \alpha - v - r V_{\text{thi}}). \] (6.3a)

For large \( \alpha \), let
\[ \sum_{v=0}^{\infty} Q^v = \frac{1}{1 - Q^\beta} \]
and with the use of the following assumptions,
\[ \text{Var}(\alpha - V_{\text{thi}}) = \text{Var}(\beta V_{\text{thi}}), \] (6.3b)
and
\[ \text{Cov}(\alpha - v V_{\text{thi}}, \alpha - v - r V_{\text{thi}}) = \text{Cov}(\beta V_{\text{thi}}, \beta - r V_{\text{thi}}) \] (6.3c)
for all \( v \) and \( r \), (6.3) becomes
\[ \text{Var}(\hat{X}_{\text{thi}}) = \frac{\text{Var}(\beta V_{\text{thi}})}{1 - Q^\beta} + \frac{2}{1 - Q^\beta} \sum_{r=1}^{\infty} Q^r \text{Cov}(\beta V_{\text{thi}}, \beta - r V_{\text{thi}}). \] (6.4)

We need to evaluate each term on the right hand side of (6.4).

To evaluate \( \text{Var}(\beta V_{\text{thi}}) \), we use the relation given in (6.2a). Thus
\[ \text{Var}(\beta V_{\text{thi}}) = Q^2 \left\{ \text{Var}(\beta, \alpha - 1) \hat{X}_{\text{thi}} + \text{Var}(\alpha - 1, \alpha) \hat{X}_{\text{thi}} \right. \\
- 2 \text{Cov}(\beta, \alpha - 1) \hat{X}_{\text{thi}}(\alpha - 1, \alpha) \hat{X}_{\text{thi}} \left. \right\} + (1 - Q^2) \text{Var}(\hat{X}_{\text{thi}}) \]
\[ + 2Q(1 - Q) \text{Cov}(\alpha, \alpha - 1) \hat{X}_{th1}, \alpha \hat{X}_{th1} \]
\[ - 2Q(1 - Q) \text{Cov}(\alpha - 1, \alpha) \hat{X}_{th1}, \alpha \hat{X}_{th1} \]. \quad (6.5)

To simplify (6.5), we define the components in
\[ \alpha X_{th1} = Q\{\alpha, \alpha - 1 \hat{X}_{th1} - \alpha - 1, \alpha \hat{X}_{th1} \} + (1 - Q) \alpha \hat{X}_{th1} \]
in terms of estimates of stratum total, \( \alpha X_{th} \), arising from each segment from each of the two sample barrios. Thus,
\[ \alpha \hat{X}_{th1} = \frac{1}{3} \left\{ \alpha \hat{X}_{th11} + \alpha \hat{X}_{th12} + \alpha \hat{X}_{th13} \right\} , \quad (6.5a) \]
\[ \alpha, \alpha - 1 \hat{X}_{th1} = \frac{1}{2} \left\{ \alpha \hat{X}_{th12} + \alpha \hat{X}_{th13} \right\} \quad (6.5b) \]
and
\[ \alpha - 1, \alpha \hat{X}_{th1} = \frac{1}{2} \left\{ \alpha - 1 \hat{X}_{th12} + \alpha - 1 \hat{X}_{th13} \right\} \quad (6.5c) \]
where
\[ \alpha \hat{X}_{th11} = \frac{m}{3} \left\{ 3(\alpha \hat{X}_{th11} + \alpha \hat{X}_{th12}) \right\} \quad (6.5d) \]
\[ \alpha \hat{X}_{th12} = \frac{m}{3} \left\{ 3(\alpha \hat{X}_{th12} + \alpha \hat{X}_{th13}) \right\} \quad (6.5e) \]
\[ \alpha \hat{X}_{th13} = \frac{m}{3} \left\{ 3(\alpha \hat{X}_{th13} + \alpha \hat{X}_{th14}) \right\} \quad (6.5f) \]

Similar definitions are given for \( \alpha \hat{X}_{th1} \) and \( \alpha \hat{X}_{th13} \).

Note that the first segment (in for the first visit) is unmatched while the second (in for the second visit) and third (in for the third visit) segments are matched. If
\[ \text{Var}(\alpha - v \hat{X}_{th1k}) = \text{Var}(\alpha \hat{X}_{th1k}) = \sigma_k^2 \quad (6.6) \]
for all \( \alpha, v \) and \( k \), where \( \sigma_k^2 \) is the variance of estimate, \( \alpha X_{th1k} \), of stratum total \( \alpha X_{th} \). This is derived from an
individual sample segment from each of the two sample barrios, then the following variances and covariances are derived:

\[
\text{Var}(\hat{X}_{\alpha}^\text{thi}) = \frac{1}{3} \sigma_k^2, \quad (6.6a)
\]

\[
\text{Var}(\alpha, \alpha_{-1}^\text{thi}) = \text{Var}(\alpha_{-1}, \alpha^\text{thi}) = \frac{1}{9} \sigma_k^2, \quad (6.6b)
\]

\[
\text{Cov}(\alpha, \alpha_{-1}^\text{thi}, \alpha^\text{thi}) = \frac{1}{3} \sigma_k^2, \quad (6.6c)
\]

\[
\text{Cov}(\alpha_{-1}, \alpha^\text{thi}, \alpha^\text{thi}) = \frac{1}{3} \rho \sigma_k^2, \quad (6.6d)
\]

\[
\text{Cov}(\alpha_{-2}, \alpha_{-1}^\text{thi}, \alpha_{-1}^\text{thi}) = \frac{1}{9} \rho \sigma_k^2, \quad (6.6e)
\]

and

\[
\text{Cov}(\alpha_{-2}, \alpha_{-1}^\text{thi}, \alpha, \alpha_{-1}^\text{thi}) = \frac{1}{4} \rho \sigma_k^2, \quad (6.6f)
\]

where

\[
\rho \text{ is the correlation one visit apart } (\alpha_{-1}, \alpha) \text{ of estimates from identical segments}
\]

and

\[
\rho \text{ is the correlation two visits apart } (\alpha_{-2}, \alpha) \text{ of estimates from identical segments.}
\]

With the use of (6.6) to (6.6f) in (6.5), (6.5) becomes

\[
\text{Var}(\alpha^\text{V\text{thi}}) = \frac{1}{3} \left\{ 1 + 2Q^2 - (Q + Q^2) \rho \right\} \sigma_k^2. \quad (6.7)
\]

To evaluate \(\text{Cov}(\alpha^\text{V\text{thi}}, \alpha_{-1}^\text{V\text{thi}})\), we use the relation given in (6.2a),

where

\[
\alpha_{-1}^\text{V\text{thi}} = Q(\alpha_{-1}, \alpha_{-2}) \hat{X}^\text{thi} - (\alpha_{-2}, \alpha_{-1}) \hat{X}^\text{thi} \nonumber
\]

\[
+ (1 - Q) \alpha_{-1} \hat{X}^\text{thi},
\]
and
\[ \alpha^v_{\text{thi}} = Q(\alpha, \alpha - 1)^{\hat{x}}_{\text{thi}} - (\alpha - 1, \alpha)^{\hat{x}}_{\text{thi}} + (1 - Q) \alpha^{\hat{x}}_{\text{thi}}, \]

with the relations given in (6.6) to (6.8f), the structure of the correlations is derived and the results are presented in Table 17.

From Table 17, we have
\[ \text{Cov}(\alpha^v_{\text{thi}}, \alpha - 1^v_{\text{thi}}) = \left\{ -\frac{1}{3}(Q - \frac{1}{4}Q^2) + \frac{1}{9}(2 + \frac{1}{2}Q + 2Q^2) \rho_1 - \frac{1}{6}(Q + \frac{1}{2}Q^2) \rho_2 \right\} \sigma_k^2 \]

(6.8)

for all \( \alpha \). By a similar method used to obtain (6.8), the \( \text{Cov}(\alpha^v_{\text{thi}}, \alpha - 2^v_{\text{thi}}) \)

may be derived with the use of Table 18. Thus,
\[ \text{Cov}(\alpha^v_{\text{thi}}, \alpha - 2^v_{\text{thi}}) = \left\{ -\frac{1}{6}Q(1 - Q) \rho_1 + \frac{1}{9}(1 - \frac{1}{2}Q - \frac{1}{2}Q^2) \rho_2 \right\} \sigma_k^2 \]

(6.9)

for all \( \alpha \).

From the nature of the rotation scheme,
\[ \text{Cov}(\alpha^v_{\text{thi}}, \alpha - 3^v_{\text{thi}}) = 0 \]

(6.10)

and in general,
\[ \text{Cov}(\alpha^v_{\text{thi}}, \alpha - v^v_{\text{thi}}) = 0 \]

(6.11)

for
\[ 3 \leq v \leq \alpha. \]

Substituting (6.7), (6.8), (6.9), (6.10) and (6.11) in (6.4) and simplifying, we get
Table 17. Structure of the correlations for $\alpha^{-1}V_{\text{thi}}$ and $V_{\text{thi}}$

<table>
<thead>
<tr>
<th>$(\alpha-1)V_{\text{thi}}$</th>
<th>$Q(\alpha-1, \alpha^{-1})\tilde{X}_{\text{thi}}$</th>
<th>$-Q(\alpha^{-1}, \alpha-1)\tilde{X}_{\text{thi}}$</th>
<th>$(1 - Q)(\alpha-1)\tilde{X}_{\text{thi}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q(\alpha, \alpha^{-1})\tilde{X}_{\text{thi}}$</td>
<td>$\frac{1}{4} Q^2 f_1$</td>
<td>$-\frac{1}{4} Q^2 f_1$</td>
<td>$\frac{1}{3} Q(1 - Q) f_1$</td>
</tr>
<tr>
<td>$-Q(\alpha^{-1}, \alpha)\tilde{X}_{\text{thi}}$</td>
<td>$-\frac{1}{4} Q^2$</td>
<td>$\frac{1}{4} Q^2 f_1$</td>
<td>$-\frac{1}{3} Q(1 - Q)$</td>
</tr>
<tr>
<td>$(1 - Q)\tilde{X}_{\text{thi}}$</td>
<td>$\frac{1}{6} Q(1 - Q) f_1$</td>
<td>$-\frac{1}{6} Q(1 - Q) f_1$</td>
<td>$\frac{2}{3}(1 - Q)^2 f_1$</td>
</tr>
</tbody>
</table>
Table 18. Structure of correlations for $\alpha^{-2V_{thi}}$ and $V_{thi}$

