Developing Component-wise Inter-censal population numbers in Two Indian Districts by Linear Projection, Multi-stage sampling and utilization of auxiliary information in Generalized Regression Technique

Arijit Chaudhuri[1] and Shankar Dihidar
Indian Statistical Institute
Kolkata, India

Abstract

Determination of demographic particulars in inter-censal periods is a vital activity. In this exercise our mission is two-fold. We undertake a suitably stratified multi-stage sample survey utilizing easily available auxiliary information to estimate certain parameters relating to the people in two Indian districts, around late December, 2003, posterior to the last National population census of India in 2001. Secondly, we intend to examine how some symptomatic data may be effectively utilized employing the generalized regression method to derive more accurate estimates, possibly closer to the projected census data for the current time-period. In adjusting inter-censal population figures at lower levels of aggregation namely, the district level totals, utilization of projection and sample survey methods together and utilization of symptomatic data by of statistical modelling should be a common practice in our country too as in USA and Canada. The methods tried turn out promising on applying certain criteria for assessment.

Keywords: Generalized regression estimation, Intercensal population numbers, Symptomatic accounting.

1. Introduction

The two districts of an Indian state, namely Assam covered in this investigation are Kamrup and Cachar. A major portion of our efforts consists in gathering household data by dint of face-to-face interviews with people on canvassing structured questionnaires. For this purpose a sophisticated scheme of sampling was adopted.

Employing traditional sampling weights estimates are obtained for various parameters relating to the people of these two districts roughly during the three months, namely October-December, 2003, several category-wise as are presented in a few Tables below. Later we modify these preliminary, rather elementary, estimates on employing principally the generalized regression (greg) method of estimation. Towards this end we gathered certain symptomatic data, namely the numbers of various schools where the local children are educated, the numbers of household children who attend them and the projected total populations in these two districts for December, 2003 utilizing the official figures available on them in April for the past Indian population censuses in 2001, 1991, 1971, 1961. Since the 1981 census

[1]Dr. Arijit Chaudhuri of Applied Statistics Unit, Indian Statistical Institute, 203, B. T. Road, Kolkata-700108 is the corresponding author. His e-mail address is: achau@isical.ac.in and fax no: +91-33-2577-3104 and +91-33-2577-6680

figures could not be gathered, they were first estimated by a projection device too. The projection formulae employed were

$$P_t = a + bt + e_t, \text{ for } t = 0, 10 \tag{1}$$

to project $\hat{P}_t$ for P_t at $t = 20$
and

$$P_t = ut + ft^2 + \epsilon_t, \text{ for } t = 0, 10, 20, 30, 40 \tag{2}$$

corresponding to 1961, 1971, 1981, 1991 and 2001, where a, b, u, f are unknown constants and e_t, ϵ_t are unkowable error terms for the population numbers P_t, ignoring the error $P_t - \hat{P}_t$ for $t = 20$. However, $\hat{P}_t$ projected for P_t using (2) was found to behave linearly in t and so the simpler formula (1) itself was used also for all $t = 0, 10, 20, 30, 40$ to find the projection rule adequate.

2. Methods of sampling and estimation

(a) Selection of houeholds covering urban and rural areas in the district of Kamrup

Besides the most important and populous city namely Gauhati there are eleven more cities and/or towns in the Kamrup district.

Gauhati is selected purposively. Other cities/towns along with their outskirts or outgrowths consist of a group of 5 for which Indian NSSO(National Sample Survey Organization)made Urban Frame Survey (UFS) blocks with their map-books available and 6 others with no UFS blocks yet framed by NSSO.

The number of UFS blocks contained in a city/town is treated as its size-measures. From the 5 cities of the first group 2 are selected following the sampling scheme given by Rao, Hartley and Cochran (RHC, 1962). This RHC scheme will be presently described. Out of the second group of 6 cities/towns, 2 are selected by Simple random sampling without replacement (SRSWOR), independently of the selection method followed for the first group. Thus we have this stratified sampling procedure adopted for the first stage units (fsu), namely the cities/towns.

The second stage units (ssu) for Gauhati and the cities with UFS blocks are the UFS blocks. From Gauhati 6 UFS blocks are selected by SRSWOR. Also 2 UFS blocks are selected by SRSWOR method independently from each of the 2 cities selected.

Finally, 25 households as the third stage units (tsu) are selected by the SRSWOR method independently from the 10 second stage units selected as above, giving a sample of 250 households (hh) from the 5 cities chosen from the first two strata as above. For each of the 2 cities with no UFS blocks in the $3rd$ stratum, separately and independently 25 households (hh) as second stage units are selected by SRSWOR schemes. Thus in all 300 households are selected by 3-stage sampling from the 2 strata with UFS blocks and by 2-stage sampling from the third stratum with no UFS blocks.

For selection of the rural households from the 17 blocks taken as fsu's, 4 blocks are selected by SRSWOR, 2 Gram Panchayats (GP) are selected by SRSWOR method independently from each selected block as the ssu's. Thus we got 8 GP's at $2nd$ stage. As the tsu's, from the

4 bigger GP's 3 villages are independently selected and from the 4 smaller GP's 2 villages are similarly chosen, each tsu got being selected by the SRSWOR scheme in both cases. From each of the 20 tsu's selected the fourth or the ultimate stage units (usu), namely the households are independently selected by the SRSWOR method. Thus 500 households are chosen from the Kamrup district.

