DOI : 10.18843/rwjasc/v6i4(1)/04
DOI URL : http://dx.doi.org/10.18843/rwjasc/v6i4(1)/04

THE USE OF PROPENSITY SCORE MATCHING IN ELIMINATING SELF-SELECTION BIAS IN MARKET SURVEYS

Joyce DL. Grajo,

Assistant Professor,
Institute of Statistics, University of the Philippines Los Baños, Institute of Statistics, UPLB, College, Laguna, Philippines

Zita VJ. Albacea,

Professor,
Institute of Statistics, University of the Philippines Los Baños, Institute of Statistics, UPLB, College, Laguna, Philippines

ABSTRACT

For most firms, the conduct of a census in market research is more manageable than a survey. Participation of concerned units in this set-up is usually independently decided so that some units may not actually provide information. As a result, unreliable estimates are usually the basis of a firm's marketing strategy if it fails to adjust for self-selection bias incorporated in the outcomes of the census. This study explores the use of propensity score matching in eliminating self-selection bias in market surveys. An experimental customer satisfaction survey was conducted to replicate the self-selection process. Results showed that with the use of propensity score matching, self-selection bias incorporated in the data was largely reduced. This suggests that given that assumptions on conditional independence and common support were attainable and with proper input and judgment from the researcher, propensity score matching can be a useful tool in minimizing the error of self-selection.

Keywords: survey participation, customer satisfaction survey, conditional independence, common support.

INTRODUCTION:

Market players have several ways of obtaining information to be used in making strategic decisions regarding their products. One of which is the use of surveys. Marketers may choose between a census which enumerates completely every unit in a population and a sample survey which consists of just a part of them. Designing a full-blown statistical survey may prove to be difficult at times especially when no definite frame of intended respondents are available. In addition, this requires economical and organizational efforts that are seen, in many cases, as burdens to small firms. As a common practice, marketers simply send out questionnaires to all intended respondents, hoping they grant answer. However, many of them choose to disregard this request. Consequently, unreliable estimates become basis of the firm's marketing strategy if it fails to adjust for self-selection bias incorporated in the results of the census. Self-selection bias is observed when the unit under study is allowed to independently choose whether or not to participate in a census, determining some amount of non-responses in the process. The units which chose to participate in the census constitute a non-probabilistic sample.

This paper intends to show the results of a study that explores the use of Propensity Score Matching (PSM) in eliminating self-selection bias in market surveys. The propensity score approach was first used by Rosenbaum & Rubin (1983) in observational research to balance treatment and control subjects. Propensity score was then defined as a "conditional probability of exposure to a treatment given observed covariates." In the case of market surveys, the propensity score developed in this paper represents the conditional probability of being a self-selected participant given specific observed covariates. Although corrective strategies demonstrated in this paper is not encouraged particularly if there is a way for market researchers to use probabilistic samples, the study is made to introduce cautious measures if using convenience samples as base for making major or even minor marketing decisions.

THE PROBLEM OF SELF-SELECTION:

Self-selection bias is observed when respondents are allowed to decide entirely for themselves whether to take part in a census or not. The units which chose to participate in the census constitute a non-probabilistic sample. To illustrate how this bias occurs, consider a finite population of N units. After the census operation, the population is basically divided into two groups: (1) participating group, and (2) non-participating group. This situation is shown in Figure 1.

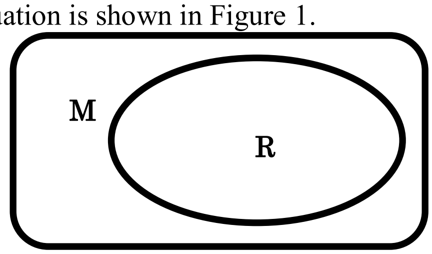

Figure 1. Illustration of self-selection in a census of finite population

Let

$R =$ Number of participating units

$M =$ Number of non-participating units

$N = R + M =$ Population size

$\bar{R} = R/N =$ Proportion of participating units in the population

$\bar{M} = M/N =$ Proportion of non-participating units in the population

Then, the population total and mean may be written as

$$Y_N = Y_R + Y_M = R\bar{Y}_R + M\bar{Y}_M \quad (1) \text{ and}$$

$$\bar{Y}_N = \bar{R}\bar{Y}_R + \bar{M}\bar{Y}_M \qquad (2)$$

where Y_R and $\bar{Y}_R$ are the total and mean of a variable of interest for participating units, respectively, whereas Y_M and $\bar{Y}_M$ are the total and mean of a variable of interest for non-participating units, respectively. If no compensation is made for non-response, the population mean will be declared as $\bar{Y}_R$ when in fact it should be $\bar{Y}_N$. The self-selection bias is computed as,