<table>
<thead>
<tr>
<th>$\alpha^{-2V_{thi}}$</th>
<th>$Q(\alpha^{-2}, \alpha^{-3})\hat{X}_{thi}$</th>
<th>$-Q(\alpha^{-3}, \alpha^{-?})\hat{X}_{thi}$</th>
<th>$(1 - Q)(\alpha^{-?})\hat{X}_{thi}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha^{V_{thi}}$</td>
<td>0</td>
<td>0</td>
<td>$\frac{1}{6}Q(1 - Q)\int_0^\alpha$</td>
</tr>
<tr>
<td>$Q(\alpha, \alpha^{-1})\hat{X}_{thi}$</td>
<td>0</td>
<td>0</td>
<td>$-\frac{1}{6}Q(1 - Q)\int_1^\alpha$</td>
</tr>
<tr>
<td>$-Q(\alpha^{-1}, \alpha)\hat{X}_{thi}$</td>
<td>0</td>
<td>0</td>
<td>$\frac{1}{6}(1 - Q)^\alpha\int_0^\alpha$</td>
</tr>
<tr>
<td>$(1 - Q)\alpha\hat{X}_{thi}$</td>
<td>0</td>
<td>0</td>
<td>$\frac{1}{6}(1 - Q)^\alpha\int_0^\alpha$</td>
</tr>
</tbody>
</table>
\[ \text{Var}(\tilde{\alpha}_{\text{th}i}) = \frac{1}{1 - Q^2} \left\{ \frac{1}{9}(1 + \frac{Q^3}{2}) \right\} \]
\[ - \frac{1}{9}(2Q^2 + 2Q^3 - Q^3 - 3Q^4) \sigma_k^2 \]
\[ - \frac{1}{9}(Q^2 + \frac{5}{2} Q^3 + Q^4) \sigma_k^2 \}\]  
\( \sigma_k^2 \).

This variance (6.12) is compared to the \( \text{Var}(\tilde{\alpha}_{\text{th}i}) \) which is equal to \( \frac{1}{3} \sigma_k^2 \) for various assumed values of \( \rho_1 \) and \( \rho_2 \) and various values of \( Q \). These comparisons are shown in Table 19. A general discussion on the choice of an 'optimum' \( Q \) will be given in a separate section.

Table 19. Ratio of the variance of \( \tilde{\alpha}_{\text{th}i} \) to the variance of \( \tilde{\alpha}_{\text{th}i} \) for various assumed values of \( \rho_1 \) and \( \rho_2 \) and some values of \( Q \)

<table>
<thead>
<tr>
<th>( \rho_1 )</th>
<th>( \rho_2 )</th>
<th>Values of ( Q )</th>
<th>0.1</th>
<th>0.3</th>
<th>0.5</th>
<th>0.7</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>0.8</td>
<td></td>
<td>1.0628</td>
<td>1.2258</td>
<td>1.3957</td>
<td>0.7026</td>
<td>*</td>
</tr>
<tr>
<td>0.8</td>
<td>0.6</td>
<td></td>
<td>1.0549</td>
<td>1.1778</td>
<td>1.2158</td>
<td>0.6081</td>
<td>*</td>
</tr>
<tr>
<td>0.7</td>
<td>0.5</td>
<td></td>
<td>1.0443</td>
<td>1.1283</td>
<td>1.0979</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>0.6</td>
<td>0.4</td>
<td></td>
<td>1.0364</td>
<td>1.0916</td>
<td>1.0084</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>0.5</td>
<td>0.3</td>
<td></td>
<td>1.0291</td>
<td>1.0584</td>
<td>0.9397</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>0.4</td>
<td>0.2</td>
<td></td>
<td>1.0287</td>
<td>1.0181</td>
<td>0.8678</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>0.3</td>
<td>0.1</td>
<td></td>
<td>1.0121</td>
<td>0.9882</td>
<td>0.8109</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

*Values in these cells are less than 0.6.
B. Estimate of Visit-to-Visit Change, \( X_{th} - X_{-1th} \), from the \( ith \) Primary

1. Structure

To estimate the visit to visit change, \( X_{th} - X_{-1th} \), with the use of the \( ith \) psu, consider

\[
D_{thi} = \hat{X}_{thi} - \hat{X}_{-1thi}
\]

(6.13)

which is the difference of the composite estimates at the \( \alpha \) and \((\alpha-1)\) visits. From (6.1) and (6.2a), (6.13) becomes

\[
D_{thi} = Q \left\{ (\alpha-1)\hat{X}_{thi} - (\alpha-\alpha)\hat{X}_{thi} \right\} + (\hat{V}_{thi} - (\alpha-1)\hat{V}_{thi})
\]

(6.13a)

Letting

\[
\alpha-1D_{thi} = \alpha-1\hat{X}_{thi} - \alpha \hat{X}_{thi}
\]

(6.13b)

and

\[
\alpha Z'_{thi} = \alpha \hat{V}_{thi} - \alpha-1 \hat{V}_{thi}
\]

(6.13c)

we have from (6.13a),

\[
D_{thi} = Q \alpha-1D'_{thi} + \alpha Z'_{thi}
\]

(6.14)

which is similar in form to (6.9). We can, therefore, write (6.14) as

\[
D_{thi} = \sum_{v=0}^{\alpha} Q^v \alpha-\alpha v Z'_{thi}
\]

(6.14a)

and its variance is
\[ \text{Var}(\alpha_{\text{D}1}) = \frac{\text{Var}(\alpha'_{\text{thi}})}{1 - Q^2} \]
\[ + \frac{\sigma^2}{1 - Q^2} \sum_{r=1}^{\infty} Q^r \text{Cov}(\alpha'_{\text{thi}}, \alpha'_{-1\text{thi}}). \]

(6.15)

2. Variance

The variance of \( \alpha'_{\text{thi}} \) is evaluated by term by term evaluation of the variances and covariances of the \( \alpha'_{\text{thi}} \)'s. But the \( \alpha'_{\text{thi}} \)'s are functions of the \( \alpha_{\text{thi}} \)'s and we can utilize the results given in Tables 17 and 18.

Using (6.13c), (6.7), and (6.8), the first term in (6.15) becomes

\[ \text{Var}(\alpha'_{\text{thi}}) = \left\{ \frac{1}{3}(Q + 2Q + \frac{2}{3}Q^2) + \frac{1}{9}(4 + 13Q + 10Q^2) \right\} \sigma_k^2. \]

(6.16)

Similarly,

\[ \text{Cov}(\alpha'_{\text{thi}}, \alpha'_{-1\text{thi}}) = \left\{ -\frac{1}{3}(1 - 2Q - \frac{3}{2}Q^2) \right\} \sigma_k^2; \]

(6.17)

\[ \text{Cov}(\alpha'_{\text{thi}}, \alpha'_{-2\text{thi}}) = \left\{ \frac{1}{3}(Q - \frac{1}{4}Q^2) - \frac{1}{9}(4 + 2Q - Q^2) \right\} \sigma_k^2; \]

(6.18)

\[ \text{Cov}(\alpha'_{\text{thi}}, \alpha'_{-3\text{thi}}) = \left\{ \frac{1}{8}(1 - Q) - \frac{1}{9}(1 - \frac{1}{2}Q - \frac{1}{2}Q^2) \right\} \sigma_k^2. \]

(6.19)

and
Cov( $\alpha'^{\text{th} \text{i}}, \alpha - \nu^{\text{th} \text{i}}$) = 0 \quad (6.20)

for

$4 \leq \nu \leq \alpha$.

Using equations (6.16), (6.17), (6.18), (6.19), and (6.20) in (6.15) and simplifying, we have

$$
\text{Var}(\alpha'^{\text{th} \text{i}}) = \frac{2}{1 - \rho^2} \left\{ \frac{1}{3}(1 - \frac{1}{4}\rho^2 - Q^3 - \frac{1}{4}\rho^4) 
- \frac{1}{18}(4 + 5\rho^2 - 4\rho^3 - 5\rho^4 + 3\rho^5) \rho_1 
+ \frac{1}{18}(\rho + \frac{1}{2}\rho^2 - 3\rho^3 - \frac{1}{2}\rho^4 + \rho^5) \rho_2 \right\} \sigma_k^2. \quad (6.21)
$$

This variance is compared to the variance of

$$
\hat{\alpha'^{\text{th} \text{i}}} = \hat{\alpha'^{\text{th} \text{i}}} - \alpha - 1 \hat{\alpha'^{\text{th} \text{i}}} \quad (6.22)
$$

which is the difference between the regular unbiased estimator at the $\alpha$ visit and at the $\alpha - 1$ visit. The variance of $\hat{\alpha'^{\text{th} \text{i}}}$ is

$$
\text{Var}(\hat{\alpha'^{\text{th} \text{i}}}) = \frac{2}{3}(1 - \frac{2}{3}\rho_1) \sigma_k^2. \quad (6.22a)
$$

The comparisons are given in Table 20, for various assumed values of $\rho_1$ and $\rho_2$ and various values of $\rho$. The choice of an 'appropriate' $\rho$ will be discussed in a later section.

If the variance of $\alpha'^{\text{th} \text{i}}$ is compared to the variance of

$$
\hat{\alpha'^{\text{th} \text{i}}} = \alpha - \alpha - 1 \hat{\alpha'^{\text{th} \text{i}}} - \alpha - 1, \alpha \hat{\alpha'^{\text{th} \text{i}}} \quad (6.22b)
$$

which is equal to

$$
\text{Var}(\hat{\alpha'^{\text{th} \text{i}}}) = (1 - \rho_1^2) \sigma_k^2 \quad (6.22c)
$$

then some changes will be made in the values given in Table
Table 20. Ratio of the variance of $\alpha_{\text{Dhi}}$ to the variance of $\alpha_{\text{Dhi}}$ for various assumed values of $\rho_1$ and $\rho_2$ and some values of $Q$.

<table>
<thead>
<tr>
<th>$\rho_1$</th>
<th>$\rho_2$</th>
<th>0.1</th>
<th>0.3</th>
<th>0.5</th>
<th>0.7</th>
<th>0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>0.8</td>
<td>1.1655</td>
<td>1.6182</td>
<td>3.8131</td>
<td>7.3665</td>
<td>*</td>
</tr>
<tr>
<td>0.8</td>
<td>0.6</td>
<td>1.0246</td>
<td>1.3330</td>
<td>2.0469</td>
<td>3.2659</td>
<td>*</td>
</tr>
<tr>
<td>0.7</td>
<td>0.5</td>
<td>1.0944</td>
<td>1.2805</td>
<td>1.6903</td>
<td>2.4956</td>
<td>*</td>
</tr>
<tr>
<td>0.6</td>
<td>0.4</td>
<td>1.0872</td>
<td>1.1820</td>
<td>1.4325</td>
<td>1.9835</td>
<td>*</td>
</tr>
<tr>
<td>0.5</td>
<td>0.3</td>
<td>1.0461</td>
<td>1.1420</td>
<td>1.2711</td>
<td>1.7930</td>
<td>*</td>
</tr>
<tr>
<td>0.4</td>
<td>0.2</td>
<td>1.0366</td>
<td>1.0906</td>
<td>1.1746</td>
<td>1.5383</td>
<td>9.7090</td>
</tr>
<tr>
<td>0.3</td>
<td>0.1</td>
<td>0.8116</td>
<td>1.0431</td>
<td>1.0907</td>
<td>1.3800</td>
<td>3.6364</td>
</tr>
</tbody>
</table>

*Value of $Q = 0.9$ is omitted since the coefficients in $\text{Var}(\alpha_{\text{Dhi}})$ are negative.

