# Investigating the Performance of Horvitz-Thompson and Generalized Regression Estimators in Small Domains with Application to Income and Expenditure of Households in Makurdi-Nigeria

**[1]Ikughur, A.J., [2]Imande, M.T. & [1]Vambe, S.**
[1]Department of Mathematics/Statistics/ Computer Science,
University of Agriculture, Makurdi-Nigeria
[2]Department of Mathematics/Computer Science,
Benue State University Makurdi-Nigeria

**Abstract**

In this study, efficiencies of Horvitz-Thompson (HT) and Generalized Regression (GREG) estimators are investigated in small domains namely Domain 1, 2 and 3. Efficiency criteria include the standard error of the estimates (SE) and coefficient of variation (CV) Computed for the study population of size 1052 households consisting of domain size of 885, 151 and 16 and domain samples sizes of 167, 29 and 3 respectively for Domains 1, 2 and 3. The estimates produced by HT estimator GREG estimator performs better than HT estimator with domain $SE_1(GREG) < SE_1(HT)$, $SE_2(GREG) < SE_2(HT)$ and $SE_3(GREG) < SE_3(HT)$. Similarly the domain Coefficient of Variation ($CV_{d^{th}}$) ;$CV_1(GREG) < CV_1(HT)$,$CV_2(GREG) < CV_2(HT)$ and $CV_3(GREG) < CV_3(HT)$. The study concluded that GREG estimator is more efficient than the HT estimator for small domain when it is clear that sample sizes of domains are small. The distribution of the study population sampled showed that Domain 1 with 84.1%of the respondents have an average monthly expenditure below₦21691.80(or less than about$2 per day World Bank poverty line).

**Introduction**

Small-Area Estimation (SAE) is a collection of statistical techniques designed for improving sample survey estimates through the use of auxiliary information. SAE offers the opportunity to improve accuracy of the current surveys through the use of statistical models, while avoiding additional respondents burden or survey costs, (Ghosh and Rao, 1994). Of course, SAE is of great importance in survey sampling due to a growing demand for reliable small area statistics from both public and private sectors for several reasons such as effective planning of health, mobilization and apportionment of resources and revenue and allocation of legislative seats as the case may be (Ghosh and Rao, 1994; Rao, 2003).

Sample sizes relating to small domains of interest turn out to be so small in practice that estimates based on them under classical sampling theory happen to be inaccurate and unreliable (especially with large measures of error in terms of large standard errors and coefficients of variation). Consequently, ways of achieving higher serviceability requires modifying the classical methods that utilizes various auxiliary pieces of information. There are several estimators that can be used in SAE when auxiliary variables are available. These include Hansen-Hurwitz estimator (HH), Rao-Hartley-Cochran estimator, Horvitz-Thompson (HT) estimator, Regression Estimator, Model-Dependent Empirical Best Linear Unbiased Predictor (EBLUP), synthetic, generalised regression (GREG) estimator etc.

In this study, we utilized Horvitz-Thompson (HT) estimator developed by Horvitz and Thompson (1952) and generalised regression (GREG) estimator developed by Sarndal, Swensson and Wretman (1992) to investigate and compare their efficiencies in estimating household expenditure in Makurdi-Nigeria considering income as auxiliary variable.

Makurdi is the capital city of Benue State and is located along the Benue River on latitude 7.73 degrees and longitude 8.52 degree. The town is mainly domiciled by those involved in Agriculture, Civil Servants and traders among others.

**Literature Review**

There are several studies in the literature (especially in Small Area) about GREG and HT estimators. Lehtonen (2006) investigated estimation for population subgroups or domains for model-assisted generalized regression (GREG) and Model-Dependent Empirical Best Linear Unbiased Predictor (EBLUP) estimators under different model choices and under unequal probability sampling. Merkouris (2009), on the other hand, explored the possibility of enhancing the precision of domain estimators by combining comparable information collected in multiple surveys of the same population. He proposed a regression method of estimation that is essentially an extended calibration procedure whereby comparable domain estimates from the various surveys are calibrated to each other. Santamaría *et al.* (2004) studied small area estimation with auxiliary variables, complex sampling designs and highlighted methods for small-area estimation including direct, synthetic and some other indirect estimators. Costa *et al.* (2006) worked on improving small-area estimation by combining surveys and investigating how to integrate information from an auxiliary survey in small-area estimation.

