Bootstrap Estimation of the Average Household Expenditure on Personal Care and Effects at Regional Level

Jachelle Anne G. Dimapilis
University of the Philippines Diliman

This study explored the use of bootstrap in estimating households' expenditure on personal care and effects at regional level. Bootstrap yields superior estimates compared to the estimates of simple random sampling without replacement (SRSWOR). Bootstrap estimates have lower variance than SRSWOR estimates. Furthermore, bootstrap estimates have smaller bias compared with SRSWOR estimates.

Keywords: survey sampling, predictive estimation, bootstrap estimation

1. Introduction

According to the result of Family Income and Expenditure Survey (FIES) conducted by the National Statistics Office (NSO) in 2006, personal care and effects (PRCRE) came sixth in the list of the major expenditure items of Filipino households with a share of 3.7% of the total expenditure. PRCRE products consist of personal hygiene and personal care like beauty aids, baby and adult care products, and others. The use of PRCRE products becomes inevitable in Filipino households nowadays. High market demand is one of the factors why PRCRE is a major industry in our country.

Determining how much Filipinos spend on PRCRE products is often carried out through surveys. However, conducting expenditure surveys across economic classes encounters some problems. Barrios (2008) cited that in skewed distribution like income, upper and lower income classes are difficult to convince to take part in surveys. Households that belong to lower income bracket might contribute minimal information because they spend little on these items. On the other hand, households that belong to higher income bracket might have the tendency to hide the true information about their income and spending.

One approach to address this problem is to use nonparametric techniques. Barrios and Kwong (2010) used nonparametric model-based estimation procedure to estimate the market size for food items and some of its components. They found out that even though the sample is drawn from the small part of the population, model-based estimates are superior or at least comparable to design-based estimates especially for small populations. They concluded that in skewed population, the model-based estimator is robust to the relationship between the target variable and auxiliary information. They also found out that bootstrap sampling errors are generally lower than the design-unbiased sampling errors.

This paper explores the use bootstrap in estimating households' expenditure on PRCRE at regional level and compares the result with that of simple random sampling without replacement (SRSWOR).

2. Methodology

This study used the microdata of 2006 FIES, particularly the total income and total expenditure for PRCRE of households across region. The collection of participating households in FIES is assumed to be the population in this study. Per region, the middle 50% of the population was identified based on PRCRE expenditure. The sample was drawn from this part of the population.

The sample per region was selected using simple random sampling without replacement.The sample size varies per region and is determined using the formula $n_h = \frac{S_h^2 Z_{\alpha/2}^2}{r^2 {\mu_h}^2}$, where S_h is the standard deviation of PRCRE expenditure in h^{th} region, μ_h is the mean PRCRE expenditure in h^{th} region, $Z_{\alpha/2}$ = 1.96, and r = 0.1.

Resampling was done by performing modified bootstrap in order to create a recreated population. Mooney and Duval (1993) define bootstrapping as a computationally intensive, nonparametric technique which employs resampling the data a large number times in order to make inferences about population parameters. This explains why recreated population is formed. Per region, each sample was replicated K number of times. K is computed using the formula $K = N/n$, where N is the number of observation and n is the sample size. Then samples were drawn from the recreated population K times. Hence, there were K samples per region. Table 1 shows the sample size and the number of replicates of sample per region.

The average PRCRE per sample and per region was computed representing the bootstrap estimates.

Table 1 Sample Size and the Number of Replicates per Region

Region	Number of Obser-vations	Sample size	Repli-cates
1	1130	23	49
2	952	22	43
3	1558	25	62
4A	1804	22	82
4B	828	34	24
5	1128	31	36
6	1358	39	35
7	1247	54	23
8	975	38	26

Region	Number of Obser-vations	Sample size	Repli-cates
9	789	46	17
10	863	47	18
11	1017	34	30
12	960	32	30
NCR	2229	21	106
CAR	766	36	21
ARMM	825	15	55
CARAGA	822	35	23

3. Results and Discussion

Table 2 shows the average PRCRE expenditure and standard deviation per region.

Table 2 Average and Standard Deviation of PRCRE Expenditureof Households per Region

Region	Mean	Standard Deviation
1	4079.681	1009.004
2	3367.342	798.2272
3	5696.164	1465.099
4A	5757.702	1389.75
4B	2719.804	804.4785
5	2824.787	802.3046
6	3045.278	973.7635
7	2777.995	1041.217
8	2739.015	856.5918

Region	Mean	Standard Deviation
9	2271.589	807.6131
10	2635.571	947.5737
11	2895.475	858.2081
12	2773.094	805.6761
NCR	7535.61	1777.891
CAR	3502.582	1074.047
ARMM	2928.924	582.6161
CARAGA	2742.263	832.2244

Tables 3 and 4 show the bootstrap estimates and SRSWOR estimates for the average PRCRE expenditure and variance per region, respectively.