\textsuperscript{a}Coefficient in $\text{Var}(\alpha_{\text{Dhi}})$ are negative.

20. These changes will depend on

a) If $\rho_1 = \frac{9}{15}$, then $\text{Var}(\alpha_{\text{Dhi}}) = \text{Var}(\alpha_{\text{Dhi}})$ and the values in the table remain the same;
b) If $\rho_1 < \frac{9}{15}$, then $\text{Var}(\alpha_{\text{Dhi}}) < \text{Var}(\alpha_{\text{Dhi}})$ and the values in the table will increase; and
c) If $\rho_1 > \frac{9}{15}$, then $\text{Var}(\alpha_{\text{Dhi}}) > \text{Var}(\alpha_{\text{Dhi}})$ and the values in the table will decrease.
C. Estimate of Visit-to-Visit a Year Ago Change, \( X_{th} - X_{th-2} \), from the ith Primary

1. Structure

To estimate the visit-to-visit a year ago change, \( X_{th} - X_{th-2} \), let us consider,

\[
D_{thi} = X_{thi} - X_{thi-2} \quad (6.23)
\]

From (6.1), (6.23) becomes

\[
D_{thi} = \sum_i (X_{thi} - X_{thi-2}) + (V_{thi} - V_{thi-2}) \quad (6.23a)
\]

Let

\[
D_{thi} = \sum_i (X_{thi} - X_{thi-2}) \quad (6.23b)
\]

and

\[
Z_{thi} = V_{thi} - V_{thi-2} \quad (6.23c)
\]

then we have for (6.23),

\[
D_{thi} = \sum_i (X_{thi} - X_{thi-2}) + V_{thi} - V_{thi-2} \quad (6.23d)
\]

which is of the same form as (6.2) and, therefore, we can write it as

\[
D_{thi} = \sum_{v=0}^{\infty} Q^v Z_{thi} \quad (6.23e)
\]

where

\( Z_{thi} \) are functions of \( V_{thi} \) and \( V_{thi-2} \).
2. Variance

The variance of \( \alpha_{D_{th1}} \) is given by

\[
\text{Var}(\alpha_{D_{th1}}) = \frac{\text{Var}(\alpha_{Z_{th1}})}{1 - \sigma^2} + \frac{\sigma^2}{1 - \sigma^2} \\
\sum_{r=1}^{\infty} \sigma^r \text{Cov}(\alpha_{Z_{th1}}, \alpha_r Z_{th1}). \tag{6.24}
\]

The variances and covariances of the \( \alpha_{Z_{th1}} \)'s are evaluated in the same manner as the corresponding variances and covariances of the \( \alpha_{Z_{th1}} \)'s. Thus,

\[
\text{Var}(\alpha_{D_{th1}}) = \frac{2}{1 - \sigma^2} \left\{ \frac{1}{3}(1 + \frac{1}{4} \xi^3 - \xi^4 - \frac{1}{4} \xi^5) \\
- \frac{1}{18}(5\xi + 8\xi^2 - 6\xi^3 - 11\xi^4 + \xi^5 + 3\xi^6) \right\} \rho_1 \\
- \frac{1}{18}(2 - 5\xi - 2\xi^2 - 2\xi^3 + \xi^4 - 5\xi^5 - \xi^6) \right\} \rho_2 \right\} \sigma_k^2. \tag{6.25}
\]

This variance may be compared to the variance of

\[
\hat{\alpha}_{D_{th1}} = \hat{\alpha}_{Z_{th1}} - \alpha \hat{\alpha}_{Z_{th1}} \tag{6.26}
\]

which is

\[
\text{Var}(\hat{\alpha}_{D_{th1}}) = \frac{2}{3}(1 - \frac{\rho_1}{3}) \sigma_k^2. \tag{6.26a}
\]

The gain in precision in the use of \( \alpha_{D_{th1}} \) to that of \( \hat{\alpha}_{D_{th1}} \) for various assumed values of \( \rho_1 \) and \( \rho_2 \) and various values of \( \xi \) is given in Table 21.

If the variance of \( \alpha_{D_{th1}} \) is compared to the variance of

\[
\hat{\alpha}_{D_{th1}} = \alpha, \alpha \hat{\alpha}_{Z_{th1}} - \alpha \hat{\alpha}, \alpha \hat{\alpha}_{Z_{th1}}
\]
Table 21. Ratio of the variance of $\hat{\alpha}_{D,th1}$ to the variance of $\alpha_{D,th1}^*$ for various values of $\rho_1$ and $\rho_2$ and some values of $Q$.

<table>
<thead>
<tr>
<th>$\rho_1$</th>
<th>$\rho_2$</th>
<th>0.1</th>
<th>0.3</th>
<th>0.5</th>
<th>0.7</th>
<th>0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>0.8</td>
<td>1.0848</td>
<td>1.3542</td>
<td>1.7393</td>
<td>1.7568</td>
<td>1.3994</td>
</tr>
<tr>
<td>0.8</td>
<td>0.6</td>
<td>1.0744</td>
<td>1.2539</td>
<td>1.4293</td>
<td>1.3516</td>
<td>1.1589</td>
</tr>
<tr>
<td>0.7</td>
<td>0.5</td>
<td>1.0613</td>
<td>1.1230</td>
<td>1.2852</td>
<td>1.1747</td>
<td>1.0362</td>
</tr>
<tr>
<td>0.6</td>
<td>0.4</td>
<td>1.0508</td>
<td>1.1383</td>
<td>1.1351</td>
<td>1.0383</td>
<td>0.9230</td>
</tr>
<tr>
<td>0.5</td>
<td>0.3</td>
<td>1.0372</td>
<td>1.0866</td>
<td>1.0651</td>
<td>0.9294</td>
<td>0.8427</td>
</tr>
<tr>
<td>0.4</td>
<td>0.2</td>
<td>1.0208</td>
<td>1.0397</td>
<td>0.9949</td>
<td>0.8465</td>
<td>0.7788</td>
</tr>
<tr>
<td>0.3</td>
<td>0.1</td>
<td>1.0181</td>
<td>1.0082</td>
<td>0.9975</td>
<td>0.7849</td>
<td>0.7253</td>
</tr>
</tbody>
</table>

*aAll ratios for $Q = 0.9$ are less than 0.6.

which is equal to

$$\text{Var}(\hat{\alpha}_{D,th1}) = \rho(1 - \rho)\sigma_k^2,$$

then changes in the values given in Table 21 will be made depending on the following:

a) If $\rho_2 = \frac{3}{4}$, then $\text{Var}(\hat{\alpha}_{D,th1}) = \text{Var}(\hat{\alpha}_{D,th1}^*)$ and the values in the table remain the same;

b) If $\rho_2 < \frac{3}{4}$, then $\text{Var}(\hat{\alpha}_{D,th1}) < \text{Var}(\hat{\alpha}_{D,th1}^*)$ and the values in the table will increase; and

c) If $\rho_2 > \frac{3}{4}$, then $\text{Var}(\hat{\alpha}_{D,th1}) > \text{Var}(\hat{\alpha}_{D,th1}^*)$ and the values in the table will decrease.
D. Sector Estimate

To estimate the sector total, $\alpha X_t$, we need to evaluate the estimator of the stratum total, $\alpha X_{th}$, from the five ordered psus or mean of the five psus. Similarly, we have to evaluate the stratum visit-to-visit change and visit-to-visit a year ago change estimates from the five ordered psus. To do this, we invoke the results of the theory presented in Chapters III and IV by identifying $\hat{\alpha} X_{thi}$, $\hat{\alpha} D_{thi}$ with $\hat{\alpha} X_{thi}$. The results will be summarized for the mean of the five psu estimates.

The stratum estimates are as follows:

a) Current total or level, $\alpha X_{th}$

Estimator: $\hat{\alpha} X_{th} = \frac{1}{m} \sum_{i=1}^{5} \hat{\alpha} X_{thi}$ \hspace{1cm} (6.27)

where $\hat{\alpha} X_{thi}$ is as defined in (6.2) and $m = 5$.

Estimate of variance: To estimate $\text{Var}(\hat{\alpha} X_{th}) = \frac{\text{Var}(\hat{\alpha} X_{thi})}{m}$, we use

$$\alpha \hat{\sigma}_{th}^2 = \frac{1}{m} \frac{1}{m-1} \sum_{i=1}^{5} (\hat{\alpha} X_{thi} - \hat{\alpha} X_{th})^2.$$ \hspace{1cm} (6.27a)

b) Visit-to-visit change, $\alpha X_{th} - \alpha X_{th-1}$

Estimator: $\hat{\alpha} D_{th} = \frac{1}{m} \sum_{i=1}^{5} \hat{\alpha} D_{thi}$ \hspace{1cm} (6.28)

where $\hat{\alpha} D_{thi}$ is as defined in (6.13).
Estimate of variance: To estimate $\text{Var}(\alpha D'_{th}) = \frac{\text{Var}(\alpha D'_{th})}{m}$, we use
\[
\sigma^2_{th} = \frac{1}{m} \frac{1}{m-1} \sum_{i=1}^{m-1} (\alpha D'_{th} - \alpha D'_{th})^2 . \quad (6.28a)
\]

c) Visit-to-visit a year ago change, $\alpha X_{th} - \alpha - \sigma X_{th}$

Estimator: $\alpha D''_{th} = \frac{1}{m} \sum_{i=1}^{m} \alpha D''_{th}$ \quad (6.29)

where $\alpha D''_{th}$ is as defined in (6.93).

Estimate of variance: To estimate $\text{Var}(\alpha D''_{th}) = \frac{\text{Var}(\alpha D''_{th})}{m}$, we use
\[
\sigma^2_{th} = \frac{1}{m} \frac{1}{m-1} \sum_{i=1}^{m-1} (\alpha D''_{th} - \alpha D''_{th})^2 . \quad (6.29a)
\]

The estimates for the sector are as follows:

a) Current total or level, $\alpha X_t$

Estimator: $\hat{\alpha} X_t = \frac{1}{m} \sum_{i=1}^{m} \hat{\alpha} X_{ti}$ \quad (6.30)

where $\hat{\alpha} X_{ti} = \sum_{h} \hat{\alpha} X_{thi}$.