Pfeffermann (2013) presented some important developments in small area estimation methods. Rai and Pandey (2013) discussed the generalized class of synthetic estimators for estimating the population mean of small domains under the information of two auxiliary variables and the results of the numerical illustration for the two auxiliary variables were compared for the synthetic ratio estimator under single and two auxiliary variables.

Münnich *et al.* (2014) developed a design which adequately supports the application of small area estimation methods that foster an accurate estimation of total population counts. Ken (2001) showed that, under mild conditions, the variance of the Horvitz-Thompson (HT) estimator depends almost entirely on first order

inclusion probabilities. Using both purely design-based and model-assisted arguments, approximate expressions and estimators were derived for this "natural variance" of the HT estimator and tentative formulae for the HT variance estimator were also provided for the most important case where the HT variance does not take its natural value, namely when sampling is systematic from a deliberately ordered population.

Grakoviča(2006) compared Horvitz-Thompson (HT) estimator with GREG estimator in farm survey. According to him, GREG estimator depends on correlation among X (auxiliary variable which was total land area of farm) and Y (study variable denoting the number of cattle) and if correlation is small, GREG estimator is not better compared to HT estimator.

Milda (2009) used generalized regression, synthetic, empirical best linear predictor and Horvitz-Thompson estimators to examine estimation for domains and small areas for enter-prise surveys. Münnich *et al.* (2011), conducted the first census after the re-unification in Germany where a register-assisted census was implemented using population register data and an additional sample. They concluded that in terms of design-based methods, the design had considerable impact on the accuracy of the estimators and simple random sampling, as expected, is very inefficient using the HT and much better for the GREG.

Haslett *et al.* (2012) produced small-area estimates of poverty and malnutrition in Cambodia at district and sub-district levels by combining survey data with auxiliary data derived from the 2008 census. Wiese (2013) investigated the best design for the Swedish Investment Survey stratifying on number of employees, Neyman allocation, Stratified Simple Random Sampling (STSRS) and HT-estimation and the alternative method was to stratify on turnover, $\pi \rho s$-sampling and GREG-estimation. Bikauskaite (2014) analyzes several different types of techniques which produce small area estimates of poverty.

**Methods**

Let the population $\mho = u_1, u_2. . .u_N$ be partitioned into domains $D = 1,2, . . .d$ and let $y_{i,d}$ be the observation on the $i^{th}$ unit of the population in the $d^{th}$ domain.

Also let $P_i = \frac{x_i}{X}$; $X = \sum_{i=1}^{N} x_i$ be the selection probability of the $i^{th}$ unit in the population

The unbiased Horvitz-Thompson estimator for domain total of $Y_d$ is given as

$$t_H = \sum_{i \in s} y_{d,i} w_i I_{d,i} \qquad 1$$

where $w_i$ is the design weight. If $w_i = \frac{1}{\pi}$, then

$$t_d = \sum_{i \in s_d} \frac{y_i}{\pi_i} \qquad 2$$

which is an unbiased estimate of population in the $d^{th}$ domain. Since $E(t_d) = Y_d$

Let $I_{d,i} = \begin{cases} 1 & if \quad i \in s_d \\ 0 & if \quad i \notin s_d \end{cases}$

be a random variable, because it results from the random experiment of drawing a sample.

