Potentiality of propensity scores methods in weighting for Web surveys: a simulation study based on a statistical register

Silvia Biffignandi and Daniele Toninelli
University of Bergamo, Dipartimento di Matematica, Statistica, Informatica e Applicazioni
Via dei Caniana, 2
Bergamo 24129, Italia
silvia.biffignandi@unibg.it

Monica Pratesi
University of Pisa, Dipartmento di Statistica e Matematica Applicata all'Economia
Via Cosimo Ridolfi, 10
Pisa 56124, Italia
m.pratesi@ec.unipi.it

1. Introduction

Self-selected Web surveys are very common in many fields of application and despite the large number of respondents they often not resemble the intended target population (Couper 2000).

In this paper we propose to compensate the self-selection bias by applying propensity scores adjustement methods (Taylor 2000, Terhanian et al, 2001). These methods are used for the self-selection bias occurred in a Web survey on establishments (Biffignandi and Pratesi, 2002). The attention is on the prospects and on the limits of propensity score as a procedure for weighting the results of the Web survey.

2. Methodology of the study

Propensity scores is a statistical technique aiming to compare two populations by simultaneously controlling for all variables that are thought to affect the comparison (Rosenbaum and Rubin, 1983).

Using a parallel statistical register (derived from the Chamber of Commerce register), the probability of being in the Web survey is estimated, based on a vector of covariates measured both in the register and in the survey. The success of the approach depends on the choice of variables used in the adjustment and on the quality of the benchmark measure (the statistical register). Potentialities and limitation of the method have been explored through a simulation study, based on a pseudo-population obtained from the register itself. The study provides evidence about the amount of bias we can correct with the method, the variability of the Web survey estimates, and the effect on the estimates of the omission and of the addition of relevant covariates in the propensity model.

3. Empirical results

In our perspective, where the convenience sample is considered as a random set of units with unknown individual probability of being included in the set, the inference problem is strictly connected with the estimation of the probability of being a respondent in the Web survey $E(R_i) = q_i$. If this probability would be known, the inference problem could be solved using the traditional tools of estimation from random samples.

Having this in mind, we estimated through propensity scores methodology the individual probabilities of participating in the Web surveys. This probability has allowed for the estimation of the moments of the y variable under study. The estimation is carried out following an "Horvitz-Thompson" type approach. In the economy of these paper we do not describe in detail the assumptions underlying our methodology of estimation and the promising results of our preliminary

research: Table 1 summarizes the empirical sampling distribution of the estimators of the mean and of the standard deviance of the target population. The procedure is able to estimate the parameters of the target with a not relevant relative bias both for the Web population and the whole population. The methodology and the complete set of results is described in Biffignandi S. and Pratesi M. (2003). For the propensity matching algorithm and its performance see Parsons L. S. (2001).

Table 1: Horvitz- Thompson approach

Estimator T_p	**Sampling distribution**		**Target: whole population**			**Target: Web population**		
	Mean	*Std Dev*	T_p / ϑ	*Bias*	*MSE*	T_p / ϑ	*Bias*	*MSE*
$\bar{y}_p$	9.88	0.80	1.159	1.36	2.48	1.021	0.05	0.64
$\sigma(y)_p$	3.27	0.38	1.005	0.07	0.15	0.939	-0.213	0.19

REFERENCES

Biffignandi S., Pratesi M., (2002), "Internet surveys: the role of time in italian firms response behaviour", *Journal of Research in Official Statistics*, Volume 5, number 2.

Biffignandi S., Pratesi M., (2003), "Potentiality of propensity scores in weighting for Web survey: a simulation study based on a statistical register", *Quaderni del Dipartimento di Matematica, Statistica, Informatica e Applicazioni,* Università di Bergamo n.1.

Couper M. P. (2000), "Web Surveys, A review of Issues and Approaches", *Public Opinion Quarterly*, vol. 64, 464-494.

Parsons LS. "Reducing Bias in a Propensity Score Matched-Pair Sample Using Greedy Matching Techniques". Proceedings of the Twenty-Sixth Annual SAS Users Group International Conference, Cary, NC: SAS Institute Inc., 2001

Rosenbaum, P.R, and D.B. Rubin, (1983), "The central role of the propensity score in Observartional studies for Causal Effect", *Biometrik*a, vol 70, 41-55

Taylor H., (2000), "Does Internet research work? Comparing Online survey Result with telephone Survey" *International Journal of Market Research*, 42 (1) 58-63

Terhanian G., R. Smith,J. Bremer, R.K. Thomas, (2001), "Exploiting Analytical Advances: Minimizing the biases Associated with Internet –Based Surveys of Non –Random Samples", ARF/ESOMAR: Worldwide Online Measurement, ESOMAR Publication Services, vol. 248, 247-272.

RESUMÉ

L'article propose de réparer l'autosélection dans les enquêtes faite voie Web avec l'application de la méthode des « propensity scores ». Les données d'une enquête prés les enterprises sont analysée. L'attention est sur les avantages et le limites des « propensity scores » afin de peser le résultats des enquêtes faites voie Web.

AKNOWLEDGMENTS

The presentation and the partecipation to the ISI meeting is supported by the 60% 97 University of Bergamo research fund and by the 40% COFIN 2001 national research project.