Estimate of variance: To estimate $\text{Var}(\hat{\alpha} X_t) = \frac{\text{Var}(\hat{\alpha} X_{ti})}{m}$, we use
\[
\hat{\alpha}^2_{t} = \frac{1}{m} \frac{1}{m-1} \sum_{i=1}^{m-1} (\hat{\alpha} X_{ti} - \hat{\alpha} X_{t})^2 . \quad (6.30a)
\]
b) Visit-to-visit change, $\alpha X_t - \alpha - 1 X_t$

Estimator: $\alpha D_t^i = \frac{1}{m} \sum_1^m \alpha D_{ti}$

where $\alpha D_{ti} = \sum_h \alpha D_{thi}$.

Estimate of variance: To estimate $\text{Var}(\alpha D_t^i) = \frac{\text{Var}(D_{ti})}{m}$, we use

$$\sigma_t^2 = \frac{1}{m} \frac{1}{m-1} \sum_1^m (\alpha D_{ti} - \alpha D_t)^2.$$ (6.31a)

c) Visit-to-visit a year ago change, $\alpha X_t - \alpha - 2 X_t$

Estimator: $\alpha D_t^{ii} = \frac{1}{m} \sum_1^m \alpha D_{ti}$

where $\alpha D_{ti} = \sum_h \alpha D_{thi}$.

Estimate of variance: To estimate $\text{Var}(\alpha D_t^{ii}) = \frac{\text{Var}(D_{ti})}{m}$, we use

$$\sigma_t^{ii} = \frac{1}{m} \frac{1}{m-1} \sum_1^m (\alpha D_{ti} - \alpha D_t^{ii})^2.$$ (6.32a)

With the use of the above techniques, estimates for the whole rural area ($t = 1$ and $t = 2$) may be obtained. If there exists a systematic correlation between the primary numbers $i = 1, i = 2, \ldots, i = 5$, then the short cut method of estimating the variances, $\sigma_t^2$, $\sigma_t^1$, and $\sigma_t^{ii}$ is an over estimate and as such the standard or 'stratified' method should be
used.

We can also apply the ratio estimation techniques developed in Chapter V to the \( i \)th PSU composite estimate, \( X_{thi} \) (6.1) for the strata and the \( i \)th PSU composite estimate, \( X_{ti} \) (6.30), for the sector. With these correspondences, four ratio estimators may be derived for the sector:

a) 'separate biased ratio of means estimate' which is similar to (5.33),

b) 'separate unbiased ratio estimate' (5.63),

c) 'combined biased ratio of means estimate' (5.38), and
d) 'combined unbiased ratio estimate' (5.66).

This approach is simple since the \( Z \) variable is presently available by precinct, by stratum and by sector. In this case, we have composite estimation first and then ratio estimation.

E. Choice of an Appropriate \( Q \)

In a general purpose survey like the PSSH where hundreds of items are processed, the practical approach is to choose a general \( Q \) which will result in a reduction of variance for most if not all of the important items in the survey rather than finding an optimum \( Q \) for each of the characteristics covered in the survey. This approach becomes more advantageous when we consider the fact that we are interested in using three types of estimates, namely: current total esti-
mate, estimate of visit-to-visit change and estimate of visit-to-visit a year ago change. Each estimator and its variance are functions of \( q \). The choice of \( q \) will, therefore, depend on the reduction of variance for most items and also for the three types of estimates.

The U. S. Bureau of the Census in its current population survey (CPS) uses a value of \( \frac{1}{6} \) for \( q \) for the within PSU variance of composite estimate. This choice of \( q \) results in smaller variances of estimates of both level and change for most characteristics, but larger gains are achieved, for most part, in the estimates of change. Experience of various authors \([7, 10]\) seem to indicate that for most items,

\[
\int \varrho^2 \leq \int \varrho^1,
\]

i.e., the magnitude of the correlation decreases as the length of the visits increases. In the PSSH, \( \int \varrho^1 \) is the correlation six months apart of estimates from identical segments and \( \int \varrho^2 \) is the correlation 12 months or one year apart of estimates from identical segments. The correlation six months apart of estimates from identical precincts (PSU) of population in interviewed HHS, \( \hat{X}_{ithj} \), for the first seven strata in urban Manila is estimated at about 0.9 and the correlation 12 months apart of estimates from identical precincts (PSU) is estimated at about 0.8. These correlations are expected to be high since the estimates are for population counts from identical precincts (PSU).
From the results in Tables 19, 20 and 21, one may obtain an idea on the appropriate choice of $Q$ for estimating current total, visit-to-visit change and visit-to-visit a year ago change. For $C_1 \geq 0.5$ and corresponding $C_2$ in the tables, fairly large gains in precision are obtained for the three types of composite estimates for $0.3 \leq Q \leq 0.5$. Empirical studies on the magnitude of $C_1$ and $C_2$ of estimates for the most important items in the survey will be made in order to indicate the choice of an appropriate $Q$ for use in the PSSH.

F. Finite Population Approach

After segmentation of the sample barrio in a given municipality, the segments are arranged at random into a rotation pattern. The number of segments, $S_{thij}$, in the sample barrio is usually small,

$$4 \leq S_{thij} \leq 10.$$  

Consider a particular rotation (R) for $S_{thij} = 5$ which is given in Figure 2 for some $\alpha$. Our main interest is the estimation of the th$^{th}$ barrio total at time $0, o^X_{thij}$. Since $S_{thij}$ is small and with

$$-\infty < \alpha \leq 0$$

our finite population will consist of the $S_{thij}$! possible R's from which we have drawn a sample of size one. In our example,

$$S_{thij} = 5$$
<table>
<thead>
<tr>
<th>Segment number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time, $\alpha$</td>
<td>-8</td>
<td>-7</td>
<td>-6</td>
<td>-5</td>
<td>-4</td>
</tr>
<tr>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

Figure 2. A particular rotation pattern for $S_{thij} = 5$

and our finite population is the $5! \ R'$s. The sample $R$ may have the pattern given in Figure 2. The usual method in the estimation of $X_{thij}$ is to draw the 3 segment out of the $S_{thij}$ segments at time 0 and consider the finite population consisting of the

$$\binom{S_{thij}}{3}$$

possible samples, and for $S_{thij} = 5$ this is

$$\binom{5}{3}$$

possible samples of size 3. In the present model, the sample consists of one sample $R$ from the possible $5! \ R'$s.
1. Structure of estimate

Assume that the same sample of hhs is observed in a given segment for any time \( \alpha \) and that there are no 'drop-outs' or additions or replacements of households. Consider as composite estimate of the thijth barrio total at time 0, \( 0^{X}_{thij} \)

\[
0^{X}_{thij} = \sum_{\alpha=0}^{\infty} \sum_{k} W_{thijk} \hat{X}_{thijk} \quad (6.33)
\]

where

\( \hat{X}_{thijk} = \frac{N_{thijk}}{n_{thijk}} \) \( X_{thijk} \) is the sample total estimate of the thijkth segment at visit \( \alpha \),

\( N_{thijk} \) is the number of hhs in the thijth segment and is independent of \( \alpha \),

\( n_{thijk} \) is the number of sample hhs in the thijth segment and is independent of \( \alpha \),

\( X_{thijk} \) is the sample total of the thijkth segment at time \( \alpha \),

\( W_{thijk} \) is the weight attached to the \( \hat{X}_{thijk} \) at time \( \alpha \),

and

\( S_{thij} = 5 \).

One substitutes \( S_{thij} \) for 5 in the following derivations in order to express the results for a general \( S_{thij} \) where \( S_{thij} \) is small. The weights, \( W_{thijk} \), in (6.33) will be determined
from the relation

\[ \hat{X}^{thij} = Q\{-1, \hat{X}^{thij} + 0, -1, \hat{X}^{thij} - 1, 0\hat{X}^{thij}\} + (1 - Q) 0\hat{X}^{thij} \]  

(6.34)

where

- \( \hat{X}^{thij} \) is the composite estimate of the \( thij \)th barrio total at time \(-1\), \(-1X^{thij}\),

\[ 0, -1\hat{X}^{thij} = \frac{5}{2} (0\hat{X}^{thij} + 0\hat{X}^{thij}) \]  

(6.35a)

is the regular unbiased estimate of \( 0X^{thij} \) based on the two matched segments for times 0 and \(-1\),

\[ -1, 0\hat{X}^{thij} = \frac{5}{2} (-1\hat{X}^{thij} + -1\hat{X}^{thij}) \]  

(6.35b)

is the regular unbiased estimate of \(-1X^{thij} \) based on the two matched segments for times \(-1\) and 0,

\[ 0\hat{X}^{thij} = \frac{5}{3} (0\hat{X}^{thij} + 0\hat{X}^{thij} + 0\hat{X}^{thij}) \]  

(6.36)

is the regular unbiased estimate of \( 0X^{thij} \) based on the three segments at time 0,

and

\[ 0 \leq Q \leq 1. \]

Using equation (6.34), \( \hat{X}^{thij} \) may be expressed in terms of \( -2\hat{X}^{thij}, -1, -2\hat{X}^{thij}, -2, -1\hat{X}^{thij} \) and \(-1\hat{X}^{thij}\). In general, (6.34) may be written as

\[ 0\hat{X}^{thij} = \sum_{\alpha = 0}^{-\infty} Q^{-\alpha} \alpha\hat{X}^{thijk} \]  

(6.37)

where

\[ \alpha\hat{X}^{thijk} = Q\{\alpha, \alpha, -1\hat{X}^{thijk} - \alpha, -1, \alpha\hat{X}^{thijk}\} + (1 - Q) \alpha\hat{X}^{thijk}. \]
For any particular $R$ and for a given time $\alpha$, the weight of a segment total estimate, $\hat{x}_{thijk}$, will depend on whether the segment is in the sample for the first visit, the second visit, the third visit, off one visit or off two visits, etc. These weights are shown in Table 22 for $S_{thij} = 5$ and for $\alpha = 0, \alpha = -1, \alpha = -2, \alpha = -3,$ and $\alpha = -4$. From the pattern of the weights in Table 22, we can say that for a given time $\alpha < 0$, the weights are

\[ -\frac{5}{3} Q^{-\alpha} (Q + \frac{1}{2}) \]  

for a segment in the sample for the first visit;

\[ \frac{5}{6} Q^{-\alpha} (Q - 1) \]  

for the second visit;

\[ \frac{5}{3} Q^{-\alpha} (\frac{Q}{2} + 1) \]  

for the third visit; and

\[ 0 \]  

for a segment off one visit or off two visits.