(b) Selection of households covering urban and rural areas from the Cachar district

Silchar is the largest city in Cachar and only in Lakshmipur among the other 7 cities in this district there are NSSO-made UFS blocks. From the UFS blocks respectively in them 8 and 2 UFS blocks are chosen by SRSWOR scheme, in independent manners. From the remaining 6 cities in Cachar 2 are selected by SRSWOR. Next, 25 households are selected by SRSWOR method as the tsu's from each of the selected ssu's, namely the UFS blocks chosen from the fsu's namely the cities of Silchar and Lakshmipur. Finally, 25 households are selected by SRSWOR method independently as the ssu's from each of the 2 selected fsu's which are the other 2 selected cities. Thus in all 300 households are selected here by 2-stage or 3-stage sampling method.

Again 500 households are selected from the rural section of the Cachar district by SRSWOR method of sampling in 4 stages as are from Kamrup district on choosing 4 blocks, 2 GP's per block, 3 villages per big selected GP and 2 villages per selected small GP and 25 households per selected village. A big or a small village is subjectively determined.

(c) Estimation Procedure

In order to describe the traditional method of estimation let us first explain the scheme of sample selection by the RHC scheme.

Let N denote the number of units in a finite survey population U of the first stage units and n be the number of distinct sampling or selection units to be chosen. Further, let x_i be the positive integer numbers treated as the size-measures of the units $i(i = 1, \ldots, N)$ in the population with X as the total of x_i's and $p_i = x_i/X$, $i \in U$. Let U be randomly divided into n disjoint groups taking N_i units in the ith group. Writing $\sum_n$ as the sum over the n disjoint groups we have $\sum_n N_i = N$. Writing Q_i as the sum of the p_i values for the units, say, $i_1, \ldots, i_j, \ldots, i_{N_i}$ falling in the ith group,

$$Q_i = p_{i_1} + \ldots + p_{i_{N_i}},\ i = 1, \ldots, n$$

Let one unit, say, i_j be chosen from the $i - th$ group with a probability p_{ij}/Q_i. Let this be repeated independently across all the n groups. The resulting selection scheme is due to RHC.

Let y be a real valued variable of interest taking values y_i for i in U. Then, $Y = \sum y_i$, is the total of y for the population U. An unbiased estimator for Y given by RHC is

$$t = \sum_n y_i \frac{Q_i}{p_i}$$

Here for simplicity (y_i, p_i) is written as the y and the size-measure values for the unit chosen from the ith group.

Let us define two numbers A and B as in the following,

$$A = \frac{\sum_n N_i^2 - N}{N(N-1)} \text{ and } B = \frac{\sum_n N_i^2 - N}{N^2 - \sum_n N_i^2}$$

Then, the formulae for the variance of t and an unbiased variance estimator of t are respectively given by

$$V(t) = A\sum_{i=1}^{N}\sum_{i'=1, i<i'}^{N} p_i p_{i'}(\frac{y_i}{p_i} - \frac{y_{i'}}{p_{i'}})^2$$

$$v(t) = B\sum_n Q_i(\frac{y_i}{p_i} - t)^2 = B[\sum_n Q_i \frac{y_i^2}{p_i^2} - t^2] = B\sum_{i=1}^{n}\sum_{i'=1, i<i'}^{n} Q_i Q_{i'}(\frac{y_i}{p_i} - \frac{y_{i'}}{p_{i'}})^2$$

writing $\sum_n \sum_n$ to denote sum over distinct non-overlapping pairs of units selected by the RHC scheme, with no duplications.

In the present case, obviously for fsu's sampled, say the ith fsu, the value y_i is not ascertained, but is itself estimated through ssu's, tsu's, usu's sampled from the ith fsu in successive stages. So, the presentation of the method of estimation in the present case needs to be explained.

Let the ith fsu be supposed to consist of M_i ssu's of which m_i are chosen by SRSWOR method, the jth ssu of ith fsu sampled, consist of T_{ij} tsu's of which t_{ij}'s are chosen by SRSWOR method, the kth usu of jth ssu of ith fsu selected, consist of P_{ijk} usu's of which p_{ijk} usu's be selected by SRSWOR and so on if further stages are added if needed. Here,

$$i = 1, \ldots, N,\ j = 1, \ldots, M_i,\ k = 1, \ldots, T_{ij} \text{ and } l = 1, \ldots, P_{ijk}$$

are the subscripts to be used to identify the units at the successive stages 1, 2, 3 and 4.

since selection from the strata is done independently, it is enough to present an estimate of a stratum total and an estimate for the variance of this estimate. A corresponding total for the population and its variance estimate is obtained respectively just by simple addition across the strata.

Let Y be unbiasedly estimated by

$$\hat{Y} = \sum_n y_i^* \frac{Q_i}{p_i},$$

noting that the fsu's are sampled by the RHC scheme as described above and y_i^* is an unbiased estimate of y_i, the y-value for the ith fsu. Now

$$y_i^* = \frac{M_i}{m_i}\sum_{j=1}^{m_i} \frac{T_{ij}}{t_{ij}}\sum_{k=1}^{t_{ij}} \frac{P_{ijk}}{p_{ijk}}\sum_{l=1}^{p_{ijk}} y_{ijkl}$$

Here y_{ijkl} denotes the y-values for the lth usu of kth tsu of jth ssu of the ith fsu and the summation symbols noted above denote addition over the sampled units of the respective stages 2, 3 and 4. Of course y_i itself may be written as

$$y_i = \sum_{j=1}^{M_i} \sum_{k=1}^{T_{ij}} \sum_{l=1}^{P_{ijk}} y_{ijkl} \text{ and } Y = \sum_{i=1}^{N} y_i$$

Let us write

$$c_{ijk} = \frac{P_{ijk}}{p_{ijk}} \sum_{l=1}^{p_{ijk}} y_{ijkl} \text{ and } \bar{c_{ijk}} = \frac{1}{p_{ijk}} \sum_{l=1}^{p_{ijk}} y_{ijkl},$$

$$a_{ij} = \frac{T_{ij}}{t_{ij}} \sum_{k=1}^{t_{ij}} c_{ijk} \text{ and } \bar{a_{ij}} = \frac{1}{t_{ij}} \sum_{k=1}^{t_{ij}} c_{ijk}$$