$$Bias = \bar{Y}_R - \bar{Y}_N = \bar{Y}_R - \left(\bar{R}\bar{Y}_R + \bar{M}\bar{Y}_M\right) = \bar{M}\left(\bar{Y}_R - \bar{Y}_M\right) \qquad (3)$$

The preceding equation (3) suggests that the 'estimated' mean $\bar{Y}_R$ is approximately unbiased for the population mean $\bar{Y}_N$ if either $\bar{M}$ is small or the mean for participating units is close to that of the non-participating units. Since there is no way to control the difference between the means for participating and non-participating units, the only way to ensure that the selection bias is small is by increasing the rate of participation (or decreasing the rate of non-participation). However, this cannot also be done, since participation is assumed to be independently decided, particularly in marketing surveys.

PROPENSITY SCORE MATCHING IN CUSTOMER SATISFACTION SURVEYS:

With the objective of eliminating self-selection bias, the following steps were formulated using the principles and concepts of propensity score matching (Grajo, 2012). For illustration, the subsequent procedure was designed for the conduct of customer satisfaction surveys.

Step 1: Data Collection on the Covariates of Participation and Outcome:

The first step in the procedure is to collect, from each customer, data relevant to their participation in surveys and censuses as this information will be used in modeling the probability of self-selection. It is also important to note that the variables to be used in modeling the propensity scores should have an effect on the outcome of interest.

Step 2: Computation of the Propensity Scores:

The second step in the procedure is to estimate the propensity score through any standard probability model. For convenience, customers who were self-selected will be referred to from here on as 'participating customers' while those who did not self-select will be referred to as non-participating customers. The simple algorithm for estimating the propensity scores which incorporates the conditional independence assumption developed by Dehejia and Wahba in 2002 was used as basis for the mechanics for computing the propensity scores given below:

a. Start with a parsimonious logit specification to compute the score.

$$P(Y=1) = \frac{e^{\alpha+\sum \beta_i X_i}}{1+e^{\alpha+\sum \beta_i X_i}} \qquad (4)$$

where Y is equal to 1 if unit *i* participated in the survey and is equal to 0, otherwise; **X** are covariates of participation, α is the regression constant; and β is the coefficient of the covariates of participation.

b. Sort data according to computed propensity scores (lowest to highest).

c. Stratify all observations such that within stratum, the computed propensity scores between participating and non-participating customers are not different. For example, start by dividing the observations into strata of equal score range (0-0.2,…, 0.8-1). For each stratum, compare propensity scores between participating and non-participating groups. Keep strata in which the groups have comparable propensity scores. If necessary, combine strata in order to achieve this balance.

d. Check the difference between the distribution of the covariates for participating and non-participating customers. Comparison of measures of central tendency and dispersion is sufficient. Ideally, the distribution of the covariates should be approximately the same across participating and non-participating groups once the propensity scores are controlled for.
   i. If covariates are balanced between participating and non-participating customers for all strata, stop.
   ii. If covariates are not balanced for some stratum, divide the stratum into finer strata and re-evaluate.
   iii. If a covariate is not balanced for many strata, the propensity score is poorly calculated. Modify the logit by adding interaction terms and/or higher order terms of the covariate and re-evaluate.

Step 3: Checking for Common Support Condition:

The third step in the procedure is to enforce the common support condition as given by Rosenbaum and Rubin (1983). This requirement is done by discarding those in the non-participating group who have propensity scores that lie outside the range of scores in the participating group.