The expectation of $\hat{x}_{thij}$ is

\[ E(\hat{x}_{thij}) = E_{5!R}\{E(\hat{x}_{thij} \mid R)\} \]  

where

\[ E(\hat{x}_{thij} \mid R) \]  

is the conditional expectation of $\hat{x}_{thij}$ given a particular $R$,

and

\[ E_{5!R} \]  

is the expectation over all $5! R$'s.
Table 22. Weights of segment total estimates depending on whether a segment is in the sample for the first visit, second visit, third visit, third visit last time, third visit second to last time for $\alpha_{thijk} = 5$ and for $\alpha = 0, -1, -2, -3, -4$ for a particular rotation

<table>
<thead>
<tr>
<th>Segment in sample for $\alpha$</th>
<th>$\alpha = 0$</th>
<th>$\alpha = -1$</th>
<th>$\alpha = -2$</th>
<th>$\alpha = -3$</th>
<th>$\alpha = -4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st visit</td>
<td>$-\frac{5}{3}(\omega - 1)$</td>
<td>$-\frac{5}{3}(\omega + \frac{1}{2})$</td>
<td>$-\frac{5}{3}(\omega + \frac{1}{2})$</td>
<td>$-\frac{5}{3}(\omega + \frac{1}{2})$</td>
<td>$-\frac{5}{3}(\omega + \frac{1}{2})$</td>
</tr>
<tr>
<td>2nd visit</td>
<td>$\frac{5}{3}(\omega + 1)$</td>
<td>$\frac{5}{3}(\omega - 1)$</td>
<td>$\frac{5}{3}(\omega - 1)$</td>
<td>$\frac{5}{3}(\omega - 1)$</td>
<td>$\frac{5}{3}(\omega - 1)$</td>
</tr>
<tr>
<td>3rd visit</td>
<td>$\frac{5}{3}(\omega + 1)$</td>
<td>$\frac{5}{3}(\omega + 1)$</td>
<td>$\frac{5}{3}(\omega + 1)$</td>
<td>$\frac{5}{3}(\omega + 1)$</td>
<td>$\frac{5}{3}(\omega + 1)$</td>
</tr>
<tr>
<td>Off 1 visit</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Off 2 visits</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\frac{1}{5} \sum_{k}^{5} \alpha_{thijk}$</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\[ \frac{1}{5} \sum_{k}^{5} \alpha_{thijk} \leq 1 \]
To evaluate (6.39), we have

\[ \begin{align*}
E & \left\{ 5! \left[ \sum_{\alpha = 0}^{\infty} \sum_{k} \alpha W_{thijk} \alpha X_{thijk} \right] \right. \\
& = E \left\{ \sum_{\alpha = 0}^{\infty} \sum_{k} \alpha W_{thijk} E(\alpha X_{thijk}) \right. \\
& = E \left\{ \sum_{\alpha = 0}^{\infty} \sum_{k} \alpha W_{thijk} X_{thijk} \right. \\
& = (6.40)
\end{align*} \]

where

\( X_{thijk} \) is the thijkth segment total at time \( \alpha \).

Finally,

\[ \begin{align*}
E & \left\{ \sum_{\alpha = 0}^{\infty} \sum_{k} \alpha W_{thijk} X_{thijk} \right. \\
& = \sum_{\alpha = 0}^{\infty} \sum_{k} E(\alpha W_{thijk}) X_{thijk} \\
& = \sum_{k} 0^{X_{thijk}} \\
& = 0^{X_{thij}} \\
& = (6.41)
\end{align*} \]

where

\( 0^{X_{thij}} \) is the thijth barrio total at time 0,

\[ E(\alpha W_{thijk}) = 1 \] for \( \alpha = 0 \), and

\[ E(\alpha W_{thijk}) = 0 \] for \( \alpha < 0 \).
This result shows that \( \hat{\alpha}^{thij} \) is unbiased. The expectation of the \( \alpha^{thijk} \) over all \( 5! \) \( \text{R}'s \) is obtained from the fact that each of the five segment totals, \( \alpha^{thijk} \), appears for a given \( \alpha \) in the five categories in Table 22 an equal number of times which is \( 4! \). Thus the expected weight of \( \alpha^{thijk} \) is obtained from the relation,

\[
\sum_{k=1}^{5} \frac{4! \alpha^{thijk}}{5!} = \sum_{k=1}^{5} \frac{\hat{\alpha}^{thijk}}{5}
\]  

which is given at the bottom of Table 22 for some \( \alpha \). Note that the summation on \( k \) is taken over the five categories into which the five segments fall on a given time \( \alpha \). In general, the number of categories is equal to the number of segments, \( S_{thij} \).

2. Variance of composite estimate

The variance of \( \hat{\alpha}^{thij} \) is given by the relation,

\[
\text{Var}(\hat{\alpha}^{thij}) = \text{E}_R \left\{ \text{Var}(\hat{\alpha}^{thij} | \text{R}) \right\} + \text{Var}_R \left\{ \text{E}_R(\hat{\alpha}^{thij} | \text{R}) \right\} .
\]  

(6.45)

We evaluate the terms in (6.45) separately. The first term is

\[
\text{E}_R \left\{ \text{Var}(\hat{\alpha}^{thij} | \text{R}) \right\} = \text{E}_R \left\{ \sum_{\alpha=0}^{\infty} \sum_{k} \alpha^{thijk} \text{Var}(\hat{\alpha}^{thijk}) \right\}
\]

\[
+ \sum_{\alpha \neq \alpha'} \sum_{k} \alpha^{thijk} \alpha'^{thijk} \text{Cov}(\hat{\alpha}^{thijk}, \hat{\alpha}'^{thijk}) \}
\]  

(6.46)
where

\[ \text{Var}(\hat{\chi}_{\text{thijk}} | R) = \sum_{\alpha=0}^{\infty} \sum_{k=1}^{5} W_{\text{thijk}}^\alpha \text{Var}(\hat{\chi}_{\text{thijk}}) \]

\[ + \sum_{\alpha \neq \alpha'} \sum_{k=1}^{5} W_{\text{thijk}}^\alpha W_{\text{thijk}}^\alpha' \text{Cov}(\hat{\chi}_{\text{thijk}}, \hat{\chi}_{\text{thijk}}'), \]

(6.46a)

\[ \text{Var}(\hat{\chi}_{\text{thijk}}) = \frac{N_{\text{thijk}}^2}{n_{\text{thijk}}} (1 - \frac{n_{\text{thijk}}}{N_{\text{thijk}}}) \alpha_{\text{thijk}}^2, \]

(6.46b)

\[ S_{\text{thijk}}^2 = \sum_{l} \left( \frac{\alpha_{\text{thijk}}^l}{N_{\text{thijk}}} - \frac{\alpha_{\text{thijk}}^0}{N_{\text{thijk}}} \right)^2, \]

\[ \alpha_{\text{thijk}}^0 = \frac{\sum_{l} \alpha_{\text{thijk}}^l}{N_{\text{thijk}}}, \]

(6.46c)

\[ \text{Cov}(\hat{\chi}_{\text{thijk}}, \hat{\chi}_{\text{thijk}}') = \frac{N_{\text{thijk}}^2}{n_{\text{thijk}}} (1 - \frac{n_{\text{thijk}}}{N_{\text{thijk}}}) \alpha_{\text{thijk}} \alpha_{\text{thijk}}', \]

and

\[ \alpha_{\text{thijk}}', \alpha_{\text{thijk}}^{\prime S} = \sum_{l} \left( \alpha_{\text{thijk}}^l - \alpha_{\text{thijk}}^0 \right) \left( \alpha_{\text{thijk}}^{l'} - \alpha_{\text{thijk}}^{0'} \right) \frac{N_{\text{thijk}}^2}{n_{\text{thijk}}} (1 - \frac{n_{\text{thijk}}}{N_{\text{thijk}}}). \]

The values of \( W_{\text{thijk}}^0 \) for \( \alpha = 0 \) and for \( \alpha < 0 \) for a given \( R \) are given in Table 9. From this table, we can evaluate

\[ E(\alpha_{\text{thijk}}^0) = \frac{\alpha}{3} (3 + \frac{3\alpha^0}{2}) \]

(6.47)

for \( \alpha = 0 \); and
Table P-3. Values of $\omega_{\text{thijk}}$ for $S_{\text{thij}} = 5$, $\alpha = 0$ and $\alpha < 0$ for a given rotation (R)

<table>
<thead>
<tr>
<th>Segment in sample for</th>
<th>$\alpha = 0$</th>
<th>$\alpha &lt; 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st visit</td>
<td>$\left(\frac{5}{3}\right)^2 \cdot \left(1 - \frac{Q}{2}\right)^2$</td>
<td>$\left(-\frac{5}{3}\right)^2 \cdot Q^{-\alpha} \cdot \left(Q + \frac{1}{2}\right)^5$</td>
</tr>
<tr>
<td>2nd visit</td>
<td>$\left(\frac{5}{3}\right)^2 \cdot \left(1 + \frac{Q}{2}\right)^2$</td>
<td>$\left(\frac{5}{3}\right)^2 \cdot Q^{-\alpha} \cdot \left(Q - 1\right)^5$</td>
</tr>
<tr>
<td>3rd visit</td>
<td>$\left(\frac{5}{3}\right)^2 \cdot \left(1 + \frac{Q}{2}\right)^2$</td>
<td>$\left(\frac{5}{3}\right)^2 \cdot Q^{-\alpha} \cdot \left(\frac{Q}{2} + 1\right)^5$</td>
</tr>
<tr>
<td>Off 1 visit</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Off 2 visits</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\[
\frac{1}{5} \sum_{k}^{5} \omega_{\text{thijk}}^2 \cdot \frac{5}{3^2} \cdot \frac{3 + 3 \cdot Q^2}{3} \cdot \frac{5}{3} \cdot Q^{-\alpha} \cdot \frac{3}{2} \cdot \left(Q^2 + Q + 1\right)
\]

\[
E \left( \omega_{\text{thijk}}^2 \right) = \frac{5}{3^2} \cdot Q^{-\alpha} \cdot \frac{3}{2} \cdot \left(Q^2 + Q + 1\right) \quad (6.48)
\]

for $\alpha < 0$.