Then,

$$y_i^* = \frac{M_i}{m_i} \sum_{j=1}^{m_i} a_{ij}$$

Let us also write

$$\bar{y_i^*} = \frac{1}{m_i} \sum_{j=1}^{m_i} a_{ij}$$

Varince estimates of these estimates may then be taken as

$$v(c_{ijk}) = P_{ijk}^2 (\frac{1}{p_{ijk}} - \frac{1}{P_{ijk}}) \frac{1}{p_{ijk} - 1} \sum_{l=1}^{p_{ijk}} (y_{ijkl} - \bar{c_{ijk}})^2$$

$$v(a_{ij}) = T_{ij}^2 (\frac{1}{t_{ij}} - \frac{1}{T_{ij}}) \frac{1}{t_{ij} - 1} \sum_{l=1}^{t_{ij}} (c_{ijk} - \bar{a_{ij}})^2 + \frac{T_{ij}}{t_{ij}} \sum_{k=1}^{t_{ij}} v(c_{ijk})$$

$$v(y_i^*) = M_i^2 (\frac{1}{m_i} - \frac{1}{M_i}) \frac{1}{m_i - 1} \sum_{j=1}^{m_i} (a_{ij} - \bar{y_i^*})^2 + \frac{M_i}{m_i} \sum_{j=1}^{m_i} v(a_{ij})$$

The estimate Y^* of the final stratum total Y and the variance estimate of Y^* are given by

$$Y^* = \sum_n y_i^* \frac{Q_i}{p_i}$$

$$v(Y^*) = B \sum_{i=1}^{n} \sum_{i'=1, i<i'}^{n} Q_i Q_{i'} (\frac{y_i}{p_i} - \frac{y_{i'}}{p_{i'}})^2 + \sum_n \frac{Q_i}{p_i} v(y_i^*)$$

An estimate for the district total is obtained by adding the Y^* values over the strata and the corresponding variance estimate is obtained by adding the $v(Y^*)$ values also across the same strata. Needless to mention, the values of N, M_i, T_{ij}, P_{ijk}; n, m_i, t_{ij}, p_{ijk} are of course all gathered in course of the field investigation and/or are as pre-assigned.

If besides y there be another variable of interest z for which also z_{ijkl}'s are the values of the $4th$ (here ultimate) stage units and z_{ijk}, z_{ij}, z_i and $Z = \sum_{i=1}^{N} z_i = \sum_{i=1}^{N} \sum_{j=1}^{M_i} z_{ij} =$

$\sum_{i=1}^{N}\sum_{j=1}^{M_i}\sum_{k=1}^{T_{ij}} z_{ijk} = \sum_{i=1}^{N}\sum_{j=1}^{M_i}\sum_{k=1}^{T_{ij}}\sum_{l=1}^{P_{ijk}} z_{ijkl}$ are defined and Z is estimated by $\hat{Z}$ following the same procedure adopted as in estimating Y by $\hat{Y}$, it may be of interest to estimate the ratio

$$R = \frac{Y}{Z}$$

If so, then, R is estimated by

$$\hat{R} = \frac{\hat{Y}}{\hat{Z}}$$

Then the Mean Square Error (MSE) of $\hat{R}$ around R is estimated by

$$v(\hat{R}) = \frac{1}{(\hat{Z})^2} v(\hat{Y})|_{y_{ijkl}=y_{ijkl}-\hat{R}z_{ijkl}}$$

This means that in the formula for $v(\hat{Y})$ we have to throughout replace y_{ijkl} by $y_{ijkl} - \hat{R}z_{ijkl}$.

Now let us turn to the utilization of symptomatic variables by dint of the application of the generalized regression (greg) method of estimating, first the totals and later the ratios of totals.

Let x denote a symptomatic characteristic we mentioned earlier like (i) the number of schools attended by the students, (ii) the number of school-going students for each of the two districts of Kamrup and Cachar and (iii) the projected total populations in these two districts.

Before introducing the greg estimators let us first start with the projection procedure needed to generate the figures in (iii) noted above.

Given the census population figures for the district of Kamrup as found from the 1961 and 1971 population censuses referring to the month of April in both cases denoted as P_0 and P_1 respectively, let us write

$$P_t = \alpha + \beta t + e_t, t = 0, 1, 2, \ldots$$

with α, β as constants and e_t as a random error. Then, solving

$$\frac{\partial S}{\partial \alpha} = 0 \text{ and } \frac{\partial S}{\partial \beta} = 0, \text{ on writing } S = \sum_{t=0}^{1} {e_t}^2 = \sum_{t=0}^{1} (P_t - \alpha - \beta t)^2$$

we get

$$\sum_{t=0}^{1} P_t = 2\alpha + \beta \text{ and } \sum_{t=0}^{1} tP_t = \alpha \sum_{t=0}^{1} t + \beta \sum_{t=0}^{1} t^2$$

So, the least squares estimates of α and β are given by

$$\hat{\alpha} = P_0 \text{ and } \hat{\beta} = P_1 - P_0$$

Then, P_t is estimated by

$$\hat{P}_t = P_0 + (P_1 - P_0)t$$

So, $\hat{P}_2 = P_0 + (P_1 - P_0)2 = 2P_1 - P_0$ is the estimated population figure for April, 1981. Ignoring the error $(P_2 - \hat{P}_2)$ and taking $\hat{P}_2$ as P_2, we shall now use P_0, P_1, P_2, P_3 and P_4 as the census figures for 1961, 1971, 1981, 1991 and 2001 and postulate the model

$$P_t = u + ft + \epsilon_t, t = 0, 1, 2, \ldots$$

with u and f as constants and ϵ_t as the random error component. We shall apply again the least squares principle to estimate $P_{4\frac{4}{15}}$ which is the census population figure for Kamrup in December, 2003, 4 decades and $2\frac{2}{3}$ years away from April, 1961.

