



**SCHOOL OF STATISTICS**  
UNIVERSITY OF THE PHILIPPINES DILIMAN



## WORKING PAPER SERIES

### **Model-Based Predictive Estimation in Samples With Frame Problems**

by

Erniel B. Barrios<sup>1</sup>

UPSS Working Paper No. 2008-04  
October 2008

School of Statistics  
Ramon Magsaysay Avenue  
U.P. Diliman, Quezon City  
Telefax: 928-08-81  
Email: [updstat@yahoo.com](mailto:updstat@yahoo.com)

<sup>1</sup>Professor, School of Statistics, University of the Philippines Diliman, Diliman, Quezon City

## **Abstract**

Probability sampling in finite populations are completely dependent on the availability of a reliable frame. Though frame updating is commonly done prior to sample selection, there are many instances where accessibility of the field location of the sampling units causes the discrepancy between the sampled population and the frame. This causes doubts on the appropriateness of design-based estimates. Under some simulated frame problems, model-based estimation is viewed in the context of re-sampling methods to estimate the population total. Even if the sample is drawn only from the middle 50% of the population distribution, model-based estimates are superior or at least comparable to design-based estimates especially for small populations. In symmetric populations, the choice of an auxiliary variable (predictor) is important but in a skewed population, performance of model-based estimator is robust to the relationship between the target variable and the auxiliary predictor. The bootstrap sampling errors are generally lower than the design-unbiased sampling errors.

## **Introduction**

The literature of finite population sampling is dominated by the design-based estimation framework. Kalton(2000), summarized the developments in design-based estimation and the model-assisted design-based framework that adopts certain features of model-based estimation. Although the design-based framework of statistical inference in finite populations is generally optimal and acceptable in the statistical community, it depends so much on the availability of a reliable frame. The sample selection process, field identification of the sampled units, and computation of the sampling weights, are completely dependent on the frame.

Both the dynamism in the target variables and the sampling units itself generally threatens the validity of the frame. In human population units, migration easily causes the frame to be outdated. In business establishments, mergers, integration, diversification of the business

would also cause the frame to be outdated. In rural areas, accessibility of the sampling units would also create a significant discrepancy between the target population and the sampling population.

What will happen if the actual sampling population is different from the target population? In development studies, whenever the sampling population deviates from the target population, a new population is defined or limitations and constraints are explicitly declared. It is however, a waste of information if some auxiliary variables are not used in generating any information about the excluded segment of the target population.

Some examples of symmetric distribution having coverage problem would include sales of firms where those with lower sales may not cooperate to mask problems with competition or those with higher sales will not cooperate to shield them from taxation. In a skewed distribution like income, those in the lower income and upper income are difficult to locate or are hard to convince to cooperate to the survey. This results to sample selection only from among the middle-income group.

There are a lot of literatures on issues relating to various frame problems. Repeated surveys, or specifically those that use rotating samples are very susceptible to frame problems. Optimum selection from among the new addition to the population is better than treating both the new and continuing units equally (Hughes, 1991). Birth of new sampling units usually arises when a list from an external source is merged into the existing frame to be updated. Dual frames and multiple frames in general, is usually addressed through independent

selection of samples from each frame, then estimation from the pooled samples are suggested, example pseudo-MLE by Skinner and Rao (1996). In the context of sample selection, Hedayat and Robieson(1998), under regularity conditions, proved that a SRS drawn from a population where “undesirable” units are assigned zero probabilities is equivalent to SRS from the desirable units only. The implication is that when separate data collection is required for a small/segment of the population, so it can be treated as certainty units, then selection from the majority of the units will not violate any rules in probability sampling.

The literature on model-based estimation is also lengthy. (Zacks, 2002) presented a good summary of model-based estimation. The review specifically focused on those that follows Basu’s framework that practically ignored the sampling design in the analysis. This is a different framework from the one presented by Kalton(2000) that is model-assisted but still invokes the design in the analysis. In 1970’s the design-based framework was challenged by model-dependent framework, but design-based model-assisted framework prevailed (Kalton, 2000). Increasing coverage error however, is posing potential problems in weight construction among complex survey designs. Several imputation procedures are available, but problems on the consistency and efficiency of the estimates continue to prevail.

