9th National Convention on Statistics (NCS)
EDSA Shangri-La Hotel
October 4-5, 2004

Monte Carlo Estimation in RCPS with Weighting Adjustment

by

Joseph Ryan G. Lansangan

For additional information, please contact:

Author's name:	Joseph Ryan G. Lansangan
Designation:	Instructor
Agency:	School of Statistics University of the Philippines - Diliman
Address:	Diliman, Quezon City
Telephone:	928-0881

Monte Carlo Estimation in RCPS with Weighting Adjustment

by
Joseph Ryan G. Lansangan
Instructor, UP School of Statistics

ABSTRACT

One of the surveys conducted regularly by Bureau of Agricultural Statistics (BAS), either semi-annually on minor palay and corn producing provinces and quarterly on major producing provinces and perhaps the most extensive nationwide survey, is the Rice and Corn Production Survey (RCPS). The RCPS, having a two-stage stratified sampling design (with replication), is aimed to characterize and estimate the national output on cereal crops. The survey targets all producing provinces that contribute a significant proportion to national production. The design constructed many years back uses partially updated frame based on the 1991 Census of Agriculture and Fisheries. This lead to estimates that produces unacceptable levels of coefficient of variation for particular domains. Weight adjustment would ideally improve the estimates once there are samples that cannot be located or are no longer part of the population. A simulation on the population values was initially conducted and different levels of nonresponse were considered under the sampling procedure. Monte Carlo estimates of bias and standard error showed that weight adjustment can be better in having a more statistically sound estimates and predictions, having relatively minimum bias and very small coefficient of variation.

Key words: RCPC, weighting adjustment, monte carlo estimation

I. Introduction

The Bureau of Agricultural Statistics (BAS) serves as the official data bank for agriculture. It covers about 78 provinces and 15 regions of the country, and conducts different surveys on agricultural commodities. Among the objectives of these surveys conducted by the agency are the generation of statistics such as production on agricultural commodities, and estimation of measurements needed by the data users such as in livestock and poultry.

One of the surveys that BAS regularly conducts is the Rice and Corn Production Survey (RCPS), which is perhaps the most extensive nationwide survey. This survey is conducted either semi-annually on minor palay and corn producing provinces or quarterly on major palay and corn producing provinces. Among its general objectives are to (1) generate quarterly data on palay and corn production and yield; (2) generate short term forecast on area and production; and (3) collect quarterly data on production utilization and fertilizer usage. The survey design, developed and piloted in 1993, uses the 1991 Census of Agriculture and Fisheries (CAF) as sampling frame.

As with most of the conducted nationwide surveys, the sampling design and the sampling frame are vital in the results of the RCPS. The RCPS, having a two-stage, with replication, stratified sampling design, is directed to characterizing and predicting the national output on cereal crops—targeting all possible provinces that are deemed to have significant turnouts in the survey. The

sampling frame, as the basis for the accumulation of the necessary inputs, defines the structure of the sampling population—of how palay and corn producers are possibly segregated across the different provinces and barangays.

As for the concern of the data users, including the producers themselves and several private or public agricultural institutions, the BAS's system on the generation of estimates and predictions is "devised" to produce accurate and reliable results. This may even involve adjustment on the levels of the data especially if these are not found to conform to the existing production patterns and trends. Though the agency utilizes statistical techniques to produce the necessary quantities, such as imputation through the IMPS and the direct design-based estimation, the "reliable" estimates are those that are data-validated by the Provincial Agricultural Officers (PASOs) and the Regional Agricultural Officers (RASOs) of which a national review is conducted to arrive at the official National Data Review validated (NDR-validated) estimates. And because a nonprobabilistic or subjective technique of final estimation is used, there is no means of measuring the accuracy, precision, and reliability of the estimates produced by the agency. With the past results of the RCPS, the estimates generated through the design-based estimation, the IMPS based estimation, and the PASOs/RASOs/NDR data validation, suggest large discrepancies. And a study conducted by the Statistical Methods and Research Division of BAS (SMRD-BAS) late October last year suggests a possible problem in its outdated sampling frame, and a possible technique of estimation that would be better compared to the current techniques done by the agency which is WEIGHT ADJUSTMENT.