So, to get the least squares estimates of u and f we have to solve

$$\frac{\partial}{\partial u}\sum_{t=0}^{4}(P_t - u - ft)^2 = 0 \text{ and } \frac{\partial}{\partial f}\sum_{t=0}^{4}(P_t - u - ft)^2 = 0$$

which leads to the normal equations

$$\sum P_t = 5u + f\sum t \text{ and } \sum tP_t = u\sum t + f\sum t^2$$

Thus, solving these two equations, we get the least squares estimates $\hat{u}$ of u and $\hat{f}$ of f. Then, $\hat{P}_t = \hat{u} + \hat{f}t$ is the least squares estimate of P_t and we calculate $\hat{P}_{4\frac{4}{15}}$ on putting $t = 4\frac{4}{15}$ in $\hat{P}_t$ so as to get an estimate of the population figure for Kamrup in December, 2003. Exactly a same procedure is followed also to estimate the population figure for the Cachar district in December, 2003.

Let us now take x as the symptomatic variable taken one after another as (i) and (ii) above and employ the following version of the Greg method of estimation.

Let us start with y_i^* as per our notations explained and x_i be the value of the auxiliary variable for ith selected unit at the $1st$ stage sampling. The total of the values of the auxiliary variable of all the N units denoted as X is also known to us.

The greg estimator for Y of urban areas is then taken as

$$\hat{Y_g} = \sum_n y_i^* \frac{Q_i}{p_i} + \beta^*(X - \sum_n x_i \frac{Q_i}{p_i})$$

on taking

$$\beta^* = \frac{\sum_n y_i^* x_i R_i}{\sum_n (x_i)^2 R_i},\ R_i = \frac{1 - \frac{p_i}{Q_i}}{\frac{p_i}{Q_i}x_i}$$

We can write $\hat{Y_g}$ in two ways. One way is

$$\hat{Y_g} = \sum_n (y_i^* - \beta^* x_i)\frac{Q_i}{p_i} + \beta^* X = \sum_n e_i^* \frac{Q_i}{p_i} + \beta^* X, \text{ with } e_i^* = y_i^* - \beta^* x_i$$

Another way is

$$\hat{Y_g} = \sum_n y_i^* \frac{Q_i}{p_i} + (\frac{\sum_n y_i^* x_i R_i}{\sum_n (x_i)^2 R_i})(X - \sum_n x_i \frac{Q_i}{p_i})$$

$$= \sum_n y_i^* \frac{Q_i}{p_i}[1 + (\frac{\frac{p_i}{Q_i} x_i R_i}{\sum_n (x_i)^2 R_i})(X - \sum_n x_i \frac{Q_i}{p_i})]$$

$$= \sum_n y_i^* \frac{Q_i}{p_i}[1 + (\frac{1 - \frac{p_i}{Q_i}}{\sum_n (x_i)^2 R_i})(X - \sum_n x_i \frac{Q_i}{p_i})]$$

i.e.

$$\hat{Y_g} = \sum_n y_i^* \frac{Q_i}{p_i} g_{1i}$$

on writing

$$g_{1i} = [1 + (\frac{1 - \frac{p_i}{Q_i}}{\sum_n (x_i)^2 R_i})(X - \sum_n x_i \frac{Q_i}{p_i})]$$

Then, an estimator of the MSE of $\hat{Y_g}$ about Y is

$$v(\hat{Y_g}) = v(\hat{Y})|_{y_i^* = e_i^*} + \sum_n \frac{Q_i}{p_i} g_{1i} v(y_i^*)$$

The greg estimator for Y of rural areas is taken as

$$\hat{Y_g} = \frac{N}{n} \sum_n y_i^* + \beta^* (X - \frac{N}{n} \sum_n x_i)$$

on taking

$$\beta^* = \frac{\sum_n (y_i^* - \bar{y^*})(x_i - \bar{x})}{\sum_n (x_i - \bar{x})^2}, \text{ with } \bar{y^*} = \frac{\sum_n y_i^*}{n} \text{ and } \bar{x} = \frac{\sum_n x_i}{n}$$

Here also we can write $\hat{Y_g}$ in two ways. One way is

$$\hat{Y_g} = \frac{N}{n} \sum_n (y_i^* - \beta^* x_i) + \beta^* X = \frac{N}{n} \sum_n e_i^* + \beta^* X, \text{ with } e_i^* = y_i^* - \beta^* x_i$$

Another way is

$$\hat{Y_g} = \frac{N}{n} \sum_n y_i^* + \frac{\sum_n (y_i^* - \bar{y^*})(x_i - \bar{x})}{\sum_n (x_i - \bar{x})^2} (X - \frac{N}{n} \sum_n x_i)$$

$$= \frac{N}{n} \sum_n y_i^* + \frac{\sum_n y_i^* (x_i - \bar{x})}{\sum_n (x_i - \bar{x})^2} (X - \frac{N}{n} \sum_n x_i)$$

$$= \frac{N}{n} \sum_n y_i^* [1 + \frac{\frac{n}{N}(x_i - \bar{x})}{\sum_n (x_i - \bar{x})^2} (X - \frac{N}{n} \sum_n x_i)]$$

i.e.

$$\hat{Y}_g = \frac{N}{n}\sum_n y_i^* g_{2i}$$

on writing

$$g_{2i} = 1 + \frac{\frac{n}{N}(x_i - \bar{x})}{\sum_n (x_i - \bar{x})^2}(X - \frac{N}{n}\sum_n x_i)$$