Table 3 Bootstrap Estimates for the Average PRCRE Expenditure and Variance

Region	Mean	Standard Deviation	Region	Mean	Standard Deviation
1	4154.343	208.8862	9	2393.563	105.8651
2	3300.59	176.812	10	2681.035	105.1835
3	5939.164	336.1678	11	3212.524	173.2218
4A	5981.771	344.8485	12	2694.85	121.1144
4B	2758.125	145.1399	NCR	7536.116	453.9411
5	2853.57	156.0731	CAR	3861.294	212.1074
6	2841.191	122.5002	ARMM	2939.985	165.964
7	2571.691	119.7942	CARAGA	2781.697	176.9434
8	2772.036	159.6849			

Table 4 SRSWOR Estimates for the Average PRCRE Expenditure and Variance

Region	Mean	Standard Deviation	Region	Mean	Standard Deviation
1	4948.27	4819.125	9	3560.683	4937.756
2	3899.373	3408.967	10	3910.941	6546.153
3	7326.794	7011.069	11	4342.958	4844.233
4A	7571.099	5736.637	12	3434.895	3214.213
4B	3447.544	2487.003	NCR	9484.879	7036.571
5	4038.397	4572.879	CAR	4733.135	4301.972
6	4077.61	3903.692	ARMM	3217.674	2220.356
7	4432.653	8757.496	CARAGA	3811.631	3994.01
8	3996.284	4929.368			

In order to compare the variance estimates, design effect (DEFF) of bootstrap estimation against SRSWOR is computed and is shown in Table 5.

Table 5 shows the small values of design effect which are all less than 0.01. This means that there is a large difference between the variance of bootstrap estimates and SRSWOR estimates. The variance of bootstrap per region is smaller than the variance of SRSWOR resulting to very small design effect of bootstrap.

In order to compare the average PRCRE estimates, the percent difference of the estimated mean of each technique from the actual mean is computed and is shown in Table 6.

The percent difference of the bootstrap mean from the actual mean is lower than the percent difference of SRSWOR mean in each region. The bootstrap estimates are close to the actual population values, with as low as 0.01 absolute percent difference in NCR and as high as 10.95 absolute percent difference in

Region XI. Table 6 also shows that bootstrap estimates are underestimates of the actual mean, except for Regions II, VI, VII, and XII. The SRSWOR estimates are all underestimates of the actual mean with high absolute percent difference ranging from 9.86 to 59.56.

Table 5 Design Effect of Bootstrap Estimation vs. SRSWOR

Region	Variance		DEFF
	SRSWOR	Bootstrap	
1	23223967	43633.44	0.001879
2	11621058	31262.49	0.00269
3	49155091	113008.8	0.002299
4A	32909009	118920.5	0.003614
4B	6185182	21065.6	0.003406
5	20911226	24358.82	0.001165
6	15238814	15006.31	0.000985
7	76693731	14350.64	0.000187
8	24298670	25499.26	0.001049
9	24381433	11207.41	0.00046
10	42852118	11063.56	0.000258
11	23466592	30005.79	0.001279
12	10331165	14668.71	0.00142
NCR	49513335	206062.5	0.004162
CAR	18506962	44989.56	0.002431
ARMM	4929982	27544.04	0.005587
CARAGA	15952119	31308.97	0.001963

Table 6 Percent Difference of Estimated Mean from Actual Mean

Region	MBE	SRSWOR	Region	MBE	SRSWOR
1	-1.83	-21.29	9	-5.37	-56.75
2	1.98	-15.80	10	-1.73	-48.39
3	-4.27	-28.63	11	-10.95	-49.99
4A	-3.89	-31.50	12	2.82	-23.87
4B	-1.41	-26.76	NCR	-0.01	-25.87
5	-1.02	-42.96	CAR	-10.24	-35.13
6	6.70	-33.90	ARMM	-0.38	-9.86
7	7.43	-59.56	CARAGA	-1.44	-39.00
8	-1.21	-45.90			

4. Concluding Remarks

In determining the average personal care and effect expenditure, the bootstrap yields superior estimates than simple random sampling without replacement. Even though the sample only comes from the middle 50% of the population, the estimates are closer to the actual population values compared to the estimates of SRSWOR where the sample is drawn from the whole population. Moreover, the estimation procedure has greater gain in precision as shown by the small bootstrap estimates of the standard errors.

REFERENCES

BARRIOS, E., 2008, Model-based predictive estimation in samples with frame problems, *Working Paper No. 2008-04*, UP School of Statistics. Available at *www.stat.upd.edu.ph/*Predictive*%20*Estimation*2008-04.pdf*

BARRIOS, E. and A. KWONG, 2010, Nonparametric model-based predicted estimation in data mining, *Working Paper No. 2010-13*, UP School of Statistics. Available at: www.nscb.gov.ph/.../ABSTRACT_Kwong_Barrios_Nonparametric%20Model%20Based%20Predictive%20Estimation.pdf

MOONEY, C. and R. DUVAL, 1993, *Bootstrapping: A Nonparametric Approach to Statistical Inference*, London: Sage