Step 4: Matching on Propensity Scores:

The fourth step in the procedure is to match the propensity scores using nearest neighbor matching technique, with replacement. In this matching technique, non-participating customer i is matched with participating customer j such that

$$\left|p_i - p_j\right| = \min_{j\in\{D=1\}} \left\{p_i - p_j\right\} \qquad (5)$$

To be precise, this involves taking each non-participating customer in turn and identifying the participating customer with the closest propensity score.

Step 5: Final Estimation:

The last step in the procedure is to estimate the outcome of interest in the study based on the matched units. The outcome associated to a non-participating customer was computed as the average of the outcomes of his matched neighbors. Suppose, if the firm intends to measure its customers' median satisfaction rating, then the median is computed from the ratings coming from participating customers and matched non-participating customers.

METHODOLOGY IN EMPIRICAL STUDY:

In order to assess the ability of PSM in correcting self-selection bias, the procedure described in the previous section was applied to a census of customer satisfaction. There were two stages of data collection in this empirical validation. The universe of the study for both stages included all registered 1,040 Statistics 1 students in a particular semester. Likewise, both stages made use of self-administered questionnaire. In the first stage of data collection, information on the covariates of survey participation and customer satisfaction was obtained through a census. All students were required to accomplish the questionnaire given to them, since the information that was obtained in this census was used to model the propensity scores. Around two weeks after the first data collection, all students were given a questionnaire asking them to rate their satisfaction on the products and services of a fast food store located near the campus. The students were allowed to choose whether or not to return the accomplished questionnaire, following the self-selection process. Categorical principal component analysis (CATPCA) and factor analysis were used to address the dimensionality problem in the satisfaction survey. An index of customer satisfaction was constructed through multiple regression analysis using optimal scaling with the components derived from CATPCA serving as predictors of customer satisfaction. The procedure developed was then applied to the data collected.

The procedure developed was then applied to the collected data. In particular, standardized difference was used to assess the distance in marginal distribution of the realized covariates. Rosenbaum and Rubin (1985) suggested that standardized difference greater than 20 percent should be considered as "large". Bias Reduction (BR) attributed to matching on propensity scores was also computed through the function

$$BR = 100\left(1 - \frac{b_m}{b_i}\right) \tag{6}$$

$$\text{with} \quad b_m = \frac{100\left|\overline{x}_{MC} - \overline{x}_{MT}\right|}{\sqrt{\left(s_{MC}^2 + s_{MT}^2\right)/2}} \quad \text{and} \quad b_i = \frac{100\left|\overline{x}_{BC} - \overline{x}_{BT}\right|}{\sqrt{\left(s_{BC}^2 + s_{BT}^2\right)/2}}$$

where b_m is the standardized difference of the matched covariate; b_i is the standardized difference of the unmatched covariate; $\overline{x}_{MC}$ is the covariate mean of the matched non-participating customers; $\overline{x}_{MT}$ is the covariate mean of the matched participating customers; s_{MC}^2 is the covariate variance of the matched non-participating customers; s_{MT}^2 is the covariate variance of the matched participating customers; $\overline{x}_{BC}$ is the covariate mean of the unmatched non-participating customers; $\overline{x}_{BT}$ is the covariate mean of the unmatched participating customers; s_{BC}^2 is the covariate variance of unmatched the non-participating customers; and s_{BT}^2 is the covariate variance of unmatched the participating customers.

For covariates of binary nature, the standardized difference was computed as

$$SB = \frac{100\left|p_C - p_T\right|}{\sqrt{\left(p_C\left(1-p_C\right) + p_T\left(1-p_T\right)\right)/2}} \tag{7}$$

where p_C is the proportion of the covariate in the non-participating customers; and p_T is the proportion of the covariate in the participating customers. Most empirical studies show that a bias reduction of 3 to 5 percent is deemed sufficient.

RESULTS OF THE EMPIRICAL STUDY:
PROPENSITY SCORES OF THE CUSTOMERS:

A total of 1,040 customers were considered for the study, of which 778 participated in the census of customer satisfaction. This gives a non-self-selection rate of about 25%. Correlation analysis was performed to identify the variables that have high association with participation. Table 1 summarizes the results of this analysis in which participation is shown to be highly related to age, sex, whether or not the customer is freshman, number of hours spent in a week on extracurricular activities, whether or not the customer believes in the usefulness of surveys in gathering information, usual daily amount spent on snacks, usual daily amount spent on dinner, and whether or not the customer prefers eating at fast food places. From the odds ratio, it can be inferred that male customers are less likely to participate in customer satisfaction surveys. The same is the case if a customer is an upper classman and older, spends more time on extracurricular activities, and prefers eating at fast food places. On the other hand, a customer who believes in the usefulness of surveys in gathering information, spends less on snacks, and spends more on dinner is more probable to participate.