Then, an estimator of the MSE of $\hat{Y}_g$ about Y is

$$v(\hat{Y}_g) = v(\hat{Y})|_{y_i^* = e_i^*} + \frac{N}{n}\sum_n g_{2i} v(y_i^*)$$

As an estimator for $R = \frac{Y}{Z}$ we shall also take

$$\hat{R}_g = \frac{\hat{Y}_g}{\hat{Z}_g}$$

and take $v(\hat{R}_g)$ as

$$v(\hat{R}_g) = \frac{1}{(\hat{Z}_g)^2} v(\hat{Y}_g)|_{y_{ijkl} = y_{ijkl} - \hat{R}_g z_{ijkl}}$$

As y and z we shall take usually as "$y = z$ multiplied by some indicator variable ". For example, y may denote the number of household members who died during the period December, 2002 - December, 2003 and z may denote the total number of household members present at the time of survey. Another example may be that y is the number of male members or female members in the household and z is the total number of household members etc.

3. Certain Comments on the findings of this investigation as presented in the Tables below in Section 4

In the Tables below as given in Section 4 we have presented estimates by the traditional versus the generalized regression (greg) methods for the totals of several demographic characteristics, ratios of totals and a few percentage distributions. In order to apply the greg method of estimation on postulating a linear regression through the origin concerning the estimated total for the characteristics of interest and symptomatic one for the first stage unit, namely the block for the rural area or the city/town in case of the urban sector, only one auxiliary characteristic, namely the number of school going students several sector-wise could be gathered as a symptomatic variable. We expected the two sets of estimates not to differ much and the estimated CV's and SE's of the greg procedures to be less than the ones by the traditional methods. This is mostly realized when the sample-size involved is not too small because a striking feature of the greg estimator is it's asymptotic design unbiasedness and equivalently asymptotic design consistency (ADU-ness and ADC respectively). There are exceptional cases, of course, covering the situations when the characteristic of interest relates to relatively small segments while the symptomatic characteristic relating to the first stage unit relates to an appreciably sizable totality. Incidentally, we show the greg estimators for estimating totals of several variables using the symptomatic variable as the (i) number of

students and except for estimating the total district populations also (*ii*) the population size. We find the latter symptomatic variable to yield better results in almost all situations in terms of the criterion of "coefficient of variation". This may be because the former variable is not accordingly ascertainable compared to the second.

4. The Findings

In the tables below we present in brief the estimates of various totals (Y) relating to the characteristics covered in this investigation and ratios (and/or percentages) of totals ($R = \frac{Y}{X}$) using (i) the traditional stratified multistage sampling weights ($\hat{Y}, \hat{R}$) and (ii) thier generalized (greg) versions ($\hat{Y}_g, \hat{R}_g$). We also present their (iii) estimated standard errors (SE) which are the positive square roots of their estimated variances and (iv) estimated coefficients of variations $CV(\hat{Y})$, $CV(\hat{Y}_g)$, $CV(\hat{R})$ and $CV(\hat{R}_g)$. The numbers of observations (n) on which these estimates are based are also indicated. The total numbers of area-wise school-going students are used as auxiliary variables in finding estimates $greg1$ and the projected populations in Dec, 2003 are used as auxiliary variables in finding estimates $greg2$ in generalized regression method. It is meaningless to find estimates of total population in $greg2$ method. So the estimates of total population in $greg2$ method are not shown in tables.

(4.1) Findings relating to the Rural areas of Kamrup District

Table 4.1.1
Total Population in Rural Areas of Kamrup District

Estimator	*Estimate*	*SE*	*CV*	*n*
Traditional ($\hat{Y}$)	2234340	357402	16.0	500
Greg1 ($\hat{Y}_g$)	2062871	303094	14.7	500

Table 4.1.2
Average household size in Rural Areas of Kamrup District

Estimator	*Estimate*	*SE*	*CV*	*n*
Traditional ($\hat{R}$)	5.4	0.13	2.4	500
Greg1 ($\hat{R}_g$)	5.1	0.15	3.0	500
Greg2 ($\hat{R}_g$)	5.1	0.15	2.9	500

Table 4.1.3
Total number of children born since 01.01.2002 in Rural Areas of Kamrup District

	Male			Female			Total		
Estimator	*EST*	*SE*	*CV*	*EST*	*SE*	*CV*	*EST*	*SE*	*CV*
Traditional ($\hat{Y}$)	25854	5830	22.5	39909	11997	30.1	65763	16800	25.5
Greg1 ($\hat{Y}_g$)	21648	3523	16.3	31661	7806	24.6	53309	9645	18.1
Greg2 ($\hat{Y}_g$)	23387	3443	14.7	34693	6705	19.3	58080	8217	14.1
n	36			46			82		

Table 4.1.4
Total number of persons died since 01.01.2002 in Rural Areas of Kamrup District

	Male			Female			Total		
Estimator	*EST*	*SE*	*CV*	*EST*	*SE*	*CV*	*EST*	*SE*	*CV*
Traditional ($\hat{Y}$)	17756	5397	30.4	6573	2256	34.3	24329	7022	28.9
Greg1 ($\hat{Y}_g$)	19679	4962	25.2	6569	2256	34.3	26248	6695	25.5
Greg2 ($\hat{Y}_g$)	17340	5339	30.8	6518	2253	34.6	23857	6964	29.2
n	21			9			30		

Table 4.1.5
Average number of hostel goers per household in Rural Areas of Kamrup District