### **On Model-Based and Predictive Estimation**

A sample selection process essentially dichotomizes the target population into the sampled segment and the non-sampled segment. Predictive estimation then forms an estimate of the total by adding the sampled values to the predicted total of the non-sampled part. Prediction

of the total in the non-sampled part utilizes the sample observations. Minimum MSE estimate of the total of the non-sampled part is just the mean of the sample multiplied by the total number of non-sampled units (similar to design-unbiased estimate). The Bayesian estimator, for squared error loss function is given by  $\hat{T}_B = \sum_{i \in S} y_i + \sum_{i \notin S} E\{y_i / (S, y_S)\}$ , where  $(S, y_S)$  is the observed sample (Zacks, 2002).

The superpopulation approach assumes that the population values  $y_1, y_2, \dots, y_N$  is a realization of some random process. Furthermore, there is an infinite superpopulation of random vectors, and that the unobserved part is predicted from the model to make inference (Zacks, 2002). The empirical characterization of the model is usually aided by the sampled part.

The model-based approach has been compared to design-based procedures and has its own limitations. Model-dependent estimates depend on the validity of the postulated model or robustness of the estimates to model failure (Kalton, 2000). The superpopulation model however, looks like a Bayesian implementation of the surveyor's background knowledge and information (Zacks, 2002).

There is an increasing interest in Bayesian predictive modeling leading to a significant contribution to model-based estimation. Rao (1999) summarized some model-based techniques in small area estimation including empirical best linear unbiased prediction (EBLUP), empirical Bayes (EB), and hierarchical Bayes (HB). Chattopadhyay and Datta (1994) used EB and HB in the simultaneous estimation of the means from different strata in

an error-in-variable superpopulation model (measurement error). EB and HB outperformed sample strata means.

### **An Alternative View on Model-Based Predictive Estimation**

Model-based estimation can be formulated in the context of resampling. Suppose a sample from the population is given:  $y_1, y_2, \dots, y_n$ . The population total is decomposed into:

$$Y = \sum_{i \in S} y_i + \sum_{i \notin S} y_i .$$

The first term is known, and to compute the estimator for the total, the total

of the non-sampled part should be estimated. Given an auxiliary variable  $x$ , a model

$$y = f(x, \theta) + \varepsilon$$

can be fitted using the sample observations. The auxiliary variable  $x$  must be

measured in the sample and should be available for the non-sampled part. The model can be

used in predicting  $y$  of the non-sampled part by substituting  $x$  into the fitted model. The

$$\text{estimate is then } \hat{Y} = \sum_{i \in S} y_i + \sum_{i \notin S} \hat{y}_{i \text{ predicted}} .$$

The estimator can be viewed as recreation of the

population, units are composed of actual  $y$ -values and of predicted  $y$ -values. This “artificial”

population is parallel to the one generated in resampling. Thus, a nonparametric bootstrap can

be used in estimating the population characteristics including the sampling errors.

For a parameter  $\theta$ , let the estimate be  $\hat{\theta}_i$   $i = 1, 2, \dots, k$  for the  $i^{\text{th}}$  replicate drawn from the

reconstructed population. The bootstrap estimate is  $\hat{\theta} = \frac{1}{k} \sum_{i=1}^k \hat{\theta}_i$ . The standard error of  $\hat{\theta}$  is

estimated from  $s.e.(\hat{\theta}) = \sqrt{\frac{1}{k-1} \sum_{i=1}^k (\hat{\theta}_i - \hat{\theta})^2}$ . Note that the monte carlo estimate of the

population total coincide the total of the recreated population for a reasonably sized replication.

The estimation procedure can also handle nonresponse and missing data. The important requirement of the estimation procedure is the availability of the auxiliary information  $x$  for all sampling units in the target population. Even if the frame does not contain information on  $x$  for all units, for as long as it can be generated with minimal cost requirement, the procedure still applies. The appropriateness of the postulated model is also important.

### **Simulation Scenarios**

Two basic shapes of the distribution will be simulated. A symmetric distribution can best represent sampling populations where the variable of interest may include sales, agricultural production, etc. On the other hand, a skewed distribution can represent sampling populations where the target variable is a count.