With (6.47) and (6.48) in the first term of (6.46), we have

\[
E \left\{ \sum_{k=0}^{\infty} \sum_{k=0}^{5} \omega_{\text{thijk}}^2 \cdot \text{Var}(\hat{x}_{\text{thijk}}) \right\} = \ldots
\]
Equation (6.49) may be simplified if we assume
\[ \text{var}(\hat{X}_{\text{thijk}}) = \text{var}(\hat{X}_{\text{thijk}}) \] (6.49a)
for all \( \alpha \), and let
\[ \sum_{\alpha=-1}^{\infty} Q^{-2\alpha} = \frac{Q^{2}}{1 - Q^{2}}. \]

After some simplification, (6.49) finally becomes
\[
\frac{5}{3^{2}} \frac{1}{1 - Q^{2}} (3 + \frac{3Q^{2}}{5}) \sum_{k} \text{var}(\hat{X}_{\text{thijk}}). \] (6.50)
\[
E \left\{ \sum_{\alpha \neq \alpha'} \sum_{k} \alpha^{W} \text{thijk} \cdot \alpha'^{W} \text{thijk} \right\} \text{Cov}( \alpha^{\hat{\text{thijk}}}, \alpha'^{\hat{\text{thijk}}} )
\]

\[
= \sum_{\alpha = 0}^{\infty} \sum_{k} E \left( \alpha^{W} \text{thijk} \cdot \alpha - 1^{W} \text{thijk} \right) \text{Cov}( \alpha^{\hat{\text{thijk}}}, \alpha - 1^{\hat{\text{thijk}}} )
\]

\[
+ 2 \sum_{\alpha = 0}^{\infty} \sum_{k} E \left( \alpha^{W} \text{thijk} \cdot \alpha - \rho^{W} \text{thijk} \right) \text{Cov}( \alpha^{\hat{\text{thijk}}}, \alpha - \rho^{\hat{\text{thijk}}} )
\]

where

\[
\text{Cov}( \alpha^{\hat{\text{thijk}}}, \alpha - v^{\hat{\text{thijk}}} ) = 0
\]

for \( v \geq 3 \).

To evaluate

\[
E \left( \alpha^{W} \text{thijk} \cdot \alpha'^{W} \text{thijk} \right)
\]

for \( \alpha \neq \alpha' \), we use the results given in Table 22 and the weights given in (6.38a) to (6.38d). Thus,

\[
E \left( 0^{W} \text{thijk} \cdot -1^{W} \text{thijk} \right) = \frac{1}{5} \left\{ \left( \frac{5}{3} \right) \left( \frac{Q}{3} + 1 \right) \left( -\frac{5}{3} \right) Q (Q + \frac{1}{2}) + \left( \frac{5}{3} \right) \left( \frac{Q}{3} + 1 \right) \left( \frac{5}{6} \right) Q (Q - 1) \right\}
\]

(6.52a)

\[
E \left( 0^{W} \text{thijk} \cdot -\rho^{W} \text{thijk} \right) = \frac{1}{5} \left\{ \left( \frac{5}{3} \right) \left( \frac{Q}{3} + 1 \right) \left( -\frac{5}{3} \right) Q (Q + \frac{1}{2}) \right\}
\]

(6.52b)

\[
E \left( \alpha^{W} \text{thijk} \cdot -1^{W} \text{thijk} \right) = \frac{1}{5} \left\{ \left( \frac{5}{6} \right) Q^{-\alpha} (Q - 1) \left( -\frac{5}{3} \right) Q^{-\alpha + 1} (Q + \frac{1}{2}) \right\}
\]

(6.52c)

+ \left( \frac{5}{3} \right) Q^{-\alpha} (Q + 1) \left( \frac{5}{6} \right) Q^{-\alpha + 1} (Q - 1) \}
\]
Substitute (6.52a), (6.52b), (6.52c) and (6.52d) into (6.51) and assume that
\[
\text{Cov}(\hat{X}_{thijk}, \hat{X}^r_{thijk}) = \int_1 \text{Var}(\hat{X}^r_{thijk}),
\]
\[
\text{Cov}(\hat{X}^r_{thijk}, \hat{X}^r_{thijk}) = \int_2 \text{Var}(\hat{X}^r_{thijk}),
\]
for all \( \alpha \neq 0 \), where
\( \int_1 \) is the correlation of estimates from identical segments one visit apart,

and
\( \int_2 \) is the correlation of estimates from identical segments two visits apart,

then after some algebra (6.51) becomes

\[
\mathbb{E} \left\{ \sum_{\alpha \neq \alpha'} \sum_{k=1}^{5} \alpha^{W_{thijk}} \alpha'^{W_{thijk}} \text{Cov}(\hat{X}_{thijk}, \hat{X}^r_{thijk}) \right\} =
-2 \left( \frac{5}{3^2} \right) \frac{1}{1 - Q^2} \left\{ (Q + Q^2 - \frac{1}{2}Q^3 - \frac{3}{2}Q^4) \int_1 + \frac{1}{2}(Q^2 + \frac{5}{2}Q^3 + Q^4) \int_2 \right\} \sum_{k=5}^{5} \text{Var}(\hat{X}^r_{thijk}).
\] (6.53)

With (6.50) and (6.53) in (6.46), we have
\[
\mathbb{E} \left\{ \text{Var}(\hat{X}^r_{thijk} \mid R) \right\} = \frac{1}{1 - Q^2} \frac{5}{3^2} \left\{ (3 + 3Q^3) - 2 \left[ (Q + Q^2 - \frac{1}{2}Q^3 - \frac{3}{2}Q^4) \int_1 \right] - \left[ (Q^2 + \frac{5}{2}Q^3 + Q^4) \int_2 \right] \right\} \sum_{k=5}^{5} \text{Var}(\hat{X}^r_{thijk}).
\] (6.54)
This is the within rotation variance and the leading term is

\[
\frac{5}{3^2} 3 \sum_{k} \text{Var}(\hat{X}_{\text{thijk}}) = \frac{5}{3} \sum_{k} \frac{N_{\text{thijk}}}{n_{\text{thijk}}} (1 - \frac{n_{\text{thijk}}}{N_{\text{thijk}}}) \sigma^2_{\text{thijk}}
\]

(6.56)

which is the within segment component for a two-stage sampling with equal probability and without replacement at each stage at time 0. For small \(S_{\text{thij}}\), the within segment component is usually the more important.

In the U. S. Bureau of the Census publication, the within psu-variance of the composite estimate is similar to that of equation (6.4) or (6.12). The between-psu variance is the same for the composite estimate as for the regular ratio estimate. The reduction in variance through the use of the composite estimate is entirely in the within-psu contribution to the variance. For many items the within-psu variance is the principal component to the variance of month-to-month change, the gains on estimates of change are worthwhile.

The variance \(\sigma^2_k\) in (6.12) denotes the variance of the straight estimator of the stratum total, \(X_{\text{th}}\), from an individual segment. The within psu variance of this estimator will in general have a between segment and within segment component. Thus,

\[
\sigma^2_k = \sigma^2_b + \sigma^2_w
\]

We shall now identify
\[
\frac{\sigma^2}{B^2} \text{ with } 5 \sum_{k}^{5} \text{Var}(\hat{X}_{\text{thijk}})
\]
in equation (6.54) where

\( B \) is the general raising factor for the stratum or is
the number of barrios in the stratum.

It will be seen that the factors applied to these two quanti-
ties agree.

The second term in (6.45) is
\[
\text{Var}\left\{E(\hat{X}_{\text{thijk}} | R)\right\} = E\left\{E(\hat{X}_{\text{thijk}} | R)\right\}^2 \nonumber
\]
\[= \left\{E \left[E(\hat{X}_{\text{thijk}} | R)\right]\right\}^2. \quad (6.56)
\]
Note that the second term on the right hand side of (6.56) is
evaluated with the use of (6.41). Thus
\[
\left\{E \left[E(\hat{X}_{\text{thijk}} | R)\right]\right\}^2 = \left[X_{\text{thijk}}\right]^2. \quad (6.57)
\]
The first term in (6.56) becomes,
\[
E\left\{\sum_{\alpha=0}^{\infty} \sum_{k}^{5} W_{\text{thijk}} X_{\text{thijk}}\right\}^2 = E\left\{\sum_{\alpha=0}^{\infty} \sum_{k}^{5} W_{\text{thijk}}^2 X_{\text{thijk}}^2\right\} \nonumber
\]
\[+ \sum_{\alpha \neq \alpha'}^{\infty} \sum_{k}^{5} W_{\text{thijk}} W_{\text{thijk}}' X_{\text{thijk}} X_{\text{thijk}}' \nonumber
\]
\[+ \sum_{\alpha=0}^{\infty} \sum_{k \neq k'}^{5} W_{\text{thijk}} W_{\text{thijk}}' X_{\text{thijk}} X_{\text{thijk}}' \nonumber
\]
\[+ \sum_{\alpha \neq \alpha'}^{\infty} \sum_{k \neq k'}^{5} W_{\text{thijk}} W_{\text{thijk}}' X_{\text{thijk}} X_{\text{thijk}}' \right\}. \quad (6.58)
\]
The four terms on the right hand side of (6.58) are evaluated separately. The first term is

$$E \left\{ \sum_{\alpha=-1}^{\infty} \sum_{k} W_{\text{thijk}}^{\alpha} X_{\text{thijk}}^{\alpha} \right\} = \sum_{\alpha=0}^{\infty} \sum_{k} E \left( W_{\text{thijk}}^{\alpha} X_{\text{thijk}}^{\alpha} \right)$$

(6.59)

With the results from (6.49), (6.59) becomes

$$E \left\{ \sum_{\alpha=0}^{\infty} \sum_{k} W_{\text{thijk}}^{\alpha} X_{\text{thijk}}^{\alpha} \right\} = \frac{5}{3} \left\{ \sum_{k} (3 + \frac{3}{2} \alpha) X_{\text{thijk}}^{\alpha} \right\}$$

$$+ \sum_{\alpha=-1}^{\infty} \sum_{k} \frac{3}{2} \alpha^{-\alpha} (\alpha + \alpha + 1) X_{\text{thijk}}^{\alpha}$$

(6.59a)

where the infinite summation is truncated to include terms involving $-3 \leq \alpha \leq 0$. The second term in (6.58) is

$$E \left\{ \sum_{\alpha \neq \alpha'} \sum_{k} W_{\text{thijk}}^{\alpha} W_{\text{thijk}}^{\alpha'} X_{\text{thijk}}^{\alpha} X_{\text{thijk}}^{\alpha'} \right\}$$

$$= \sum_{\alpha \neq \alpha'} \sum_{k} E \left( W_{\text{thijk}}^{\alpha} W_{\text{thijk}}^{\alpha'} X_{\text{thijk}}^{\alpha} X_{\text{thijk}}^{\alpha'} \right)$$

(6.60)

Equation (6.60) is approximated by truncating the infinite summation. Thus,

$$E \left\{ \sum_{\alpha \neq \alpha'} \sum_{k} W_{\text{thijk}}^{\alpha} W_{\text{thijk}}^{\alpha'} X_{\text{thijk}}^{\alpha} X_{\text{thijk}}^{\alpha'} \right\} =$$

$$- \frac{5}{2} \left\{ \sum_{k} (\frac{3}{2} \alpha^2 + \alpha \alpha + \alpha) X_{\text{thijk}}^{\alpha} - X_{\text{thijk}}^{\alpha} \right\}$$