	Male			Female			Total		
Estimator	*EST*	*SE*	*CV*	*EST*	*SE*	*CV*	*EST*	*SE*	*CV*
Traditional ($\hat{R}$)	0.04	0.01	23.7	0.02	0.005	32.8	0.06	0.01	21.1
Greg1 ($\hat{R}_g$)	0.03	0.01	32.5	0.01	0.005	35.7	0.05	0.01	26.7
Greg2 ($\hat{R}_g$)	0.04	0.01	28.1	0.02	0.005	32.5	0.05	0.01	23.3
n	25			10			35		

Table 4.1.6

Percentage distribution of households by types of houses made in Rural Areas of Kamrup District

Type of house	*EST* (%)			*SE*			*CV*			n
	trad	greg1	greg2	trad	greg1	greg2	trad	greg1	greg2	
Pucca	23.2	23.3	22.7	2.6	2.1	2.5	10.5	10.6	10.8	121
Katcha	56.9	55.4	53.9	4.3	4.5	4.0	6.9	7.14	7.4	282
Semi Pucca	19.9	16.2	18.0	4.1	4.3	3.8	18.4	24.1	21.3	97
Total	100.0									500

Table 4.1.7

Percentage distribution of households by availability of electricity in Rural Areas of Kamrup District

Availability of electricity	*EST* (%)			*SE*			*CV*			n
	trad	greg1	greg2	trad	greg1	greg2	trad	greg1	greg2	
Yes	50.4	44.3	48.9	5.3	6.1	5.8	11.4	13.0	12.1	251
No	49.6	50.7	46.7	6.4	6.2	5.9	11.5	11.4	12.6	249
Total	100.0									500

Table 4.1.8

Percentage distribution of households by sources of drinking water in Rural Areas of Kamrup District

Sources of drinking water	*EST* (%)			*SE*			*CV*			n
	trad	greg1	greg2	trad	greg1	greg2	trad	greg1	greg2	
Municipal tap	0.2	0.2	0.2	0.2	0.2	0.2	111.6	102.1	83.7	1
Tubewell	46.4	32.9	41.9	14.0	15.0	14.6	30.6	44.8	34.7	238
Well	25.3	29.2	24.5	7.6	7.7	7.8	30.2	26.5	31.9	119
Neighbour's	10.4	10.8	10.3	2.0	2.0	2.1	19.9	19.7	20.6	59
Public facilities	17.0	21.1	17.1	6.0	7.0	6.6	37.9	31.5	38.4	81
River/canal	0.6	0.7	0.6	0.6	0.6	0.6	100.8	84.4	101.8	2
Pond	–	–	–	–	–	–	–	–	–	0
Total	100.0									500

Table 4.1.9

Percentage distribution of persons by different age-groups in Rural Areas of Kamrup District (Traditional Estimates)

Age group (years)	Male			Female			Total		
	EST	*SE*	*CV*	*EST*	*SE*	*CV*	*EST*	*SE*	*CV*
< 1	1.0	0.34	34.3	2.0	0.39	19.8	1.5	0.32	21.1
$1 - 4$	4.3	0.45	10.3	6.4	0.64	10.1	5.4	0.31	5.9
$5 - 14$	20.1	0.95	4.7	22.8	1.7	7.3	21.6	1.19	5.5
$15 - 24$	20.7	2.0	9.5	20.0	0.78	3.9	20.3	0.95	4.7
$25 - 34$	19.0	1.6	8.4	18.2	1.7	9.3	18.6	1.6	8.7
$35 - 44$	14.3	0.87	6.1	14.3	1.5	10.3	14.3	0.92	6.5
$45 - 54$	11.4	0.8	7.0	8.8	0.35	3.9	10.0	0.54	5.4
$55 - 64$	5.8	0.25	4.3	5.2	0.68	12.9	5.5	0.29	5.4
$65 - 74$	2.6	0.58	22.4	1.8	0.28	15.3	2.2	0.33	15.0
≥ 75	0.75	0.26	35.3	0.5	0.21	46.8	0.6	0.21	35.0
Total	100.0			100.0			100.0		

Table 4.1.10

Percentage distribution of persons by different age-groups in Rural Areas of Kamrup District (Greg1 Estimates)

Age group (years)	Male			Female			Total		
	EST	*SE*	*CV*	*EST*	*SE*	*CV*	*EST*	*SE*	*CV*
< 1	0.8	0.35	45.0	1.5	0.44	28.9	1.2	0.35	29.8
$1 - 4$	4.0	0.43	10.8	5.5	0.73	13.2	4.8	0.36	7.6
$5 - 14$	19.1	0.97	5.1	22.6	1.7	7.4	20.9	1.18	5.6
$15 - 24$	20.3	2.0	9.8	18.7	0.77	4.1	19.5	0.93	4.8
$25 - 34$	16.4	1.9	11.4	15.8	1.9	12.1	16.1	1.9	11.6
$35 - 44$	13.6	0.82	6.0	14.6	1.5	10.3	14.1	0.92	6.5
$45 - 54$	10.5	0.87	8.3	8.2	0.36	4.4	9.3	0.59	6.3
$55 - 64$	5.4	0.23	4.2	4.6	0.75	16.4	5.0	0.34	7.0
$65 - 74$	2.1	0.6	28.4	1.8	0.28	16.0	1.9	0.34	17.5
≥ 75	0.75	0.27	35.6	0.4	0.21	52.8	0.6	0.21	36.6
Total									

Table 4.1.11

Percentage distribution of persons by different age-groups in Rural Areas of Kamrup District (Greg2 Estimates)