The shape of the sampling population can possibly influence the distance between the estimated total and the true total. The normal distribution will represent the symmetric distribution, while poisson represents the skewed distribution. In each distribution, homogeneous and heterogeneous cases will be simulated. The part of the frame that prevents access to sampling units may also affect the estimate and to this effect, four sampling scenarios will be simulated. S1 is when simple random selection (SRS) will be made only in the bottom 25% of the population, S2 SRS in the middle 50%, S3 SRS in the upper 25% and

S4 SRS from the whole population. S1 and S3 are extreme cases, but it will help in the assessment of the performance of the estimation procedure in extreme cases. S2 should represent a typical frame constraint, though a little bit exaggerated. Normally, the problems with the frame would affect only small proportions at the tails, so that if the procedure performs well in S2, then most probably it will perform well in many cases. Sampling rates were simulated at 1%, 3%, 5%, and 10%.

Only one auxiliary variable is used. The relationship between  $y$  and  $x$  is controlled through the simulation of  $y$  and the error term and computing the values of  $x$ . A multiplier on the error component would pull down the prediction ability of  $x$  on  $y$  and this serves as the control on the relationship. For the symmetric population, normal regression model was used, while poisson regression with log link function is used in the skewed population.

## **Results and Discussions**

The percent difference between the estimated total from the true total are summarized in Tables 1 to 4.

In a homogeneous normal population, the sampling rate have minimal effect on the closeness of estimates to the true total. There is also a minimal difference between small and large populations. A strong association between  $y$  and  $x$  ensures a minimal distance between the estimate and the true total, especially in large populations. Even in the worst case of the relationship between  $y$  and  $x$ , the distance is generally short (less than 10%). There is expectedly an underestimation in S1 and overestimation in S3. S2 and S4 are comparable to the design-unbiased estimates of the population total.



In the heterogeneous normal population, the sampling rate affected only the design-unbiased estimates in small population. The role of the relationship between  $y$  and  $x$  is highlighted on the distance between the estimate and the true population total. The estimates from S2 even for low  $R^2$  cases are close to the true values especially in large populations. Even in small populations with low  $R^2$ , estimates in S2 are still very close to the true totals. Estimates in S1 and S3 are good only when  $R^2$  is high. Estimates from S2 and S4 are better than the design-unbiased estimates.

In skewed populations, the sampling rate has no effect on the distance between the estimates and the true population totals. The population size and the relationship between  $y$  and  $x$  also have minimal effect on the distance. S2, S4 and the design-unbiased estimates are generally comparable. In small population however, the estimates from S2 are superior.

Generally, the bootstrap estimates of the standard errors are lower in the proposed estimation procedure than in the design-unbiased estimates resulting to very low coefficient of variation (see Tables 5-8).

## **Conclusions**

The estimation procedure performs well in small populations. Even for large populations, the estimates are comparable to the design-unbiased estimates. Even if samples are drawn only from the middle 50% of the distribution, regardless of the shape, model-based predictive estimate is either superior or comparable to the design-unbiased estimates. In symmetric

populations, a good choice of the auxiliary variable is important but not so in skewed populations.

The common problems with the frame that contaminate measurement or prevent access/identification of units in the tails/extreme parts can be resolved through model-based predictive estimation.

If frame problems occur, it is important that an auxiliary variable ( $x$ ) is available or can be easily measured for units where access to the sampling unit to measure the target variable is not feasible. The relationship between  $y$  and  $x$  in the sample can be used in recreating the population to perform nonparametric inference in a finite population.

## **References**

Kalton, G, 2000, Developments in Survey Research in the Past 25 Years, Survey Methodology, Vol. 26, No. 1, pp. 3-10.

Zacks, S, 2002, In the Footsteps of Basu: The Predictive Modelling Approach to Sampling From Finite Population, Sankhya: The Indian Journal of Statistics, Vol. 64 Ser. A, No. 3, pp. 532-544.

Rao, J, 1999, Some Recent Advances in Model-Based Small Area Estimation, Survey Methodology, Vol. 25, No. 2, pp. 175-186.