The variance of HT estimator $t_d$ in the $d^n$ domain is given as

$$V_{HT}(t_d) = \sum y_{di}^2 \left(\frac{1-\pi_i}{\pi_i}\right) + \sum_{i \neq} \sum_j y_{di} y_{dj} \left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j}\right) \quad 3$$

because $E(I_{sij}) = \pi_{ij} = \sum_{s \ni i.j} p(s)$, the inclusion probability of both $i$ and $j$ in a sample according to the design $p$, while $\pi_i$ is the inclusion probability with $E(I_i) = \pi_i$

In case every sample has a fixed number of units, each distinct, $V_{HT}(t_d)$ has the alternative form

$$V_{HT}(t_d) = \sum_{i<}^{N_d} \sum_j^{N_d} (\pi_i \pi_j - \pi_{ij}) \left(\frac{y_{di}}{\pi_i} - \frac{y_{dj}}{\pi_j}\right)^2 \quad 4$$

and if $\pi_{ij} > 0 \forall i,j \in U_d$ then,

$$V_{HT}(t_d) = \sum_{i<} \sum_{j \in s_d} \left(\frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}}\right) \left(\frac{y_{di}}{\pi_i} - \frac{y_{dj}}{\pi_j}\right)^2 \quad 5$$

is the unbiased estimator for $V_{HT}(t_d)$
Yates and Grundy (1953) and Sen (1953) gave equivalent formula for the HT variance
If p(s) > 0 is of fixed sample size, then $V_{HT}(t_d)$ and $\hat{V}_{HT}(t_d)$ have the following expression,

$$V_{HT}(t_d) = \sum_{U_d} \sum_{U_d} \Delta_{ij} \left(\frac{y_i I_{di}}{\pi_i} - \frac{y_j I_{dj}}{\pi_j}\right)^2 \quad 6$$

which can be estimated by

$$\hat{V}_{HT}(t_d) = \sum_{s_d}\sum_{s_d} \check{\Delta}_{ij}\left(\frac{y_i I_{di}}{\pi_i} - \frac{y_j I_{dj}}{\pi_j}\right)^2 I_{ij} \qquad 7$$

where$\Delta_{ij} = \pi_{ij} - \pi_i\pi_j$ , $\check{\Delta}_{ij} = \frac{\Delta_{ij}}{\pi_{ij}}$ $\forall i,j \in$
$U_d$(Chaudhuri, 2012)

**GREG Estimator**

The GREG estimator for $d^{th}$domain total is given by

$$t_{gd} = \sum_{i\in s_d}\frac{y_i}{\pi_i} + \hat{\beta}_{Qd}\left(X_d - \sum_{i\in s_d}\frac{x_i}{\pi_i}\right) \qquad 8$$

where

$$X_d = \sum_{i\in U_d} x_i, \quad \hat{\beta}_{Qd} = \frac{\sum_{i\in s_d} Q_i x_i y_i}{\sum_{i\in s_d} Q_i x_i^2}, \quad Q_i(>0)s_d$$

and $Q_i = \frac{1}{x_i}$,

For small n the variance of $t_{gd}$ is given by

$$V_g(t_{gd}) = \sum_{i\in s_d} e_i^2\frac{1-\pi_i}{\pi_i^2} + \sum_{i\neq}\sum_j e_i I_{di} e_j I_{dj}\frac{\pi_{ij}-\pi_i\pi_j}{\pi_{ij}} \qquad 9$$

or

$$V_g(t_{gd}) = \sum_{i<}\sum_{j\in s_d}\left(\frac{\pi_i\pi_j-\pi_{ij}}{\pi_{ij}}\right)\left(\frac{e_i I_{di}}{\pi_i} - \frac{e_j I_{dj}}{\pi_j}\right)^2 \qquad 10$$

where$e_i = y_{di} - \hat{\beta}_{Qd}x_i$

**Theorem:** For the estimates of population mean, the variance is given as

$$\hat{V}(\hat{\bar{t}}_d) = V\left(\frac{t}{N}\right) = \frac{1}{n^2}V(t) \qquad 11$$