The weight adjustment technique is not a new tool in the analysis of survey data. Weight adjustment is inherently sampling weights tagged to the sampled respondents in the survey. That is, it assigns sampled respondents greater importance relative to the size of their strata or group classifications, or possibly relative to the type of information gathered from these respondents. In general, the appropriate weight adjustment depends on the level of the analysis. In the RCPS data, weight adjustment is applied in order to compensate for the non-responding households. The weight adjustment for the RCPS data is based on the sample data, that is, the weights are obtained on the basis of the number of households to be sampled based on the survey design, and of the number of households actually sampled during the survey. These are obtained across the strata for all the provinces. That is, for each stratum in each province, the responding households are weighted up by a factor that makes the total number of actually sampled households for a given stratum equal to the total number of households that must be sampled based on the survey design. In general, the non-adjustment for the non-response in the RCPS data results in under-estimation. Thus, weight adjustment can be a better procedure than existing techniques in having more statistically sound estimates and predictions.

II. Objectives and Limitations

This study is undertaken primarily to determine the characteristics of the weight adjusted estimates for the July 2003 RCPS on the 78 coverage provinces

and compare it with the original design-based estimates. Specifically, the author would want to determine the accuracy and the precision, in terms of the Bias and the Standard Error (or coefficient of variation or CV), of the weight-adjusted estimator for production and compare these properties with the original design-based estimator for production. To do these, the author considers different scenarios, with regards to the percentage of non-response and observes how the weight-adjusted estimator behaves when there is a high (low) non-response percentage using Monte Carlo methods. The use of Monte Carlo methods to address the objective allows the author to examine more complex scenarios than what is currently observed.

The data structure came from SMRD-BAS. The agency provided the sampling design for the provinces (71 for the Palay Production Survey and only 55 for the Corn Production Survey), including the different replicates for the different strata and the number of households to be sampled based on the design. To meet the objectives, simulations of the population quantities of the provinces were made. That is, for each province, each barangay included in the survey design assumed its own quantity of production. This way, the parameters of interest, the "actual" production of rice and corn for each province, were known. Then, households were sampled from each province, as indicated in the survey design, and at different non-response rates. The sampling procedure at a given percentage of non-response was repeated 100 times (initial runs suggested that the behavior of the estimators were already defined even for fewer than 100 repetitions) and based on these 100 samples, the Monte-Carlo weight-adjusted and design-based estimates will be computed. The simulations were used to describe the survey design's system as a whole, and hence based on the computed estimates, the estimators were characterized.

III. Review of the RCPS Sampling Design

The design of the RCPS is based on the 40 major palay-producing provinces and the 27 major corn-producing provinces. Survey operation is done quarterly in the major producing provinces and semi-annually for the remaining provinces. The provinces are the domains of the survey. A four-replicate two-stage stratified sampling design is used. The primary sampling units (psu) are the barangays that are to be selected using probability proportional to size sampling (pps). The psus are first divided into strata of approximately equal sizes relative to the total farm area devoted to palay for the Palay Production Survey (PPS) and the total farm area devoted to corn for the Corn Production Survey (CPS).

The secondary sampling units (ssu) are the palay and/or corn farming households in the sample barangays. The sample households are drawn systematically from each sample barangay. The result is a self-weighted sampling scheme that facilitates estimation of survey characteristics. For economic reasons, sample sizes at the ssu level are set to a minimum of 4 and a maximum of 25 households. To cope with this constraint, household weights are instituted. The uniform household weights for the i^{th} barangay in the h^{th} stratum of the k^{th} province (W_{khi}) is computed as follows:

$$W_{khi} = 1.00 \quad if \quad 4 \le n_{khi} \le 25$$

$$W_{khi} = \frac{n_{khi}}{4} \quad if \quad n_{khi} < 4$$

$$W_{khi} = \frac{n_{khi}}{25} \quad if \quad n_{khi} > 25$$

$$W_{khi} = \frac{n_{khi}}{N_{khi}} \quad if \quad n_{khi} < 25 \; and \; n_{khi} > N_{khi}$$

where n_{khi} is the number of sample households in the i^{th} barangay of the h^{th} stratum for the k^{th} province, and N_{khi} is the total number of farm households in the i^{th} sample barangay of the h^{th} stratum for the k^{th} province. Here, there are 78 provinces, at most 10 strata for each province, and at most 4 barangays per stratum, each with 4 to 25 sample households.

IV. The Data

The data for this study is a simulation from the July 2003 round of the palay and corn production survey. The data points are simulated based on the design or the current sampling frame for the survey, and are taken as ten thousands in units. Different sets of simulated data points are generated, more or less capturing the actual schemes for the survey. Hence, making the results more reliable. Tables 1 and 2 show the number of strata and replicates for the provinces that are included in the PPS and the CPS, respectively, and the total number of households that are to be sampled and the corresponding raising factor for each province.

The numbers of strata for both the PPS and the CPS take values at either 10 or 40. The number of replicates for the major producing provinces is fixed at 4, while the minor producing provinces have replicates equal to 2. The number of sample households varies in each barangay/stratum across provinces in a region. This ranges from four to twenty five based on the design. The actual numbers of households sampled are determined only after the sampling procedures are applied (and across the 100 repetitions, it is very much likely to get different sets households for the samples). The raising factors are fixed across provinces.

After the simulation, the production parameters are identified. Then samples are generated based on the design. The sampling procedures cover different scenarios of non-response, i.e., the analyses include estimating the production parameters when there is a 10% non-response, a 20% non-response, and so on, up to a 90% non-response.

V. Estimation

The computations for the rice and corn production are as follows:

$$Weight\text{-}Adjusted\ production = \frac{production\ *\ adjusted\ household\ weights\ *Rk*4}{Number\ of\ replicates}$$

$$Design\text{-}Based\ production = \frac{production\ *\ original\ household\ weights\ *Rk*4}{Number\ of\ replicates}$$

where Rk is the raising factor for the k^{th} province.

Households in the same barangay receive the same household weight. The original household weights are computed from the design while the adjusted household weights are computed based on the sampling done (the original household weight multiplied by the number of original households in the design divided by the actual number of households generated in the sampling procedure). Sampling is done per barangay to assure nonzero responses for all barangays (otherwise weights become more complicated).

VI. Results and Discussions

The parameter values for the simulation of the production of the barangays across all provinces are made to vary across the provinces, ranging from 10 to 222404, and with standard deviations ranging from 5 to 50000. The simulations are done independently across provinces. More than one set of population parameters are generated for the study, but fewer number of repetitions are done due to computer capacity and time constraint. Still, the Monte Carlo methods produce similar results as to using only one set of population parameters. For the subsequent discussions, 100 repetitions of sampling are made across this set and at different levels of non-response.

Tables 3, 4, 5 and 6 give the results of the estimation for a particular major producing province. Tables 7, 8, 9 and 10 give the results for a particular minor producing province. The total production parameter for the major producing province is at 497480.08, while that of the minor producing province is at 1037.7. Based on Tables 3 and 7, the weight-adjusted estimates on the average are far better than the design-based estimates. Using the design-based estimator, consistent across all levels of non-response, on the average, underestimates the true production value. Tables 4 and 8 also suggest that the weight-adjusted estimator is far less biased than the design-based estimator is. Noticeably, for the major producing province, across the different non-response rates, the weight-adjusted estimates are different from the parameter on the average by at most 0.47%; while the design-based estimates are biased on the average by at least 4.68% and at most 86.35%. And this is also the case for the minor producing province: the weight-adjusted estimates are biased on the average by at most 1.48%; while the design-based estimates, by 3.83% to 84.60%. From tables 5,6, 9 and 10, the 100 generated samples indicate that the behavior of both estimators are very much stable, and the coefficients of variation for the 100 samples at every level of non-response show that the weight-adjusted estimator is more stable than the design-based estimator. All these results point to the

weight-adjusted estimator as a better estimator for production compared to the original design-based estimator.