Chattopadhyay, M and Datta, G, 1994, Finite Population Prediction For Stratified Sampling Under Error-in-Variables Superpopulation Models, Sankhya: The Indian Journal of Statistics, Vol 56, Ser. B, No. 3, pp. 415-433

Hughes, P, 1991, Design for Composite Estimation with Changing Survey Frames, Journal of Official Statistics, Vol. 7 No. 1, pp. 77-91

Skinner, C, and Rao, J, 1996, Estimation in Dual Frame Surveys with Complex Designs, Journal of the American Statistical Association, Vol. 91, No. 433 (Theory and Methods), pp. 349-355

Hedayat, A, and Robieson, W, 1998, Exclusion of an Undesirable Sample from the Support of a Simple Random Sample, The American Statistician, Vol. 52, No. 1, pp. 41-43.

Table 1: Percent Difference of Estimated Total from Actual Total in Homogeneous Normal Population:  $N(100,25)$

	$R^2$	96.48	96.48	96.48	96.48	52.05	52.05	52.05	52.05	21.08	21.08	21.08	21.08
	Sampling Rate	1%	3%	5%	10%	1%	3%	5%	10%	1%	3%	5%	10%
N=1,000	Sample 1	2.46	1.59	0.84	0.64	4.96	5.07	5.62	5.40	6.87	6.63	6.94	6.68
	Sample 2	-0.30	-0.09	0.01	0.04	0.86	-0.33	0.56	-0.91	-1.22	-0.35	-0.02	-0.59
	Sample 3	-0.95	-0.96	-1.07	-1.01	-10.44	-6.97	-6.52	-6.83	-9.72	-7.95	-7.60	-7.62
	Sample 4	-0.08	-0.11	0.15	-0.06	-0.83	-0.49	-0.10	0.43	-0.53	-1.57	0.54	0.16
	Design-SRS	-1.04	1.56	0.60	0.76	-0.89	-0.34	0.00	0.23	-0.81	-0.83	1.01	0.16
	$R^2$	96.17	96.17	96.17	96.17	49.39	49.39	49.39	49.39	19.07	19.07	19.07	19.07
	Sampling Rate	1%	3%	5%	10%	1%	3%	5%	10%	1%	3%	5%	10%
N=10,000	Sample 1	0.75	0.70	0.71	1.10	4.44	4.91	4.77	4.88	5.77	5.69	5.55	5.44
	Sample 2	0.00	0.14	0.04	0.06	-0.29	-0.11	0.21	-0.15	0.06	-0.25	-0.14	0.07
	Sample 3	-0.89	-0.49	-0.74	-0.70	-5.26	-4.89	-4.98	-5.07	-6.36	-5.36	-5.84	-5.68
	Sample 4	0.08	0.12	-0.04	0.00	-0.14	-0.18	-0.28	-0.05	-0.32	0.54	0.23	-0.05
	Design-SRS	0.02	0.09	0.02	0.15	-0.67	-0.11	-0.21	0.09	0.03	0.54	-0.08	0.03

Table 2: Percent Difference of Estimated Total from Actual Total in Heterogeneous Normal Population: N(100,2500)

N=1,000	R <sup>2</sup>	99.96	99.96	99.96	99.96	52.03	52.03	52.03	52.03	21.08	21.08	21.08	21.08
	Sampling Rate	1%	3%	5%	10%	1%	3%	5%	10%	1%	3%	5%	10%
Sample 1	0.48	-0.45	0.90	0.49	37.92	52.84	46.18	49.05	61.25	70.70	67.15	60.97	
Sample 2	-0.28	-0.10	0.20	0.14	-5.30	-5.56	2.38	-1.00	5.03	-2.59	-0.99	-2.62	
Sample 3	0.08	0.27	-0.33	-0.23	-43.12	-52.95	-46.04	-50.01	-62.30	-52.78	-64.83	-65.02	
Sample 4	-0.22	-0.21	-0.22	0.00	-0.15	12.66	-0.66	2.49	-19.41	-1.53	0.55	2.74	
Design-SRS	59.88	4.76	0.50	-4.56	-9.63	11.14	-6.29	1.93	-20.22	-2.78	4.53	1.88	
N=10,000	R <sup>2</sup>	99.96	99.96	99.96	99.96	49.39	49.39	49.39	49.39	19.07	19.07	19.07	19.07
	Sampling Rate	1%	3%	5%	10%	1%	3%	5%	10%	1%	3%	5%	10%
Sample 1	-0.02	0.34	0.02	-0.02	50.29	52.34	48.56	47.43	60.13	59.28	59.49	60.23	
Sample 2	-0.18	-0.01	-0.07	-0.09	3.08	1.82	-1.02	0.07	-0.18	0.25	0.47	-0.56	
Sample 3	-0.06	0.25	0.12	-0.05	-48.85	-52.54	-50.95	-52.37	-65.03	-57.60	-60.97	-61.22	
Sample 4	-0.18	-0.06	-0.04	-0.01	3.90	-0.06	-2.33	-1.11	-2.78	-3.26	-3.74	2.56	
Design-SRS	-1.67	-2.74	1.10	-1.00	-1.10	-2.70	0.79	-0.29	-0.47	-2.71	-4.79	2.97	