where n is the sample size

**Measure of Efficiency**

Basically, an estimator $t_{gd}$ is more efficient than another estimator $t_d$ if and only if $V_g(t_{gd}) < V_{HT}(t_d)$. Alternatively, the relative efficiency (RE) criteria is used and is defined as

$$RE = \frac{V_g(t_{gd})}{V_{HT}(t_d)} < 112$$

or

$$RE = \frac{V_g(t_{gd})}{V_{HT}(t_d)} \times 100 < 100 \qquad 13$$

whichcan also be expressed as

$$V_g(t_{gd}) - V_{HT}(t_d) < 0 \qquad 14$$

so that$t_{gd}$ is more efficient than $t_d$

Other criteria for detecting the more efficient estimator include Standard Error (SE) and Coefficient of Variation (CV) defined as

$$S.E(t_{g_d}) = \sqrt{V_g(t_{gd})} \qquad 15$$

$$C.V(t_d) = \frac{S.E(t_d)}{t_d} \times 100 \qquad 16$$

$$C.V(t_{g_d}) = \frac{S.E(t_{g_d})}{t_{g_d}} \times 100 \qquad 17$$

where$S.E(t_d) = \sqrt{V_{HT}(t_d)}$ $\qquad 18$

**The Study Population and Samples**

Data for this study was obtained from field surveys conducted on household income and expenditure pertaining to Civil Servants in Makurdi and environs by administering structured questionnaire to one thousand and fifty two randomly selected households.

The variables in the data are: Y, average monthly household income per month and X, average household expenditure per month. The correlation coefficient for the study and auxiliary variable is 0.966; The sample size utilized in this study is determined using the expression:

$$n = \frac{N\hat{p}_{srs}\hat{q}_{srs}}{\frac{d^2}{1.96^2}(N-1)+\hat{p}_{srs}\hat{q}_{srs}} \qquad 19$$

where $\hat{p}_{srs}$= the estimated proportion = 0.8, $\hat{q}_{srs} = 1 - \hat{p}_{srs} = 0.2$, $d^2$= desired absolute precision (degree of precision) = 0.05, N= 1052. The sample size for each domain is then allocated using proportional to population size.

$$n_{d^{th}} = \frac{n}{N}N_d \qquad 20$$

WhereN₁= population size of domain 1, N₂= population size of domain 2, N₃= population size of domain 3, $n_1$= sample size of domain 1, $n_2$= sample size of domain 2, $n_3$= sample size of domain 3.

According to Renaissance Capital (2011), the Nigerian Middle Class average monthly income is in the range ₦75,000 - ₦100,000. Thus, for the study population under consideration, the data was grouped into three domains: below ₦75,000 as lower income (or domain 1), ₦75,000 – ₦100,000 as middle income (or domain 2) and above ₦100,000 as high income (or domain 3).

**Results**

The domain sizes and their respective sample sizes proportional to their domain sizes are given in Table 1.Table 2 below shows the total and aveage expenditure of persons in the three small domains in Makurdi using the Renaissance classifications of 2011when information about their incomes are available. Tables 3 and 4 show the results of the efficiency measures.

**Table 1:** Domain Sizes and their Respective Sample Sizes

| Domain | $N_d$ | $n_d$ |
|---|---|---|
| 1 | 885 | 167 |
| 2 | 151 | 29 |
| 3 | 16 | 3 |

**Table 2:** Estimates of Total Expenditures, Averages and Variances using HT and GREG Estimators.

| | Domain | HT | GREG |
|---|---|---|---|
| $\hat{y}$ | 1 | 20081766.72 | 18255724.60 |
| | 2 | 9435671.27 | 8607876.36 |
| | 3 | 1795022.97 | 1675787.97 |
| $\hat{\bar{y}}$ | 1 | 22691.26 | 20627.94 |
| | 2 | 62487.89 | 57005.80 |
| | 3 | 112188.94 | 104736.75 |
| $V(\hat{\bar{y}})$ | 1 | 10979832.49 | 4721163.20 |
| | 2 | 467438218.62 | 125850877.95 |
| | 3 | 16350042919.67 | 681525145.10 |