VII. Conclusion

Based on the above results, the weight adjustment for the estimation of the production for the different provincess gives better results than the original design-based estimation. Both the design-based and weight-adjusted estimates remain stable even at different levels of non-response; however, the weight-adjusted estimates are closer to the true parameter value being estimated. The bias for using the weight adjustment is smaller compared to the bias for using the design-based estimation. Furthermore, the coefficient of variation is smaller for the weight-adjusted estimator compared to that of the design-based estimator. The design-based estimation is shown to be effective only for the case of very low non-response or a near complete coverage of the households. The design-based estimation in general underestimates the true parameter value.

Assuming that the sampling design for the survey is still reliable, that is, the structure as to how the classifications are made for the barangays at the time the design is made still holds at present, then the agency can be confident of their estimates. Once the problem of non-response is treated, then the estimates and predictions for the RCPS can become reliable and statistically correct. These can be treated by the agricultural officers as good estimates as the survey data and may then impose their subjective inputs to “strengthen” these quantities.

Table1. Information on Provinces included in the PPS

Province	No. of Strata	No. of Rep.	No. of Hholds	Raising Factor	Province	No. of Strata	No. of Rep.	No. of Hholds	Raising Factor
Abra	10	2	100	50	***Laguna***	40	4	262	100
Agusan Norte	10	2	73	75	***Lanao Norte***	40	4	370	100
Agusan Sur	40	4	406	100	***Lanao Sur***	40	4	428	100
Aklan	40	4	402	100	***Leyte***	40	4	331	350
Albay	40	4	361	150	***Maguindanao***	40	4	369	250
Antique	40	4	410	150	***Marinduque***	10	2	166	50
Aurora	10	2	103	25	***Masbate***	40	4	369	150
Basilan	10	2	67	25	***Mindoro Occ***	40	4	597	50
Bataan	10	2	137	25	***Mindoro Or***	40	4	385	100
Batangas	40	4	264	250	***Misamis Occ***	10	2	107	100
Benguet	10	2	100	50	***Misamis Or***	10	2	66	150
Biliran	10	2	90	25	***Mt. Province***	10	2	153	25
Bohol	40	4	364	300	***Negros Occ***	40	4	317	250
Bukidnon	40	4	290	250	***Negros Or***	10	2	70	250
Bulacan	40	4	395	100	***North Cotabato***	40	4	374	250
Cagayan	40	4	442	200	***N. Samar***	40	4	378	100
Camarines Nor	10	2	109	75	***Nueva Ecija***	40	4	435	300
Camarines Sur	40	4	442	300	***Nueva Viscaya***	40	4	481	50
Camiguin	10	2	130	25	***Palawan***	40	4	504	100
Capiz	40	4	443	150	***Pampanga***	40	4	386	100
Catanduanes	10	2	115	75	***Pangasinan***	40	4	424	400
Cavite	10	2	109	75	***Quezon***	40	4	310	150
Cebu	10	2	62	400	***Quirino***	10	2	110	25
Davao City	10	24	121	75	***Rizal***	10	2	123	25
Davao Nor	40	4	235	250	***Romblon***	10	2	91	25
Davao Oriental	10	2	71	100	***Eastern Samar***	10	2	113	75
Davao Sur	40	4	272	250	***Sarangani***	10	2	108	75
Guimaras	10	2	116	25	***Siquijor***	10	2	72	25
Ifugao	10	2	118	25	***Sorsogon***	40	4	457	100
Ilocos Norte	40	4	441	100	***South Cotabato***	40	4	319	200
Ilocos Sur	40	4	497	100	***Southern Leyte***	10	2	68	100
Iloilo	40	4	403	300	***Sultan Kudarat***	40	4	402	100
Isabela	40	4	368	300	***Sulu***	10	2	67	75
Kal. Apayao	40	4	467	50	***Surigao Nor***	40	4	382	100
La Union	40	10	437	100	***Surigao Sur***	40	4	320	100