Table 1: Covariates of participation and their odds ratios

Variable	Odds Ratio
Age	0.895
Sex (1 – Male, 0 – Female)	0.745
Classification (1 – Freshman, 0 – Otherwise)	1.511
College Address (1 – Within UPLB/*Batong Malake* 0 – Outside UPLB/*Batong Malake*)	1.356
Membership to organization (1 – Yes, 0 – No)	0.760
Number of hours in a week spent studying	1.005
Number of hours in a week spent on extracurricular activities	0.988
Surveys are useful in gathering information. (1 – Yes, 0 – No)	2.745
Surveys require time. (1 – Yes, 0 – No)	0.942
Surveys are not my concern. (1 – Yes, 0 – No)	0.895
Surveys are fun and informative. (1 – Yes, 0 – No)	1.101
Surveys are tedious to do. (1 – Yes, 0 – No)	0.837
Conducted an actual survey? (1 – Yes, 0 – No)	1.163
Survey participation (how many times have you participated in an actual survey?)	0.821
Survey refusal (How many times have you refused to participate in an actual survey?)	0.765
Willingness to participate in another survey (1 – Yes, 0 – No)	1.392
Customer satisfaction survey experience (1 – Yes, 0 – No)	1.090
Usual weekly food allowance	1.000
Usual daily amount spent on breakfast	1.002
Usual daily amount spent on snacks	0.995
Usual daily amount spent on lunch	1.000
Usual daily amount spent on dinner	1.019
Preferred place to eat at (1 – Fast food place, 0 – Elsewhere)	0.599
Enjoy eating fast food (1 – Yes, 0 – No)	0.956
Frequency of eating fast food (1 – More than thrice a week 0 – At most thrice a week)	0.806
Preferred fast food place (1 – fast food under study , 0 – Others)	1.131

In the model-building process for the probability of participation, all initial models resulted to propensity scores ranging from 0.25 to 1; hence stratification on propensity scores was arbitrarily defined by the divisions 0.25 to 0.5, 0.5 to 0.75 and 0.75 to 1. The final logistic model for predicting the probability of participation as presented in Table 2 is a function of only four variables, namely: *age*, *number of hours in a week spent on extracurricular activities*, *whether or not the customer believes in the usefulness of surveys in gathering information*, and *usual daily amount spent on dinner*. The other variables were omitted from the model upon unbalanced distribution across participating and non-participating groups.

Table 2: Parameters for the logistic model of participation

Covariate	Parameter
Constant	2.011
Age	-0.114
Number of hours in a week spent on extracurricular activities	-0.013
Whether or not the customer believes in the usefulness of surveys in gathering information	1.064
Usual daily amount spent on dinner	0.006
Pseudo $R^2 = 0.262$	

The final model shows that consistent to the odds ratio in Table 1, age and number of hours in a week spent on extracurricular activities have a negative effect on the propensity to participate while belief in the usefulness of surveys in gathering information, and usual daily amount spent on dinner have a positive effect on the propensity to participate. The propensity scores were computed as:

$$P(Y=1)=\frac{e^{2.011-0.114(Cov1)-0.013(Cov2)+1.064(Cov3)+0.006(Cov4)}}{1+e^{2.011-0.114(Cov1)-0.013(Cov2)+1.064(Cov3)+0.006(Cov4)}} \quad (8)$$

where Cov1 is the customer's age; Cov2 is the number of hours in a week the customer spends on extracurricular activities; Cov3 is whether or not the customer believes in the usefulness of surveys in gathering information; and Cov4 is the usual daily amount the customer spends on dinner.