**Table 3:** The Relative Efficiency of the Estimators

| Domain | $V(t_{gd})$ | $V(t_d)$ | RE(%) |
|---|---|---|---|
| 1 | 4721163.20 | 10979832.49 | 42.9985 |
| 2 | 125850877.95 | 467438218.62 | 26.9235 |
| 3 | 681525145.10 | 16350042919.67 | 4.1683 |

**Table 4:** Standard Error and Coefficient of Variation of the Estimates

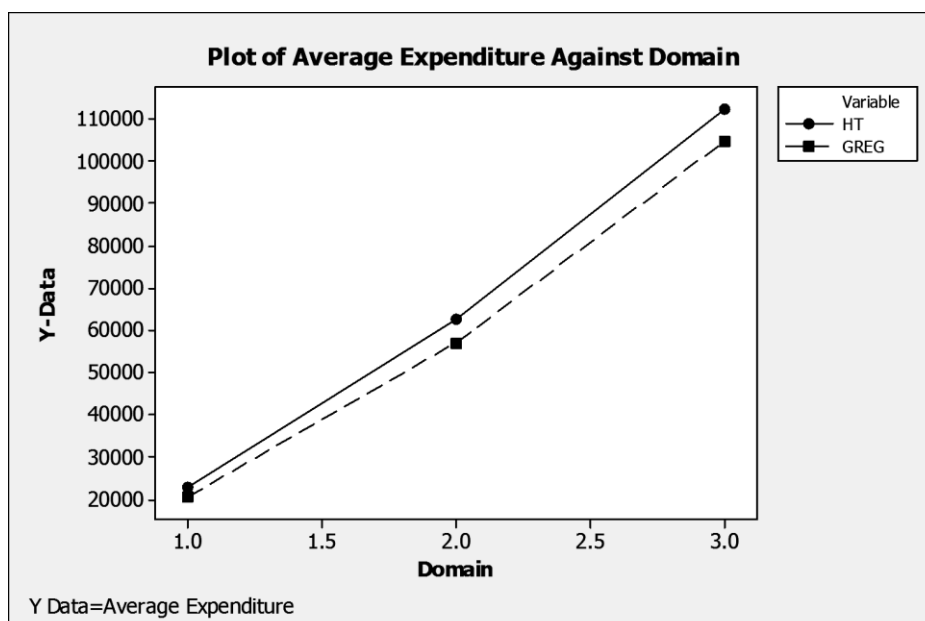| | Domain | HT | GREG |
|---|---|---|---|
| S.E | 1 | 3313.58 | 2172.82 |
| | 2 | 21620.32 | 11218.33 |
| | 3 | 127867.29 | 26106.04 |
| C.V | 1 | 14.6029 | 10.5334 |
| | 2 | 34.5992 | 19.6793 |
| | 3 | 113.9749 | 24.9254 |

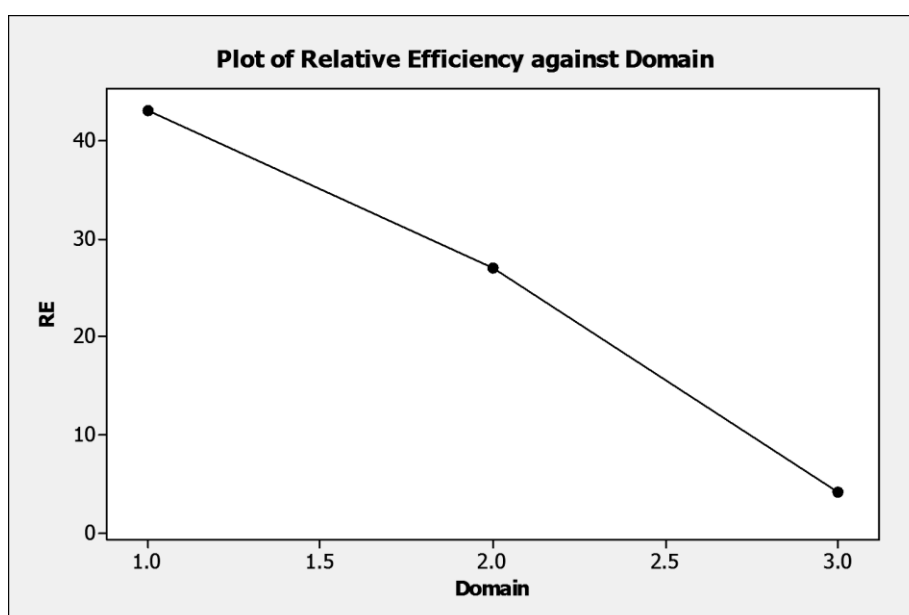**Figure 1:** Plot of Average Expenditure for the three domains