Province	No. of Strata	No. of Rep.	No. of Hholds	Raising Factor	Province	No. of Strata	No. of Rep.	No. of Hholds	Raising Factor
Tarlac	40	4	450	150	*Zambo. Norte*	40	4	358	200
Western Samar	40	4	370	100	*Zambo. Sur*	40	4	351	350
Zambales	10	2	122	50					
Zambo. City	10	2	98	25					

Table2. Information on Provinces included in the CPS

Province	No. of Strata	No. of Rep.	No. of Hholds	Raising Factor	Province	No. of Strata	No. of Rep.	No. of Hholds	Raising Factor
Agusan Sur	40	4	538	100	*Mindoro Occ*	10	2	65	100
Albay	40	4	305	150	*Mindoro Or*	10	2	40	100
Batangas	10	2	57	250	*Misamis Occ*	40	4	291	100
Biliran	10	2	73	25	*Misamis Or*	40	4	371	150
Bohol	40	4	365	300	*Mt. Province*	10	2	83	25
Bukidnon	40	4	439	250	*Negros Occ*	40	4	335	250
Bulacan	10	2	41	100	*Negros Or*	40	4	350	200
Cagayan	40	4	314	200	*North Cotabato*	41	4	362	250
Camarines Sur	40	4	277	300	*N. Samar*	10	2	45	100
Capiz	10	2	45	150	*Nueva Ecija*	10	2	48	300
Cebu	40	4	436	400	*Nueva Viscaya*	10	2	65	50
Davao City	40	4	407	250	*Palawan*	40	4	357	100
Davao Oriental	40	4	340	100	*Pampanga*	10	2	51	100
Davao Prov	40	4	388	250	*Pangasinan*	40	4	218	400
Davao Sur	40	4	407	250	*Quezon*	40	4	218	150
Guimaras	10	2	102	25	*Samar*	10	2	79	100
Ifugao	10	2	54	25	*Sarangani*	40	4	353	75
Ilocos Norte	10	2	59	100	*Sorsogon*	10	2	51	100
Ilocos Sur	10	2	45	100	*South Cotabato*	40	4	328	200
Iloilo	40	4	304	300	*Sultan Kudarat*	40	4	377	100
Isabela	40	4	315	300	*Tarlac*	10	2	42	150
Kal. Apayao	10	2	47	50	*Western Samar*	10	2	79	100
Lanao Norte	40	4	386	100	*Zambo. City*	10	2	95	350
Lanao Sur	40	4	363	100	*Zambo. Norte*	40	4	414	200
Leyte	40	4	270	350	*Zambo. Sur*	40	4	390	350
Maguindanao	40	4	395	250					
Masbate	40	4	442	150					

Table 3. Estimated Production for a Major Producing Province

	Estimated Production At Different Percentages of Nonresponse								
Non-response	*90%*	*80%*	*70%*	*60%*	*50%*	*40%*	*30%*	*20%*	*10%*
Weight-Adjusted	495131.3	497894.7	495325.6	497084.6	497685.5	498155.9	497267.7	498038.7	497568.7
Design-Based	67899.33	112744.4	168710.2	215544.6	264258.6	313092.3	372785.4	416210.6	474215

Table 4. Bias of Estimator for Production for a Major Producing Province

	Bias At Different Percentages of Nonresponse								
Non-response	*90%*	*80%*	*70%*	*60%*	*50%*	*40%*	*30%*	*20%*	*10%*
Weight-Adjusted	-2348.78	414.62	-2154.46	-395.44	205.38	675.8	-212.35	558.63	88.66
Design-Based	-429581	-384736	-328770	-281936	-233222	-184388	-124695	-81269.4	-23265.1