The final logistic model has a percent of correct prediction of around 75%. Likewise, from Table 3, it can be seen that the propensity scores are adequately modeled since the covariate means per stratum are almost equal, particularly in the third stratum. The distributions of the covariates are therefore balanced between participating and non-participating groups for each defined stratum. The resulting propensity scores from the final fitted model range from 0.388 to 0.954 with a mean of 0.748 and a variance of 0.006. The distribution of the propensity scores is skewed to the left, implying that some customers have especially low propensity scores.

Table 3: Covariate means of participating and non-participating groups within strata

Covariates	Participants	Non-Participants
Stratum 1: Propensity scores 0.25 to 0.5		
Age	18.286	19.154
Number of hours in a week spent on extracurricular activities	15.000	22.308
Whether or not the customer believes in the usefulness of surveys in gathering information	0.000	0.077
Usual daily amount spent on dinner	34.286	30.385
Stratum 2: Propensity scores 0.5 to 0.75		
Age	18.983	18.774
Number of hours in a week spent on extracurricular activities	15.079	15.538
Whether or not the customer believes in the usefulness of surveys in gathering information	0.881	0.830
Usual daily amount spent on dinner	36.347	32.217
Stratum 3: Propensity scores 0.75 to 1		
Age	17.453	17.643
Number of hours in a week spent on extracurricular activities	5.8118	5.951
Whether or not the customer believes in the usefulness of surveys in gathering information	1.000	1.000
Usual daily amount spent on dinner	55.756	55.916

MATCHING THE PROPENSITY SCORES:

Since the propensity scores of those in the participating group lie between 0.4731 and 0.9540 while the propensity scores of those in the non-participating group lie between 0.3884 and 0.8585, nine student customers who have propensity scores below 0.4731 were dropped from the analysis. This is done to enforce the common support condition.

Nearest neighbor matching with replacement was performed on the remaining propensity scores. Table 4 gives a quick look of the matching results which is summarized in four cases. The first case is called exact matches in which the propensity scores of the recipient and donor are equal. The second case is single matches where one recipient is matched to one donor. The third case is multiple matches where one recipient is matched to several donors. The last case is just a single match but with notably high absolute difference between propensity scores. In fact, the last case shown in Table 4 is the farthest match in the data set. In summary, the matching procedure produced 142 single matches and 111 multiple matches, of which 146 are exact.

Table 4: Overview of the matching results

Case	Non-Participant Id Number	Propensity Score	Participant Id Number	Propensity Score
1	59028	0.5531	12419	0.5531
2	67813	0.8245	12277	0.8247
3	37667	0.7408	13970 33434 56890	0.7407
4	96026	0.5609	56534	0.5556

Once the matches were completed, a non-participating student customer was assigned the customer satisfaction index of his participating match. In case of multiple matches, the customer satisfaction index of a non-participating customer is computed as the average of the indices of his participating matches. The results of these assignments were summarized in Table 5. Results suggest that there is no obvious difference in the mean customer satisfaction index of participating and non-participating student customers although if no matching was done, the mean index will just be equal to 0.227, the mean index of participating student customers. However, after matching, the resultant overall mean index is equal to 0.226. The only noticeable difference between participating and non-participating groups is in the 1st quartile index. Clearly, non-participating student customers have lower satisfaction indices.

Table 5: Description of the customer satisfaction index after matching

	Minimum	Maximum	Mean	Median	Q1	Q3
Participating	-1.000	1.000	0.227	0.275	0.039	0.439
Non-participating	-1.000	0.967	0.221	0.243	0.103	0.358
Overall	-1.000	1.000	0.226	0.268	0.067	0.417

ASSESSING THE REDUCTION IN SELF-SELECTION BIAS:

Table 6 provides some diagnostic checks on the performance of the developed procedure. Standardized differences between participating and non-participating groups were used to illustrate the reduction in bias brought about by matching on propensity scores and to check the balance between the groups. Before matching, it is evident that there is a large difference in the variable *whether or not the customer believes in the usefulness of surveys in gathering information* between the participating and non-participating groups in the original data, and all standardized differences have values larger than 13%. This is not surprising since one cannot expect individuals in the participating group to resemble the non-participating group in general. These differences are considerably reduced after nearest neighbor matching with replacement except for the covariate *number of hours in a week spent on extracurricular activities*, which was reduced by a minimal degree amounting to 1.835%. Bias reductions were observed to be very large particularly in the variable *usual daily amount spent on dinner*.