**Figure 2:** Plot of Relative Efficiency against the Domain

**Discussion of Results**

This section discusses the results of data presented in the previous section as shown in Tables 1, 2, 3and 4 as well as Figures1and 2.Table 1 shows the domain sizes and their respective sample sizes. Table 2 shows the estimated total expenditures, average expenditures and also their variances for the three domains using Horvitz-Thompson and GREG estimators.

The estimates of average expenditure in the three domains in Table 2 reveals that 885 out of 1052 (84.1%) of the respondents have average monthly expenditure of ₦22,691.26, 14.4% of respondents have average monthly expenditure

of ₦62,487.89 while 1.5% have average monthly expenditure of ₦112,188.94 when Horvitz-Thompson estimator is used. When GREG estimator is used in estimation, the average monthly expenditure for the domains are ₦20,627.94, ₦57,005.80 and ₦104,736.75 for domain 1, domain 2 and domain 3 respectively exhibiting similar pattern with the HT estimator. The population distribution of income earners has a wider base for lower income and smaller peak for highest income group.

The estimated variances for each domain and hence the standard error (SE) of the estimates and the coefficient of variations, show that when HT estimator is utilized, domain 1

gives S.E=3,313.58 and C.V= 14.6029%, domain 2 gives S.E=21,620.32 and C.V= 34.5992% while domain 3 gives S.E=127,867.29 and C.V=113.9749%.

Thus, estimates in domain 1 appear to be more precise than those of other domains. It is also noticeable that the SE for domain 3 is far larger than others and has a CV of 113.9749%. This can be attributed to the size of the sample which is actually small.

When we consider the GREG estimator as shown in Table 4,it is obvious in domain 1, 2 and 3 that the SE's(i.e2172.82, 11,218.33 and 26,106.04)and CV(i.e10.5334%, 19.6793% and 24.9254%)   are increasing from domain 1 to domain 3 The behaviour of these estimates is similar to those of HT estimates and they agree with the findings of Lehtonen (2006) and Munnich *et al.* (2011) who concluded that the accuracy of estimates is improved with increase in sample size as shown by the results.

However when we compare the performance of the two estimators, it is clear that GREG estimator is more efficient than HT estimator as shown in Table 3. Generally, the results are consistent with the literatures that posited that when some auxiliary or design variables are available, the model-based estimator, that utilizes the relationship between the survey outcome and these auxiliary or design variables, can improve the efficiency compared to the design-based estimators. The same way, the result agrees with that of Siobhan (2003) who noted that the model assisted procedure is much more efficient than the Horvitz-Thompson estimator. This is also shown by Tikkiwal and Pandey (2007), Pandey and Tikkiwal (2010), and Pandey (2010), who also established that when an auxiliary variable is closely related with the variable under study, the small area estimators based on auxiliary information perform better than those which do not use auxiliary information.