Age group (years)	Male			Female			Total		
	EST	*SE*	*CV*	*EST*	*SE*	*CV*	*EST*	*SE*	*CV*
< 1	0.9	0.4	39.8	1.7	0.4	24.3	1.3	0.3	25.5
$1 - 4$	4.2	0.4	10.6	5.8	0.7	12.2	5.0	0.3	6.9
$5 - 14$	19.0	1.0	5.1	21.7	1.7	7.6	20.4	1.2	5.8
$15 - 24$	19.9	2.0	10.1	19.0	0.8	4.1	19.4	0.9	4.8
$25 - 34$	17.4	1.8	10.2	16.8	1.8	11.0	17.1	1.8	10.5
$35 - 44$	13.7	0.8	6.1	13.9	1.5	10.3	13.8	0.9	6.4
$45 - 54$	10.6	0.9	8.2	8.3	0.4	4.3	9.4	0.6	6.2
$55 - 64$	5.5	0.2	4.3	4.8	0.7	15.4	5.1	0.3	6.5
$65 - 74$	2.5	0.6	24.0	1.8	0.3	16.0	2.1	0.3	15.9
≥ 75	0.8	0.3	33.6	0.5	0.2	45.1	0.7	0.2	33.6
Total									

(4.2) Findings relating to the Urban areas of Kamrup District

Table 4.2.1
Total Population in Urban Areas of Kamrup District

Estimator	*Estimate*	*SE*	*CV*	*n*
Traditional ($\hat{Y}$)	799118	71475	8.9	300
Greg1 ($\hat{Y}_g$)	661265	53604	8.1	300

Table 4.2.2
Average household size in Urban Areas of Kamrup District

Estimator	*Estimate*	*SE*	*CV*	*n*
Traditional ($\hat{R}$)	4.7	0.08	1.7	300
Greg1 ($\hat{R}_g$)	3.9	0.15	3.97	300
Greg2 ($\hat{R}_g$)	4.7	0.08	1.7	300

Table 4.2.3
Total number of children born since 01.01.2002 in Urban Areas of Kamrup District

	Male			Female			Total		
Estimator	*EST*	*SE*	*CV*	*EST*	*SE*	*CV*	*EST*	*SE*	*CV*
Traditional ($\hat{Y}$)	12575	2662	21.2	10650	3463	32.5	23225	2284	9.8
Greg1 ($\hat{Y}_g$)	15651	2106	13.5	4780	1535	32.1	23225	2284	9.8
Greg2 ($\hat{Y}_g$)	11816	2071	17.5	10792	2274	21.1	22608	2008	8.9
n	27			17			44		

Table 4.2.4
Total number of persons died since 01.01.2002 in Urban Areas of Kamrup District

	Male			Female			Total		
Estimator	*EST*	*SE*	*CV*	*EST*	*SE*	*CV*	*EST*	*SE*	*CV*
Traditional ($\hat{Y}$)	7281	2521	34.6	4536	2107	46.4	11817	3495	29.6
Greg1 ($\hat{Y}_g$)	4570	2135	46.7	1882	1607	85.4	6451	2223	34.5
Greg2 ($\hat{Y}_g$)	7231	1867	25.8	4589	1772	38.6	11820	2049	17.3
n	12			6			18		

Table 4.2.5
Average number of hostel goers in Urban Areas of Kamrup District

	Male			Female			Total		
Estimator	*EST*	*SE*	*CV*	*EST*	*SE*	*CV*	*EST*	*SE*	*CV*
Traditional ($\hat{R}$)	0.06	0.01	20.0	0.04	0.01	32.1	0.1	0.02	21.9
Greg1 ($\hat{R}_g$)	0.06	0.01	22.9	0.04	0.01	32.5	0.1	0.02	24.1
Greg2 ($\hat{R}_g$)	0.06	0.01	20.2	0.04	0.01	32.3	0.1	0.02	22.2
n	18			8			26		

Table 4.2.6

Percentage distribution of households by types of houses made in Urban Areas of Kamrup District

Type of house	*EST* (%)			*SE*			*CV*			n
	trad	greg1	greg2	trad	greg1	greg2	trad	greg1	greg2	
Pucca	66.4	70.5	66.7	7.1	7.6	7.4	10.7	10.8	11.1	186
Katcha	16.4	13.2	16.0	3.1	3.6	3.3	19.0	27.2	20.5	66
Semi Pucca	17.3	3.1	18.3	6.6	7.6	6.7	38.0	242.7	36.9	48
Total	100.0									300

Table 4.2.7

Percentage distribution of households by availability of electricity in Urban Areas of Kamrup District

Availability of electricity	*EST* (%)			*SE*			*CV*			n
	trad	greg1	greg2	trad	greg1	greg2	trad	greg1	greg2	
Yes	86.4	77.3	87.4	3.5	3.4	3.7	4.1	4.4	4.3	255
No	13.6	9.5	13.5	3.5	3.9	3.7	25.9	40.7	27.2	45
Total	100.0									300

Table 4.2.8

Percentage distribution of households by sources of drinking water in Urban Areas of Kamrup District

Sources of drinking water	*EST* (%)			*SE*			*CV*			n
	trad	greg1	greg2	trad	greg1	greg2	trad	greg1	greg2	
Municipal tap	1.5	2.0	1.5	1.2	1.2	1.2	79.3	59.6	85.0	7
Tubewell	56.5	29.3	57.0	11.6	13.8	12.0	20.6	47.1	21.1	195
Well	27.5	40.1	27.7	9.9	11.0	10.3	36.0	27.2	37.1	62
Neighbour's	3.4	1.0	3.7	2.0	2.3	2.1	60.1	–	55.7	6
Public facilities	9.4	13.0	9.2	6.8	7.0	7.1	72.4	53.7	76.9	30
River/canal	–	–	–	–	–	–	–	–	–	0
Pond	–	–	–	–	–	–	–	–	–	0
Total	100.0									300

Table 4.2.9

Percentage distribution of persons by different age-groups in Urban Areas of Kamrup District (Traditional Estimates)