Table 6: Bias reduction due to matching on propensity scores

Covariates	Standardized Difference[a] (%)		Bias Reduction (%)
	Unmatched	Matched	
Age	17.049	13.115	23.077
Number of hours in a week spent on extracurricular activities	15.143	14.865	1.835
Whether or not the student believes in the usefulness of surveys in gathering information	25.836	11.150	56.841
Usual daily amount spent on dinner	13.908	2.610	81.236

[a] Standardized difference is the size of the difference in means of a conditioning variable, scaled by the square root of the average of two associated variances and multiplied by 100.

ASSESSMENT OF THE MATCH THROUGH BOOTSTRAP RESAMPLING:

In assessing the performance of propensity score matching in correcting self-selection bias, bootstrap resampling technique was implemented in order to approximate the sampling distribution of the average customer satisfaction index. From a pseudo-population of 1,031 index values, a sample of size *m* was drawn 250, 500, 750 and 1,000 times. For this study, bootstrap resamples corresponding to 10%, 15% and 20% of the pseudo-population was obtained. For each bootstrap sample, the average customer satisfaction index was computed. The mean and variance of the sampling distribution of this average index is presented in Table 7. Not one estimator is exactly unbiased for the true mean index. However, it can be seen that bias is small and almost equal to zero across sample sizes and number of resamples. In fact, when *t*-test was employed, bias is shown to be not significantly different from zero, except for three scenarios where $(B=500,\ m=20)$, $(B=750,\ m=15)$ and $(B=750,\ m=20)$. From this, it can be safely concluded that the estimates are approximately unbiased. The variance of the sampling distribution is very small, such that in the long run, its value approaches zero as sample size is increased. Still, it cannot be stated that the estimates are consistent because the behavior of bias across sample sizes and number of resamples seems to be without obvious direction and in no way it is clearly approaching zero as sample size is increased.

Table 7. Mean and variance of the sampling distribution of the average customer satisfaction index with corresponding bias at different values of sample size *m* and at different number of bootstrap resamples *B*

B	*M*	Average Customer Satisfaction Index			
		Mean	Variance	Bias	*P*-Value
250	104	0.226	0.00074	0.00014	0.937
	155	0.225	0.00051	-0.00055	0.699
	207	0.227	0.00041	0.00151	0.240
500	10	0.225	0.00090	-0.00059	0.660
	15	0.227	0.00059	0.00144	0.184
	20	0.224	0.00038	-0.00173	0.049*
750	10	0.225	0.00076	-0.00010	0.917
	15	0.224	0.00055	-0.00180	0.036*
	20	0.224	0.00037	-0.00162	0.021*
1000	10	0.225	0.00086	-0.00048	0.606
	15	0.225	0.00051	-0.00037	0.609
	20	0.224	0.00039	-0.00107	0.086

* significant at 5% level of significance

CONCLUSION:

From this study, Propensity Score Matching was inferred to be a useful tool in minimizing the error of self-selection provided that there is proper input and judgment from the researcher. The use of probabilistic samples always takes precedence in research works and application of corrective strategies that are demonstrated in this study are only to be done when information gathered through census are marred with self-selection problems. Nonetheless, the empirical validation has shown that prior to PSM, large differences in the values of the covariates of participation in the customer satisfaction survey were observed. These differences are considerably reduced upon matching propensity scores which lead to a more acceptable estimate.

REFERENCES:

[1] AGRESTI, A. (2002). *Categorical Data Analysis*. John Wiley & Sons, Inc., Hoboken, New Jersey. ISBN 0471360937

[2] BRYSON, A., R. DORSETT and S. PURDON. (2002). *The Use of Propensity Score Matching in the Evaluation of Active Labour Market Policies.* A study carried out on behalf of the Department for Work and Pensions, United Kingdom. ISBN 1843880431.

[3] D'AGOSTINO, R.B. Jr. and D.B. RUBIN. (2000). *Estimating and Using Propensity Scores with Partially Missing Data*. Journal of American Statistical Association, 95:451, pp. 749-759.