Table 5. Coefficient of Variation of Estimator for Production for a Major Producing Province

		Coefficient of Variation (CV) in Percent for the Resamples At Different Percentages of Nonresponse								
Non-response		*90%*	*80%*	*70%*	*60%*	*50%*	*40%*	*30%*	*20%*	*10%*
Weight-	*Min*	1.34	2.4	3.72	4.93	5.99	7.06	8.67	9.84	11.42
Adjusted	*Max*	1.86	3	4.38	5.55	6.81	7.95	9.26	10.32	11.62
Design-	*Min*	9.88	10.62	11	11.37	11.3	11.24	11.57	11.79	11.99
Based	*Max*	13.54	13.29	12.89	12.84	12.85	12.67	12.36	12.36	12.19

Table 6. Coefficient of Variation Relative to Parameter of Estimator for Production for a Major Producing Province

	Percent Variability Relative to the Production Parameter At Different Percentages of Nonresponse								
Non-response	*90%*	*80%*	*70%*	*60%*	*50%*	*40%*	*30%*	*20%*	*10%*
Weight-Adjusted	5.48%	4.06%	2.62%	2.35%	1.90%	1.59%	1.16%	0.85%	0.42%
Design-Based	5.29%	3.92%	2.57%	2.33%	1.87%	1.57%	1.16%	0.84%	0.41%

Table 7. Estimated Production for a Minor Producing Province

	Estimated Production At Different Percentages of Nonresponse								
Non-response	*90%*	*80%*	*70%*	*60%*	*50%*	*40%*	*30%*	*20%*	*10%*
Weight-Adjusted	1037.23	1053.07	1041.38	1041	1039.76	1031.99	1035.54	1039.14	1041.61
Design-Based	159.78	236.24	381.01	467.42	542.01	668.83	772.32	906.5	997.91

Table 8. Bias of Estimator for Production for a Minor Producing Province

	Bias At Different Percentages of Nonresponse								
Non-response	*90%*	*80%*	*70%*	*60%*	*50%*	*40%*	*30%*	*20%*	*10%*
Weight-Adjusted	-0.47	15.37	3.68	3.3	2.06	-5.71	-2.16	1.44	3.91
Design-Based	-877.92	-801.46	-656.69	-570.28	-495.69	-368.87	-265.38	-131.2	-39.79

Table 9. Coefficient of Variation of Estimator for Production for a Minor Producing Province

		Coefficient of Variation (CV) in Percent for the Resamples At Different Percentages of Nonresponse								
Non-response		*90%*	*80%*	*70%*	*60%*	*50%*	*40%*	*30%*	*20%*	*10%*
Weight-	*Min*	1.55	2.28	4.39	5.14	6.32	8.34	9.58	11.76	13.33
Adjusted	*Max*	4.14	4.51	6.8	8.03	8.69	10.86	12	13.42	14.2
Design-	*Min*	9.88	10.62	11	11.37	11.3	11.24	11.57	11.79	11.99
Based	*Max*	10.96	10.27	12.34	11.56	12.13	12.94	12.84	13.54	13.94

Table 10. Coefficient of Variation Relative to Parameter of Estimator for Production for a Minor Producing Province

	Percent Variability Relative to the Production Parameter At Different Percentages of Nonresponse								
Non-response	*90%*	*80%*	*70%*	*60%*	*50%*	*40%*	*30%*	*20%*	*10%*
Weight-Adjusted	16.46%	13.06%	7.78%	6.97%	6.11%	4.85%	4.03%	2.62%	1.55%
Design-Based	13.83%	12.60%	7.16%	6.59%	6.16%	4.61%	4.01%	2.47%	1.49%

References:

1. “The Effect of Weighting Adjustment on the July 2003 Round of Palay and Corn Production in Selected Regions” by SMRD-BAS, October 30, 2003.

2. Enumerators’ Manuals, Palay and Corn Production Survey

3. Rice and Corn Production Survey, July 2003 Round