Table 3: Percent Difference of Estimated Total from Actual Total in Low-Mean Poisson Population: Po(10)

N=1,000	P-Value ( $\hat{\beta}$ )	<.0001	<.0001	<.0001	<.0001	0.0441	0.0441	0.0441	0.0441	0.3454	0.3454	0.3454	0.3454
	Sampling Rate	1%	3%	5%	10%	1%	3%	5%	10%	1%	3%	5%	10%
Sample 1	35.99	29.49	35.01	30.83	31.29	39.25	35.78	34.75	40.58	34.47	31.19	36.57	
Sample 2	-2.58	-3.31	-0.51	0.20	1.67	0.71	0.54	0.29	-5.60	-0.17	1.40	0.39	
Sample 3	-43.42	-41.26	-37.36	-39.01	-31.81	-38.15	-39.62	-39.30	-42.60	-42.97	-42.52	-37.50	
Sample 4	15.08	-9.75	59.47	-2.26	-11.86	4.24	-1.49	0.12	1.16	2.70	10.17	2.21	
Design-SRS	8.75	-9.63	0.33	-1.98	6.75	4.74	-0.27	-0.37	-0.27	2.74	10.16	2.24	
N=10,000	P-Value ( $\hat{\beta}$ )	<.0001	<.0001	<.0001	<.0001	0.0202	0.0202	0.0202	0.0202	0.5415	0.5415	0.5415	0.5415
	Sampling Rate	1%	3%	5%	10%	1%	3%	5%	10%	1%	3%	5%	10%
Sample 1	31.46	33.25	32.91	32.46	35.20	35.49	34.50	34.96	34.27	33.61	34.27	34.24	
Sample 2	1.31	-0.25	0.23	0.44	-0.51	0.26	0.16	0.58	0.71	-0.13	0.99	0.51	
Sample 3	-39.25	-35.34	-37.81	-37.09	-32.48	-37.34	-38.12	-37.52	-38.51	-37.46	-37.46	-37.43	
Sample 4	3.62	0.51	1.63	-0.49	4.06	3.28	-0.61	1.03	3.70	-0.03	0.55	2.45	
Design-SRS	3.79	90.01	90.12	-0.15	3.79	3.25	-0.56	1.04	90.34	0.29	0.61	2.40	

Table 4: Percent Difference of Estimated Total from Actual Total in High-Mean Poisson Population: Po(100)