Similarly, the result, also agrees with the findings of Statistics Sweden (2008) on the disadvantage of HT Estimator, that, it is not the most efficient estimator when compared with model based design. In other words, the variance for an HT-estimation is sometimes unnecessary big. Similarly Holmberg (2003) who investigated the combination of probability proportional and GREG argued that adding auxiliary information to a survey will highly improve the quality of the estimators. According to him, using an estimator outside the GREG family may probably not reduce the variance. In general, the variance of the design-based estimators is larger than the model-based and HT estimator is a member of design based estimators.

Generally, the survey results reveal that out of the 1052 respondents about 84% of respondents expend an average of between ₦665.00 and ₦731.00 per day for the lowest income class identified as the "Poor" population, about 14% of members of the study population earn what is termed "Average" income while less than 2%of members of the population earn what is termed "Highest" income. The implication of the lower average expenditure in terms of the $2per day (with ₦350.00 per Dollar exchange rate) World Bank poverty line indicates that the average income of persons in the study domains is not capable of meeting the needs of a household of two or more family members. Similarly, most of the households fall below the middle class in terms of household income and expenditure as defined by Renaissance Capital (2011).

**Conclusion and Recommendation**

Following the results from the analysis, we can conclude that in order to have quality estimates of population characteristics in small domains, the use of auxiliary variables that are positively correlated is necessary more so, GREG estimator provides better estimates of population characteristics than Horvitz-Thompson estimator just as Särndal (2007) also pointed out. The GREG estimator although is nearly unbiased, it has a considerably smaller variance than the Horvitz-Thompson estimator.

Furthermore, the estimates for the sample population show that over 80% members of the

study population live below the World Bank's poverty line of $2per day. The implication is that it tends to have negative effects on the socio-economics and even cultural values in such community and the domains of study.

It is therefore worth to recommend that when estimating population parameters for small domains, it is advisable to use auxiliary information and GREG estimator, a model based estimator rather than the design based HT estimator. It further recommends the need to narrow income gap to increase household ability to spend so as to reduce poverty especially among the lower class, an effort which stakeholders need to take deliberate steps to achieve.

**References**

Bikauskaite A. (2014) Estimation of Poverty in Small Areas.Statistika, Eurostat, Luxembourg94 (4): 43-54

Chaudhuri A. (2012). Developing small domain statistics: Modelling in survey sampling. 1- 44

Costa, A., Satorra, A., and Ventura, E. (2006).Improving small area estimation by combining surveys: new perspectives in regional statistics. SORT (Statistics and Operations Research Transactions), 30: 101–122.

Pfeffermann, D (2013). New Important Developments in Small Area Estimation. Southampton Statistical Sciences Research Institute, University of Southampton, Southampton, SO17 1BJ, United Kingdom, and Department of Statistics, Hebrew University of Jerusalem, Jerusalem. Statistical science, 28:40-68

Ghosh, M. and Rao, J. N. K. (1994). Small area estimation: An appraisal (with discussion). Statistical Science, 9: 65–93

Grakoviča O. (2006) Analysis of GREG Estimator in Farm Survey.The Central Statistical Bureau of Latvia University of Latvia.e-mail:olga.grakovica@csb.gov.lv1 -15Grazyna Dehnel, ElzbietaGolata (2006) Attempts to Estimate Basic Information for Small Business in Poland. *Statistics in Transition,* **7**(4): 785—803

Haslett S. A., Jones G. and Alison S. (2012) Small-area Estimation of Poverty and Malnutrition in Cambodia. Institute of Fundamental Sciences–Statistics Massey University New Zealand. 1- 50

Holmberg, A. (2003) Essays on model assisted survey planning. Ph.D thesis. Comprehensive the Faculty of Social Sciences 126, department of statistics. ActaUniversitaisUpsaliensis, Uppsala. 1-83

Horvitz, D.G. and Thompson, D.l. (1952).A generalization of sampling without replacement from finite-universe. *Journal American Statistical Association*, 47:663-685

Ken B. (2001) Deriving and Estimating an Approximate Variance for the Horvitz Thompson Estimator Using Only First Order Inclusion Probabilities. Australian National University. Australian Capital Territory 0200, Australia. Ken.Brewer.anu.edu.au. 1417 - 1422

Lehtonen R. (2006) The Role of Models In Model-Assisted And Model-Dependent Estimation for Domains and Small Areas. University of Helsinki, Finland e-mail: risto.lehtonen@helsinki.fi. 1-10

Merkouris, T. (2009), "Efficient small-domain estimation by combining information from multiple surveys through regression," Working paper, HSMD-2006-007E, Statistics Canada, 3419-3425.