[4] DEHEJIA, R. (2005). *Practical Propensity Score Matching: A Reply to Smith and Todd*. Journal of Econometrics, 125:1-2, pp. 355-364.

[5] DEHEJIA, R. and S. WAHBA (2002). *Propensity-score matching Methods for Non-experimental Causal Studies*. The Review of Economic and Statistics,84:1, pp. 151-161.

[6] ESQUIVEL, G. and A. HUERTA-PINEDA. (2006). *Remittances and Poverty in Mexico: A Propensity Score Matching Approach*. A paper written under the auspices of the Inter-American Development Bank.

[7] GITMAN, L. and C. MCDANIEL (2005). *The Future of Business: The Essentials*. Mason, Ohio: South-Western. ISBN 0324320280.

[8] GRAJO, J.D.L (2012). *Development of a Procedure to Correct Self-Selection Bias in Market Research.* Unpublished Master's Thesis. University of the Philippines Los Baños. pp. 33-38.

[9] GREER, T.V., N. CHUCHINPRAKARN and S. SWSHADRI. (2000). *Likelihood of Participating in Mail Survey Research: Business Respondents' Perspectives*. Industrial Marketing Management, 29, pp. 97–109.

[10] HECKMAN, J. (1979). *Sample Selection Bias as Specification Error*. Econometrica, 47, pp. 153-161.

[11] HECKMAN, J., H. ICHIMURA and P. TODD (1998). *Matching as an Econometric Evaluation Estimator*. The Review of Economic Studies, 65:2, pp. 261-294.

[12] HEINRICH, C., A. MAFFIOLI and G. VAZQUEZ (2010). *A Primer for Applying Propensity Score Matching*. Impact-Evaluation Guidelines, Inter-American Development Bank Technical Notes.

[13] IMBENS, G. and J. WOOLDRIDGE (2009). *Recent Developments in the Econometrics of Impact Evaluation*. Journal of Economic Literature, 47, pp. 5-86.

[14] LALONDE, R. (1986). *Evaluating the Econometric Evaluations of Training Programs*. American Economic Review, 76:4, pp. 604-620.

[15] LECHNER, M. (2001). *Program Heterogeneity and Propensity-score matching: An Application to the Evaluation of Active Labor Market Policies*. The Review of Economics and Statistics, 84, pp. 205-220.

[16] LEE, W.S. (2011). *Propensity Score Matching and Variations on the Balancing Test*. Empirical Economics. Springer, pp. 1-34.

[17] NICOLINI G. and L. DALLA VALLE. (2011). *Errors in Customer Satisfaction Surveys and Methods to Correct Self-Selection Bias*. Quality Technology & Quantitative Management, 4:2, pp. 167-181.

[18] OBERSKI, D. (2008). *Self-selection Bias Versus Nonresponse Bias in the Perceptions of Mobility Survey*. The Hague. The Netherlands Institute for Social Research. ISBN 9789037703436

[19] PHAU, I. and M. BAIRD. (2008). *Complainers Versus Non-complainers Retaliatory Responses Towards Service Dissatisfactions*. Marketing Intelligence and Planning, 26:6, pp. 587-604.

[20] ROSENBAUM, P. and D. RUBIN. (1983). *The Central Role of the Propensity Score in Observational Studies for Causal Effects*. Biometrika, 70, pp. 41-55.

[21] ROSENBAUM, P. and D. RUBIN. (1985). *Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score*. American Statistician, 3, pp. 33-38.
[22] RUBIN, D. (1973). *Matching to Remove Bias in Observational Studies*. Biometrics, 29, pp. 159–183.
[23] SMITH, J. and P. TODD. (2005). *Rejoinder to Dehejia*. Journal of Econometrics, 125, pp. 365-375.
[24] SOLIVAS, E., G. RAMIREZ and A. MANALO. (2007). *The Propensity Score Matching for Correcting Sample Selection Bias*. A paper presented at the 10th National Convention on Statistics, EDSA Shangri-La Hotel, Mandaluyong City, Philippines.
[25] XIANGHONG, L. (2004). *Three Applications of Propensity Score Matching in Microeconomics and Corporate Finance: US International Migration; Seasoned Equity Offerings; Attrition in a Randomized Experiment*. PhD Dissertation, Ohio State University.