	P-Value ( $\hat{\beta}$ )	0.0009	0.0009	0.0009	0.0009	0.0322	0.0322	0.0322	0.0322	0.6200	0.6200	0.6200	0.6200
	Sampling Rate	1%	3%	5%	10%	1%	3%	5%	10%	1%	3%	5%	10%
N=1,000	Sample 1	13.91	14.84	13.44	13.71	13.78	13.92	14.44	13.92	11.41	13.63	13.08	14.01
	Sample 2	0.31	-0.22	-1.32	-0.55	1.12	-0.47	0.35	-0.46	1.89	0.74	-0.68	-0.72
	Sample 3	-13.27	-14.12	-14.30	-14.34	-13.73	-15.28	-15.04	-13.41	-11.76	-14.69	-13.81	-14.25
	Sample 4	0.76	0.75	2.89	-0.83	-6.58	3.90	-0.68	-0.84	-4.26	-0.38	-1.36	-1.53
	Design-SRS	0.55	0.81	2.86	-0.73	-5.89	3.90	-0.76	-0.76	-3.58	-0.73	-1.32	-1.53
N=10,000	P-Value ( $\hat{\beta}$ )	<0.0001	<0.0001	<0.0001	<0.0001	0.0159	0.0159	0.0159	0.0159	0.7251	0.7251	0.7251	0.7251
	Sampling Rate	1%	3%	5%	10%	1%	3%	5%	10%	1%	3%	5%	10%
	Sample 1	14.31	14.19	14.59	14.24	14.47	14.08	14.09	14.19	14.89	14.22	14.43	14.23
	Sample 2	0.23	1.02	0.68	0.35	0.19	0.16	0.60	0.56	0.25	0.03	0.18	0.35
	Sample 3	-13.22	-13.50	-13.61	-13.40	-13.78	-13.81	-13.26	-13.44	-13.62	-13.58	-13.52	-13.52
	Sample 4	1.62	0.27	-0.49	0.10	-0.23	-0.34	-3.22	0.14	-1.08	-0.76	0.31	0.01
	Design-SRS	1.17	0.23	-0.49	0.09	-0.22	-0.33	-0.23	0.14	-0.97	-0.72	0.31	0.00

Table 5: Coefficient of Variation of Estimated Total in a Homogeneous Normal Population: N(100,25)

	R <sup>2</sup>	96.48	96.48	96.48	96.48	52.05	52.05	52.05	52.05	21.08	21.08	21.08	21.08
	Sampling Rate	1%	3%	5%	10%	1%	3%	5%	10%	1%	3%	5%	10%
N=1,000	Sample 1	0.01	0.02	0.03	0.05	0.01	0.01	0.01	0.02	0.00	0.01	0.01	0.02
	Sample 2	0.01	0.02	0.03	0.05	0.01	0.01	0.02	0.01	0.00	0.01	0.01	0.01
	Sample 3	0.01	0.03	0.03	0.05	0.00	0.01	0.01	0.01	0.00	0.00	0.01	0.01
	Sample 4	0.02	0.03	0.04	0.05	0.01	0.02	0.03	0.04	0.01	0.02	0.02	0.02
	Design-SRS	1.59	1.07	0.68	0.44	2.23	0.97	0.67	0.43	2.17	0.93	0.74	0.45
	BS Replication	500	500	500	500	500	500	500	500	500	500	500	500
	BS Sample Size	10	30	50	100	10	30	50	100	10	30	50	100
N=10,000	R <sup>2</sup>	96.17	96.17	96.17	96.17	49.39	49.39	49.39	49.39	19.07	19.07	19.07	19.07
	Sampling Rate	1%	3%	5%	10%	1%	3%	5%	10%	1%	3%	5%	10%
	Sample 1	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
	Sample 2	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Sample 3	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Sample 4	0.00	0.01	0.01	0.02	0.00	0.01	0.01	0.01	0.00	0.00	0.01	0.01
	Design-SRS	0.45	0.29	0.23	0.15	0.46	0.27	0.23	0.15	0.52	0.31	0.21	0.16
	BS Replication	500	500	500	500	500	500	500	500	500	500	500	500
BS Sample Size	100	300	500	1000	100	300	500	1000	100	300	500	1000	

Table 6: Coefficient of Variation of Estimated Total in a Heterogeneous Normal Population:  
N(100,2500)