MildaŠličkutė-Šeštokienė(2009) Estimation for Domains and Small Areas. Statistics Lithuania, Lithuania e-mail: milda.slickute-sestokiene@stat.gov.lt56- 78

Mohammadi M. and M. Salehi M. (2012), Horvitz-Thompson Estimator of Population Mean under inverse Sampling Designs Bulletin of the Iranian Mathematical Society. **38**(2): 333-347.

Münnich R., Jan Pablo Burgard, Siegfried Gabler, Matthias Ganninger, and Jan-Philipp Kolb (2011), Small Area Estimation in the German Census 2011.*Statistics in Transition new series and Survey Methodology Joint Issue: Small Area Estimation***17**(1): 25–40

Pandey, K. K. (2010), Aspects of small area estimation using auxiliary information". 1st edition, Germany, VDM VerlagDr. Muller GmbH & Co. KG,. (ISBN: 978-3-639-31569-1). 1-209

Pandey, K. K. and Tikkiwal, G. C., (2010).,Generalized class of synthetic estimators for small area under systematic sampling design" *Statistics in Transition-new series*, Poland, **11**(1): 75-89

Rai P. K. and Pandey K. K. (2013).Synthetic Estimators Using Auxiliary Information In Small Domains. Statistics In Transition-new series. **14**(1): 31–44

Rao, J.N.K. (2003). Small Area Estimation. New York: Wiley. pp1-115

Renaissance Capital (2011), A survey of the Nigerian middle class. 1-36 http://research.rencap.com/eng/default.asp or http://research.rencap.com, 26 September 2011

Santamaría, L., Morpales, D., and Molina, I. (2004).A comparative study of small area estimators. SORT, **28**:215–230

Sarndal, C.-E., B. Swensson, and J. Wretman (1992).Model Assisted Survey Sampling. Springer-Verlag Inc., New York 1-152

Sen, A.R.(1953). Present status of Probability Sampling and its use in estimation of farm characteristics. *J. Indian Soc. Agric. Statist.* 5:119-127.

Siobhan E. S. (2003). Nonparametric Survey Regression Estimation in two-Stage Spatial Sampling. Master's Project, Department of Statistics, Colorado State University, Fort Collins, COSpring 2003. 1-51

Statistics Sweden.(2008) Urval–frånTeori till Praktik SCB-Tryck, Örebro 2008. 03. *Statistics Sweden, Research and Development.* Pp 1-59

Tikkiwal, G. C. and Ghiya, A., (2000). A Generalized Class of Synthetic Estimators with Application to Crop Acreage Estimation for Small Domains. *Biometrical Journal*, 42 (7): 865-876.

Tikkiwal, G. C. and Pandey, K. K., (2007).On Some Aspects of Small Area Estimation Using Auxiliary Information. Ph.D. Thesis. Department of Mathematics and Statistics, J. N. V. University Jodhpur Rajasthan. 1-71

Wiese Linda (2013) Stratification, Sampling and Estimation: Finding the best design for the Swedish Investment Survey. Stockholm University. Department of Statistics. 1- 78

Wilford, B. M. (2011). Sample design for small area estimation. Doctor of philosophy thesis centre for statistical and survey methodology, University of Wollongong. 1-19.

Yates, F.and Grundy, P.M. (1953). Selection without replacement from within strata with Probability Proportional to Size. *J. Roy. Statist. Soc,* (B) 15:253-261.