(6.60a)
\[ + \sum_{k}^{5} (\alpha^{5} + \alpha^{4} - \gamma \alpha^{3})X_{\text{thijk}} - 3X_{\text{thijk}} \]

\[ + \sum_{\alpha = -1}^{-2} \sum_{k}^{5} \xi_{\alpha}^{-2} \left( \frac{\alpha^{3}}{2} - \frac{3}{2} \right) X_{\text{thijk}} \alpha^{-1}X_{\text{thijk}} \]

\[ + \sum_{\alpha = 0}^{-1} \sum_{k}^{5} \xi_{\alpha}^{-2} \left( \frac{\alpha^{4}}{2} + \frac{5\alpha}{2} + \frac{\alpha^{3}}{2} \right) X_{\text{thijk}} \alpha^{-2}X_{\text{thijk}} \} \]

\[ (\xi.60a) \]

The third term in (6.59) is

\[ \mathbb{E}_{5! \infty} \left\{ \sum_{\alpha = 0}^{-\infty} \sum_{k \neq k'} \xi_{\text{thijk}} \xi_{\text{thijk'}} \xi_{\text{thijk}} \xi_{\text{thijk'}} \right\} \]

\[ = \sum_{\alpha = 0}^{-\infty} \sum_{k \neq k'} \mathbb{E}_{5! \infty} \left( \xi_{\text{thijk}} \xi_{\text{thijk'}} \right) \xi_{\text{thijk}} \xi_{\text{thijk'}} \cdot \]

\[ (\xi.61) \]

Now

\[ \mathbb{E}_{5! \infty} \left( 0 \xi_{\text{thijk}} 0 \xi_{\text{thijk'}} \right) = \frac{1}{2} \xi \frac{5}{3} \left\{ 3 - \frac{3}{4} \xi \right\} \]

\[ (\xi.61a) \]

and

\[ \mathbb{E}_{5! \infty} \left( \xi \xi_{\text{thijk}} \xi \xi_{\text{thijk'}} \right) = - \xi^{-2} \xi \frac{5}{3} \left\{ \frac{3}{4} \xi \xi + \xi + 1 \right\} \]

\[ \alpha < 0 \]

\[ (\xi.61b) \]

With (6.61a) and (6.61b) in (6.61) and truncating the infinite sum, (6.61) is approximated by
\[ E \left\{ \sum_{\alpha=0}^{\infty} \sum_{k=k'}^{5} \alpha W_{\text{thijk}} \alpha W_{\text{thijk'}} \alpha X_{\text{thijk}} \alpha X_{\text{thijk'}} \right\} = \]

\[ \frac{1}{3} \sum_{k=k'}^{5} \left( 3 - \frac{3\alpha}{4} \right) X_{\text{thijk}} X_{\text{thijk'}} \]

\[ = \frac{1}{3} \sum_{k=k'}^{5} \sum_{\alpha=-1}^{5} \frac{3\alpha}{4} (\alpha + \zeta + 1) X_{\text{thijk}} X_{\text{thijk'}}. \]

The fourth term in (6.58) is

\[ E \left\{ \sum_{\alpha \neq \alpha'}^{\infty} \sum_{k=k'}^{5} \alpha W_{\text{thijk}} \alpha W_{\text{thijk'}} \alpha X_{\text{thijk}} \alpha X_{\text{thijk'}} \right\} = \]

\[ \sum_{\alpha \neq \alpha'}^{\infty} \sum_{k=k'}^{5} E \left( \sum_{\alpha=0}^{\infty} \sum_{k=k'}^{5} \alpha W_{\text{thijk}} \alpha W_{\text{thijk'}} \alpha X_{\text{thijk}} \alpha X_{\text{thijk'}} \right). \]

Equation (6.63) is simplified by truncating the infinite sum. Thus, (6.63) is approximated by

\[ \sum_{\alpha \neq \alpha'}^{\infty} \sum_{k=k'}^{5} E \left( \sum_{\alpha=0}^{\infty} \sum_{k=k'}^{5} \alpha W_{\text{thijk}} \alpha W_{\text{thijk'}} \alpha X_{\text{thijk}} \alpha X_{\text{thijk'}} \right) = \]

\[ \sum_{\alpha=0}^{\infty} \sum_{k=k'}^{5} E \left( \alpha W_{\text{thijk}} \alpha W_{\text{thijk'}} \alpha X_{\text{thijk}} \alpha X_{\text{thijk'}} \right) \]

\[ + \sum_{\alpha=0}^{\infty} \sum_{k=k'}^{5} E \left( \alpha W_{\text{thijk}} \alpha W_{\text{thijk'}} \alpha X_{\text{thijk}} \alpha X_{\text{thijk'}} \right) \]

\[ + \sum_{k=k'}^{5} E \left( 0 W_{\text{thijk}} -3 W_{\text{thijk'}} \right) X_{\text{thijk}} -3 X_{\text{thijk'}}. \]

(6.64)
To evaluate (6.64), the following expectations are derived:

\[ E_{5!R} \left( \alpha \bar{W} \theta_{ijkl} - \bar{W} \theta_{ijkl}' \right) = \frac{1}{2} \frac{5}{3} \left( \frac{Q^3}{4} + Q^2 + Q \right), \quad (6.64a) \]

\[ E_{5!R} \left( \alpha^2 \bar{W} \theta_{ijkl} - \bar{W} \theta_{ijkl}' \right) = \frac{1}{2} \frac{5}{3} \left( \frac{Q^4}{4} + \frac{5Q^3}{2} + \frac{Q^2}{2} \right), \quad (6.64b) \]

\[ E_{5!R} \left( \alpha \bar{W} \theta_{ijkl} - 3 \bar{W} \theta_{ijkl}' \right) = \frac{1}{2} \frac{5}{3} \left( \frac{Q^5}{4} + \frac{Q^4}{2} - Q^3 \right), \quad (6.64c) \]

\[ E_{5!R} \left( \alpha^2 \bar{W} \theta_{ijkl} - 3 \bar{W} \theta_{ijkl}' \right) = \frac{1}{2} \frac{5}{3} \left( \frac{Q^6}{4} + \frac{5Q^5}{4} + \frac{Q^4}{2} \right), \quad (6.64d) \]

\[ E_{5!R} \left( \alpha^2 \bar{W} \theta_{ijkl} - 3 \bar{W} \theta_{ijkl}' \right) = \frac{1}{2} \frac{5}{3} \left( \frac{Q^7}{4} + \frac{5Q^6}{4} + \frac{Q^5}{2} \right), \quad (6.64e) \]

\[ E_{5!R} \left( \alpha^2 \bar{W} \theta_{ijkl} - 3 \bar{W} \theta_{ijkl}' \right) = \frac{1}{2} \frac{5}{3} \left( \frac{Q^8}{4} + \frac{5Q^7}{4} + \frac{Q^6}{2} \right). \quad (6.64f) \]

With (6.64a) to (6.64f) in (6.64), (6.64) is approximated by

\[
E_{5!R} \left\{ \sum_{\alpha \neq k} \sum_{k \neq k'} \alpha \bar{W} \theta_{ijkl} - \bar{W} \theta_{ijkl}' \right\} + \sum_{k \neq k'} \left\{ \left[ \left( \frac{Q^3}{4} + Q^2 + Q \right) \alpha \bar{X} \theta_{ijkl} - \bar{X} \theta_{ijkl}' \right] + \left[ \left( \frac{Q^4}{4} + \frac{5Q^3}{2} + \frac{Q^2}{2} \right) \alpha^2 \bar{X} \theta_{ijkl} - \bar{X} \theta_{ijkl}' \right] + \left[ \left( \frac{Q^5}{4} + \frac{Q^4}{2} - Q^3 \right) \alpha \bar{X} \theta_{ijkl} - 3 \bar{X} \theta_{ijkl}' \right] + \left[ \left( \frac{Q^6}{4} + \frac{5Q^5}{4} + \frac{Q^4}{2} \right) \alpha \bar{X} \theta_{ijkl} - 3 \bar{X} \theta_{ijkl}' \right] + \left[ \left( \frac{Q^7}{4} + \frac{5Q^6}{4} + \frac{Q^5}{2} \right) \alpha \bar{X} \theta_{ijkl} - 3 \bar{X} \theta_{ijkl}' \right] \right\}.
\]
With (6.57), (6.59a), (6.60a), (6.62), (6.65) in (6.56)
and after some simplification (6.58) finally becomes,

\[
\text{Var}
\left[
\frac{1 \sum_{i}^{5} X_{\text{thij}}}{R}
\right]
\geq \frac{5}{3}
\left[
\begin{array}{c}
\frac{2}{5} + \frac{3}{5} Q_0^2
\end{array}
\right] S_{\text{thij}}^2
\]

+ \frac{5}{3}
\sum_{\alpha = -3}^{-1}
(\frac{5}{2}) Q_0^{-2\alpha} (Q_0^2 + Q_0 + 1) S_{\text{thij}}^2

- \frac{5}{3}
\sum_{\alpha = 0}^{-1}
\frac{5}{3} Q_0^3 + \frac{4}{3} Q_0^2 + \frac{4}{3} Q_0^2

- \frac{5}{3}
\sum_{\alpha = -2}^{-1}
(\frac{5}{3} Q_0^3 - \frac{2}{3} Q_0^2 + \frac{3}{3})

\begin{align*}
\text{where} & \\
\alpha_{S_{\text{thij}}}^2 & = \frac{1}{4}
\bigg[
\sum_{k} X_{\text{thijk}}^2 - \frac{\alpha_{S_{\text{thij}}}^2}{5}
\bigg], \\
\alpha, \alpha_{-v_{S_{\text{thij}}}} & = \frac{1}{4}
\bigg[
\sum_{k} X_{\text{thijk}} \alpha_{-v_{X_{\text{thijk}}}} - \frac{\alpha_{S_{\text{thij}}} \alpha_{-v_{X_{\text{thijk}}}}}{5}
\bigg] \\
\text{and} & \\
\alpha_{X_{\text{thijk}}} & = \sum_{k} X_{\text{thijk}}.
\end{align*}

At \( \alpha = 0 \), equation (6.66) becomes

\[
\frac{5}{3} \frac{2}{5} \frac{3}{5} \alpha_{S_{\text{thij}}}^2
\]
which is the between segment component in a two-stage sampling scheme with equal probability and without replacement at each stage at time 0.