N=1,000	<b>R<sup>2</sup></b>	<b>99.96</b>	<b>99.96</b>	<b>99.96</b>	<b>99.96</b>	<b>52.03</b>	<b>52.03</b>	<b>52.03</b>	<b>52.03</b>	<b>21.08</b>	<b>21.08</b>	<b>21.08</b>	<b>21.08</b>
	<b>Sampling Rate</b>	<b>1%</b>	<b>3%</b>	<b>5%</b>	<b>10%</b>	<b>1%</b>	<b>3%</b>	<b>5%</b>	<b>10%</b>	<b>1%</b>	<b>3%</b>	<b>5%</b>	<b>10%</b>
	Sample 1	0.17	0.30	0.38	0.53	0.17	0.17	0.27	0.41	0.20	0.18	0.29	0.34
	Sample 2	0.16	0.31	0.38	0.51	0.03	0.07	0.05	0.12	0.03	0.06	0.07	0.07
	Sample 3	0.17	0.29	0.38	0.53	0.04	0.07	0.09	0.12	0.01	0.07	0.04	0.07
	Sample 4	0.17	0.29	0.37	0.52	0.20	0.23	0.29	0.32	0.02	0.16	0.17	0.23
	Design-SRS	40.24	10.48	8.52	4.46	5.70	9.72	7.33	4.43	18.14	8.84	7.57	4.80
	BS Replication	500	500	500	500	500	500	500	500	500	500	500	500
	BS Sample Size	10	30	50	100	10	30	50	100	10	30	50	100
N=10,000	<b>R<sup>2</sup></b>	<b>99.96</b>	<b>99.96</b>	<b>99.96</b>	<b>99.96</b>	<b>49.39</b>	<b>49.39</b>	<b>49.39</b>	<b>49.39</b>	<b>19.07</b>	<b>19.07</b>	<b>19.07</b>	<b>19.07</b>
	<b>Sampling Rate</b>	<b>1%</b>	<b>3%</b>	<b>5%</b>	<b>10%</b>	<b>1%</b>	<b>3%</b>	<b>5%</b>	<b>10%</b>	<b>1%</b>	<b>3%</b>	<b>5%</b>	<b>10%</b>
	Sample 1	0.05	0.09	0.11	0.17	0.03	0.06	0.08	0.12	0.01	0.05	0.05	0.08
	Sample 2	0.05	0.09	0.11	0.16	0.01	0.02	0.02	0.03	0.00	0.01	0.01	0.02
	Sample 3	0.05	0.09	0.11	0.16	0.01	0.01	0.02	0.03	0.00	0.01	0.01	0.02
	Sample 4	0.05	0.09	0.11	0.16	0.04	0.06	0.09	0.12	0.02	0.04	0.05	0.09
	Design-SRS	5.29	2.96	2.26	1.48	5.22	2.84	2.28	1.48	4.58	2.84	2.08	1.63
	BS Replication	500	500	500	500	500	500	500	500	500	500	500	500
	BS Sample Size	100	300	500	1000	100	300	500	1000	100	300	500	1000

Table 7: Coefficient of Variation of Estimated Total in a Low-Mean Poisson Population:  
Po(10)

N=1,000	<b>P-Value (<math>\hat{\beta}</math>)</b>	<b>&lt;.0001</b>	<b>&lt;.0001</b>	<b>&lt;.0001</b>	<b>&lt;.0001</b>	<b>0.0441</b>	<b>0.0441</b>	<b>0.0441</b>	<b>0.0441</b>	<b>0.3454</b>	<b>0.3454</b>	<b>0.3454</b>	<b>0.3454</b>
	<b>Sampling Rate</b>	<b>1%</b>	<b>3%</b>	<b>5%</b>	<b>10%</b>	<b>1%</b>	<b>3%</b>	<b>5%</b>	<b>10%</b>	<b>1%</b>	<b>3%</b>	<b>5%</b>	<b>10%</b>
	Sample 1	0.02	0.04	0.05	0.09	0.02	0.03	0.04	0.07	0.02	0.02	0.03	0.08
	Sample 2	0.01	0.01	0.01	0.04	0.01	0.01	0.02	0.03	0.01	0.01	0.01	0.03
	Sample 3	0.01	0.01	0.03	0.04	0.01	0.02	0.02	0.04	0.01	0.02	0.02	0.04
	Sample 4	0.03	0.05	0.28	0.10	0.07	0.05	0.07	0.12	0.02	0.03	0.06	0.10
	Design-SRS	7.37	6.14	4.61	2.60	9.84	6.06	4.28	3.37	11.39	5.66	4.99	0.98
	BS Replication	500	500	500	500	500	500	500	500	500	500	500	500
	BS Sample Size	10	30	50	100	10	30	50	100	10	30	50	100
N=10,000	<b>P-Value (<math>\hat{\beta}</math>)</b>	<b>&lt;0.0001</b>	<b>&lt;0.0001</b>	<b>&lt;0.0001</b>	<b>&lt;0.0001</b>	<b>0.0202</b>	<b>0.0202</b>	<b>0.0202</b>	<b>0.0202</b>	<b>0.5415</b>	<b>0.5415</b>	<b>0.5415</b>	<b>0.5415</b>
	<b>Sampling Rate</b>	<b>1%</b>	<b>3%</b>	<b>5%</b>	<b>10%</b>	<b>1%</b>	<b>3%</b>	<b>5%</b>	<b>10%</b>	<b>1%</b>	<b>3%</b>	<b>5%</b>	<b>10%</b>
	Sample 1	0.01	0.01	0.02	0.03	0.00	0.01	0.01	0.02	0.00	0.01	0.01	0.02
	Sample 2	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01
	Sample 3	0.00	0.00	0.01	0.02	0.00	0.00	0.01	0.01	0.00	0.00	0.01	0.01
	Sample 4	0.01	0.02	0.03	0.05	0.00	0.01	0.02	0.03	0.00	0.01	0.01	0.03
	Design-SRS	3.56	18.47	13.88	0.98	3.14	1.83	1.39	0.95	33.77	1.91	0.06	0.99
	BS Replication	500	500	500	500	500	500	500	500	500	500	500	500
	BS Sample Size	100	300	500	1000	100	300	500	1000	100	300	500	1000