Age group (years)	Male			Female			Total		
	EST	*SE*	*CV*	*EST*	*SE*	*CV*	*EST*	*SE*	*CV*
< 1	1.8	0.55	31.2	1.8	0.6	31.7	1.8	0.23	13.1
$1 - 4$	4.2	1.2	29.6	6.3	1.1	17.0	5.2	0.95	18.2
$5 - 14$	17.8	1.4	7.9	17.7	2.3	13.2	17.8	1.7	9.5
$15 - 24$	19.2	2.3	12.2	19.9	2.4	12.0	19.5	1.8	9.4
$25 - 34$	17.8	1.3	7.0	18.3	1.6	8.8	18.1	1.2	6.8
$35 - 44$	17.8	1.1	6.4	15.4	1.5	9.4	16.7	1.2	7.1
$45 - 54$	10.8	1.0	8.9	10.1	1.2	19.7	10.4	1.1	10.3
$55 - 64$	6.1	1.5	24.5	5.4	0.9	16.1	5.8	1.1	18.1
$65 - 74$	3.6	0.6	16.7	4.0	1.0	25.2	3.8	0.75	19.9
≥ 75	1.0	0.49	46.6	0.72	0.32	45.4	0.9	0.3	33.4
Total	100.0			100.0			100.0		

Table 4.2.10

Percentage distribution of persons by different age-groups in Urban Areas of Kamrup District (Greg1 Estimates)

Age group (years)	Male			Female			Total		
	EST	*SE*	*CV*	*EST*	*SE*	*CV*	*EST*	*SE*	*CV*
< 1	2.4	0.61	25.9	0.7	0.7	100.0	1.5	0.23	14.9
$1 - 4$	4.9	1.3	26.2	6.3	1.2	18.4	5.6	1.0	18.0
$5 - 14$	12.4	1.9	15.2	14.7	2.5	16.2	13.5	1.9	14.3
$15 - 24$	11.6	3.1	26.4	14.5	2.5	17.2	14.2	2.2	15.3
$25 - 34$	15.5	1.3	8.4	15.4	1.7	11.0	15.4	1.3	8.5
$35 - 44$	16.1	1.2	7.2	9.6	1.5	15.6	15.8	1.2	7.5
$45 - 54$	9.8	1.0	10.2	7.7	2.0	26.7	8.8	1.1	12.7
$55 - 64$	4.2	1.5	37.2	3.6	1.0	27.4	3.9	1.1	29.2
$65 - 74$	3.6	0.6	16.6	4.6	1.1	22.9	4.1	0.78	18.8
≥ 75	1.5	0.53	34.1	0.71	0.34	47.5	1.1	0.3	27.7
Total									

Table 4.2.11

Percentage distribution of persons by different age-groups in Urban Areas of Kamrup District (Greg2 Estimates)

Age group (years)	Male			Female			Total		
	EST	*SE*	*CV*	*EST*	*SE*	*CV*	*EST*	*SE*	*CV*
< 1	1.7	0.6	33.3	2.0	0.6	30.3	1.8	0.2	13.3
$1 - 4$	4.1	1.3	31.4	6.5	1.1	17.4	5.3	1.0	18.9
$5 - 14$	18.0	1.4	8.0	17.7	2.4	13.8	17.9	1.8	9.9
$15 - 24$	19.6	2.4	12.3	20.0	2.5	12.5	19.8	1.9	9.6
$25 - 34$	18.0	1.3	7.3	18.4	1.7	9.1	18.2	1.3	7.1
$35 - 44$	18.2	1.2	6.7	15.8	1.5	9.8	17.0	1.3	7.4
$45 - 54$	10.8	1.0	9.3	10.1	2.1	20.5	10.5	1.1	10.8
$55 - 64$	6.3	1.6	24.9	5.6	0.9	16.1	6.0	1.1	18.3
$65 - 74$	3.6	0.6	17.3	4.1	1.0	25.7	3.9	0.8	20.4
≥ 75	1.0	0.5	50.3	0.7	0.3	48.8	0.9	0.3	35.9
Total									

Similar Tables have been prepared for the other district namely Cachar as well. To avoid repetitions we omit them here to save space.

5. Concluding Remarks

Let us first note the following published data on some of the characteristics covered in this report concerning some of the demographic features of the two districts of Kamrup and Cachar in Assam. They are quoted from Statistical Handbook, Assam, 2002 relating to the Population Census of India 2001 giving figures pertaining to April, 2001.

Looking at the two sets of data namely the population census 2001 figures relating to the rural and urban figures for the districts of Kamrup and Cachar and the comparable ones based on our investigation presented in the tables referring to December, 2003 there is only a rough agreement. To achieve a narrower level of discrepancies on the one hand, suitably projected figures relating to December, 2003 should be derived from the Population Census data and contrarily better use of symptomatic data should be made with a proper modelling to derive more accurate generalized regression estimates. In this investigation with limited resources further improvement could not be achieved. Some comparison is cited below.

Table 5.1

Comparison of total population among census projection and the sample survey figure of Kamrup District

Area	Projected Census	Traditional Estimate	Greg1 Estimate
Rural	1696858	2234340	2062871
Urban	879265	799118	661265
Total	2576123	3033458	2724136

Table 5.2

Comparison of total population among census projection and the sample survey figure of Cachar District

Area	Projected Census	Traditional Estimate	Greg1 Estimate
Rural	1282030	1621094	1513495
Urban	186098	192726	191823
Total	1468128	1813820	1705318

References

Statistical Handbook, Assam, 2002.
National Family Health Survey (NFHS - 2), 1998 – 99.

Acknowledgements

We are grateful to some of the Administrators in Indian Statistical Institue, Kolkata for their help in clearing bottlenecks at various stages.

The research of Dr. Arijit Chaudhuri is partially supported by CSIR Grant No.21(0539)/02/EMR-II.