With the use of (6.54) and (6.66) in (6.45), and with rearrangement of terms then,

\[
\text{Var}(\hat{X}_{thi}) = \left\{ \frac{1}{1 - \frac{5}{3}} \left[ (3 + \frac{3}{2} \alpha^3) - \sigma(Q + \frac{1}{2} Q^3 + \frac{1}{3} \alpha^4) \right] \right\} \\
- (\sigma^2 + \frac{5\sigma^3}{3} + \frac{\sigma^4}{4}) \sum_k \text{Var}(\hat{X}_{thijk}) \\
+ \frac{5^2}{3} \left\{ \left( \frac{2}{5} + \frac{1}{2} \alpha^2 \right) \sigma^2_{thi} + \sum_{\alpha=-1}^{-3} \left( \frac{1}{2} \sigma^2 \alpha (Q^2 + \alpha + 1) \right) \right\} \\
- \frac{1}{2} \frac{1}{3} \left[ (Q^3 + 4Q^2 + 4Q) \right]_{0,-1} S_{thij} \\
+ (2Q^5 + 2Q^4 - 4Q^3)_{0,-3} S_{thij} \\
+ \sum_{\alpha=0}^{-1} \sigma^2 \alpha (2Q^4 + 5Q^3 + 2Q^2) \left[ \alpha, \alpha^{-2} S_{thij} \right] \\
+ \sum_{\alpha=-1}^{-2} \sigma^2 \alpha (Q^3 - 2Q^2 + Q) \left[ \alpha, \alpha^{-1} S_{thij} \right] \right\}.
\]

The identification of the corresponding terms for the \( \sigma^2_d \) component in \( \sigma^2_k \) with the second part of (6.67) is more difficult.
VII. SUMMARY AND RECOMMENDATIONS

Development of multistage designs for statistical surveys in the Philippines is presented. These surveys are collectively known as the Philippine Statistical Survey of Households (PSSH).

The present basic design of the PSSH (Chapter III) is known as multistage with complete replacement of primaries (psus). The psus were drawn with pps in the rural area and with equal probability in the urban area. Five numbered psus were drawn from each stratum. The stratification techniques, methods of sample selection and details of the methods of estimation are described. To simplify the estimation procedure for totals, a self-weighting estimator is utilized. In order to make the self-weighting estimator unbiased, apart from rounding errors, two restrictions were imposed on the design. These were the conditions on $R_t$, the overall raising factor for the $t^{th}$ sector ($t = 1, 2, 3, 4$) and on $f_{thi}$, the $th$ psu constant for the rural areas ($t = 1, 2$) or on $f_{th}$, the $th$ stratum constant for the urban areas ($t = 3, 4$). Since the psus were drawn with complete replacement and were numbered for each stratum, the estimation of variance of estimate consisted of the same ordinary sample variance form for the stratum, the sector and for the whole country.

If the numbering of the psus, $i = 1, 2, ..., i = 5$, is independent and at random within each stratum, then $\hat{\sigma}^2_{x_t}$, the 'short cut method' of variance estimate of the $t^{th}$ sector
(t = 1, 2, 3, 4) is unbiased, but of a rather low precision since it is computed from only 4 degrees of freedom (df) for any given sector. On the other hand, the standard or 'stratified' variance is given by

$$s_{xt}^2 = \sum_h s_{th}^2$$

and is computed from a pooled 4 \( L_t \) df since each stratum estimate \( s_{th}^2 \) is derived from 4 df. An analysis of variance is presented to indicate the differences between the two estimates of variance. If a systematic correlation existed between the numbering, \( i = 1, i = 2, \ldots, i = 5 \), then \( \hat{\sigma}_{xt}^2 \) will be an over estimate, since a component of variance is introduced into the between primary number source of variation.

For the May and October, 1956 rounds, the 'short cut' method of variance estimate, \( \hat{\sigma}_{xt}^2 \), was found to give consistently higher estimates than the 'stratified' variance, \( s_{xt}^2 \), in the rural areas (barrio and poblacion sectors) but consistently lower estimates in the urban areas (capitals and cities and urban Manila sectors). Some systematic bias may have been introduced into the initial numbering procedure of the psus within strata.

Under the requirements of the present basic design of the PSSH, the same panel, except for minor replacements, of sample hhs in the sa (barrio, poblacion or precinct) is interviewed from visit to visit and a complete listing of the
number of hhs in each sa is executed for every visit. These requirements resulted in increased 'response' resistance from panel hhs and in a relatively high cost of listing. To remedy these situations, a segmentation procedure of the sa is introduced and a rotation scheme is developed and instituted into the basic design (Chapter IV). The sa (barrio, poblacion or precinct) is dissected into three or more segments and then the segments are arranged in rotation groups. In general, each rotation group consisted of three segments; two segments or 67 per cent of the segments were common from visit to visit and one segment or 33 per cent was common from year to year. A given segment was in the sample for three consecutive visits and then dropped from the sample. The basic idea is to split the barrio or poblacion (ssu) and the precinct (psu) into segments and to carry out the listing operation only for a sample of three of the segments to be rotated from visit to visit. Thus the cost of listing is reduced inversely with the number of segments in the sa. This technique of segmentation and rotation is found to be feasible and applicable in both rural and urban areas. The rotation scheme is expected to reduce 'response resistance' of panel hhs.

The segmentation of the sa and the subsequent listing of only three segments in a given rotation group reduced the cost of listing from 15 per cent to about 6 per cent of the overall field cost. In addition, the introduction of a simplified
field operation which eliminated a second visit to the sa for the actual interview, resulted also in a reduction in cost of the return travel for the actual interview of sample hhs.

The sampling rates within segments (srws) are adjusted such that the regular self-weighting estimation procedures are retained in the design. Equal srws is used in the urban areas and in the poblacion sector of the rural areas whereas unequal srws is introduced in the barrio rural sector. An attempt is made to equalize the size of segment in the urban areas and in the poblacion sector. This technique introduced some control in minimizing the between segment component in the urban area estimates of variance. Equalization of size of segment was not feasible in the barrio rural sector. Thus, unequal srws but equal take in the number of sample hhs is used in the barrio sector design. This technique is expected to provide some control on the variation of estimates from visit to visit and has the advantage of having equal work loads within segments.

To improve the barrio sample and to bring the barrio sector estimates, particularly those for the region to a higher level of precision, an attempt was made prior to the segmentation procedure to control the variation in the size of the barrio through the equalization of size of population in barrio. This redefinition of the barrio brought about considerable reduction in the coefficients of variation of
regional estimates for the barrio, the barrio sector and the whole country. These aspects of the PSSH design are discussed in Chapter V.

In the urban area, the psus were drawn with complete replacement and sampled at each stage with equal probability. There is a slight loss in precision represented by not applying the finite population correction and a larger loss in precision arises through sampling the psu with equal probability.

To increase the precision of estimates in the urban areas, ratio estimation is introduced into the urban design. This is discussed in Chapter V. It was noted that an auxiliary variable, Z, the count of registered voters which was correlated with population characteristics, is available by precinct, by stratum and by sector with little or not cost to the survey. The survey urban precinct was identical to the registered voters precinct. The $i^{th}$ psu estimate, $\hat{x}_{ith}$, of the stratum total, $x_{ith}$, at the $\alpha$ visit is assumed to be a random observation from an infinite population of such estimates. The same assumption is made for the $i^{th}$ psu estimate of the Z variable. Thus standard techniques of ratio estimation are applied with the $i^{th}$ psu paired estimate, say, $(\hat{x}_{ith}, \hat{z}_{ith})$ as the $i^{th}$ pair of random observations. Three forms of ratio estimators for population in listed hhs and occupation of head of hhs were available, namely:
a) complete ratio estimation - where the segment for the list of population in hhs was identical to the segment for list of registered voters;
b) estimation at the primary stage - where the segment was for the list of hhs only and the precinct (psu) was for the list of registered voters; and
c) no segmentation - where the precinct (psu) was completely listed and the precinct for the list of population in hhs was identical to precinct for the list of registered voters.

Three forms were also derived for a general characteristic observed at the sample hhs stage. Empirical results indicated that a gain in precision of 90 to 95 per cent is obtained with the use of ratio estimators for population in listed hhs and for population in interviewed hhs for urban Manila. This technique will be utilized in the rural area design if found feasible and applicable.

To further improve the precision of estimates, a composite estimation procedure is presented (Chapter VI). Composite estimates are introduced into the design for estimating current total or level, visit-to-visit change and visit-to-visit a year ago change. The composite estimation procedures utilized the rotation scheme which was described earlier (Chapter IV). The form of the estimates was similar to the one used by the U. S. Bureau of the Census in its Current
Population Survey (CPS). With the rotation of sample in the PSSH, the approximate variances of estimates are derived and these are compared to the variances of corresponding estimates obtained with the use of the regular unbiased self-weighting estimates. The precision of the composite estimate is given for some assumed values of \( Q, r_1 \) and \( r_2 \). Substantial gains are obtained for estimates of the visit-to-visit change and visit-to-visit a year ago change for moderate to large values of \( r_1 \) and \( r_2 \) and for values of \( Q \) between 0.3 and 0.5.

The choice of the weight, \( Q \), will depend on the values of \( r_1 \) and \( r_2 \), the correlations of a characteristic observed in hhs of matched segments one visit apart and two visits (one year) apart, respectively. For a multipurpose survey like the PSSH, the empirical values of \( r_1 \) and \( r_2 \) for most of the important characteristics of the survey are needed to help determine the 'appropriate' value of \( Q \) which will result in gains in precision of the estimates for most of these characteristics.

A finite population model consisting of the \( S_{thij} \) rotations in the barrio is developed. An unbiased composite estimate of the barrio total at time 0 is derived. The variance of this estimate is made up of two parts, namely, the within rotation component and the between rotation component. The within rotation component was compared to the corresponding component of the within psu variance of the straight esti-
A series of recommendations based on the development of the design of the Philippine Statistical Survey of Households (PSSH) as described in this work will be submitted to the Philippine Government for consideration and implementation. The subjects of these recommendations will be on the following:

a) undertaking of an independent and random numbering of primaries within stratum; the effect of this procedure on the short cut method of sector variance estimate;

b) utilization of the segmentation procedure and the adoption of the rotation scheme;

c) introduction of ratio estimation into the urban and rural designs; and

d) possible introduction of composite estimation into the design for the most important characteristics of the survey.
VIII. LITERATURE CITED


27. Stevens, W. L. Samples with the same number in each stratum. Biometrika 39:414-417. 1952.


IX. ACKNOWLEDGMENT

The author wishes to extend his sincere thanks to Dr. H. O. Hartley of the Department of Statistics, Iowa State University of Science and Technology, for his help and guidance throughout the preparation of this dissertation.