Table 8: Coefficient of Variation of Estimated Total in a High-Mean Poisson Population:  
 $Po(100)$

N=1,000	<b>P-Value (<math>\hat{\beta}</math>)</b>	<b>0.0009</b>	<b>0.0009</b>	<b>0.0009</b>	<b>0.0009</b>	<b>0.0322</b>	<b>0.0322</b>	<b>0.0322</b>	<b>0.0322</b>	<b>0.6200</b>	<b>0.6200</b>	<b>0.6200</b>	<b>0.6200</b>
	<b>Sampling Rate</b>	<b>1%</b>	<b>3%</b>	<b>5%</b>	<b>10%</b>	<b>1%</b>	<b>3%</b>	<b>5%</b>	<b>10%</b>	<b>1%</b>	<b>3%</b>	<b>5%</b>	<b>10%</b>
	Sample 1	0.01	0.01	0.02	0.02	0.01	0.01	0.01	0.02	0.00	0.01	0.01	0.02
	Sample 2	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.01
	Sample 3	0.01	0.00	0.01	0.02	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.02
	Sample 4	0.01	0.01	0.02	0.04	0.01	0.01	0.02	0.04	0.01	0.01	0.02	0.03
	Design-SRS	2.34	20.91	1.53	1.00	2.61	2.60	0.19	1.11	7.52	1.89	1.42	0.97
	BS Replication	500	500	500	500	500	500	500	500	500	500	500	500
	BS Sample Size	10	30	50	100	10	30	50	100	10	30	50	100
N=10,000	<b>P-Value (<math>\hat{\beta}</math>)</b>	<b>&lt;0.0001</b>	<b>&lt;0.0001</b>	<b>&lt;0.0001</b>	<b>&lt;0.0001</b>	<b>0.0159</b>	<b>0.0159</b>	<b>0.0159</b>	<b>0.0159</b>	<b>0.7251</b>	<b>0.7251</b>	<b>0.7251</b>	<b>0.7251</b>
	<b>Sampling Rate</b>	<b>1%</b>	<b>3%</b>	<b>5%</b>	<b>10%</b>	<b>1%</b>	<b>3%</b>	<b>5%</b>	<b>10%</b>	<b>1%</b>	<b>3%</b>	<b>5%</b>	<b>10%</b>
	Sample 1	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01
	Sample 2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Sample 3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Sample 4	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.01	0.00	0.00	0.01	0.01
	Design-SRS	1.02	0.56	0.43	0.29	1.00	0.55	0.43	0.30	0.97	0.57	0.44	0.30
	BS Replication	500	500	500	500	500	500	500	500	500	500	500	500
	BS Sample Size	100	300	500	1000	100	300	500	1000	100	300	500	1000