Journal of Business Research 66 (2013) 1272–1278

Contents lists available at SciVerse ScienceDirect

Journal of Business Research

Comparing methods to separate treatment from self-selection effects in an online banking setting ☆

Sonja Gensler [a,*], Peter Leeflang [a,b,1], Bernd Skiera [c,2]

[a] Department of Marketing, University of Groningen, Post Box 800, 9700 AV Groningen, The Netherlands
[b] LUISS Guido Carli, Rome, Italy
[c] Electronic Commerce, Department of Marketing, University of Frankfurt, Grueneburgplatz 1, 60323 Frankfurt am Main, Germany

ARTICLE INFO

Article history:
Received 1 September 2011
Received in revised form 1 November 2011
Accepted 1 January 2012
Available online 3 March 2012

Keywords:
Self-selection effects
Matching methods
Instrumental variables method
Control function method
Online use

ABSTRACT

The literature discusses several methods to control for self-selection effects but provides little guidance on which method to use in a setting with a limited number of variables. The authors theoretically compare and empirically assess the performance of different matching methods and instrumental variable and control function methods in this type of setting by investigating the effect of online banking on product usage. Hybrid matching in combination with the Gaussian kernel algorithm outperforms the other methods with respect to predictive validity. The empirical finding of large self-selection effects indicates the importance of controlling for these effects when assessing the effectiveness of marketing activities.

1. Introduction

In 2005, Bank of America claimed that not only were its 12.6 million online customers 27% more profitable than their offline counterparts, but these online users also carried higher balances (Tedeschi, 2005). This statement might motivate other bank managers to conclude that moving customers to online channels might improve customer profitability by stimulating product usage. However, the statement only indicates that online banking customers are more profitable and carry higher balances; not that online banking makes customers more profitable or leads to higher balances. Customer characteristics, such as age, might drive the adoption of an online channel and also cause differences in customer profitability and balances held (Shankar, Smith, & Rangaswamy, 2003). If so, the difference in balances likely reflects self-selection effects, not the effect of online use.

A significant stream of research in economics and econometrics proposes methods to control for self-selection effects by using instrumental variables, control functions, or matching methods (see, e.g., the *Journal of Econometrics*, Issue 125, 2005). Yet surprisingly, most studies that consider self-selection effects use only one method without comparing that method to alternative approaches (e.g., Leenheer, Van Heerde, Bijmolt, & Smidts, 2007). The few studies that compare different methods find ambiguous results with respect to the performance of those methods (e.g., Blundell, Dearden, & Sianesi, 2005; Heckman & Navarro-Lozano, 2004; Zhao, 2004).

Further, studies in economics and econometrics mostly use extensive survey data that has up to 20 consumer characteristics to control for self-selection effects (e.g., Dehejia & Wahba, 2002). However, marketers have access primarily to cross-sectional transaction data that usually contain only a limited number of characteristics beyond a customer's buying behavior. This data structure raises the question about which method performs best with only limited information about customers in order to control for self-selection. None of the previous studies addresses this question. The answer to this question is critical because self-selection effects can affect many marketing decisions such as the decision whether to stimulate customers to move to another channel or use the loyalty card. Ignoring self-selection effects might result in inaccurate managerial decisions.

The objective of this research is to theoretically compare different methods that control for self-selection effects (matching, instrumental variables, and control functions) and empirically assess the performance of these methods in a situation in which large self-selection effects are likely and few customer characteristics are available. For this purpose, the authors use cross-sectional transaction data from a sample of 200,000 customers of a large European retail bank and investigate the effect of online banking on product usage. Thus, this article aims to contribute to the literature by providing state-of-the-art knowledge on how to control for self-selection effects in

☆ The authors thank the two anonymous reviewers and Tammo Bijmolt for their feedback on prior versions of this manuscript. We also thank the E-Finance Lab at the House of Finance in Frankfurt, Germany, for financial support.

* Corresponding author. Tel.: +31 50 363 9086; fax: +31 50 363 3720.
E-mail addresses: S.Gensler@rug.nl (S. Gensler), p.s.h.leeflang@rug.nl (P. Leeflang), skiera@skiera.de (B. Skiera).
[1] Tel.: +31 50 363 3696; fax: +31 50 363 3720.
[2] Tel.: +49 69 798 34649; fax: +49 69 798 35001.

0148-2963/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jbusres.2012.02.025

situations with little additional information about customers to control for self-selection.

2. Definition of treatment and self-selection effect

In this empirical study, the treatment effect refers to the effect of using online banking on different banking services such as checking account balances or brokerage account turnover (outcome variables). The main problem in identifying the treatment effect is that the outcome variables appear in either the treated or untreated condition but never in both. For example, researchers might observe checking account balances for customers who use online banking (treated condition) but not the potential balances of these same customers if they did not use online banking (untreated condition). An argument exists that the observed value of the outcome variables for the customers who do not use online banking can serve as an estimate for the counterfactual outcome that is missing in the untreated condition. But the average online and offline banking customers might differ in their characteristics because they self-select whether to use or not to use online banking respectively. Thus, simply measuring the average differences of the outcome variables between online and offline customers actually capture both the effect of using online banking (treatment effect) and the difference in characteristics (self-selection effect).

Estimating the treatment effect at the individual level is impossible; thus, the focus must center on the average treatment effect. The average treatment effect for customers who participate in the treatment (average treatment on the treated effect [ATTE]) is the variable of interest. In this study, ATTE refers to the effect of using online banking on product usage for customers who actually use online banking:

$$ATTE_k = E_i\left(y^1_{i,k}|d_i = 1\right) - E_i\left(y^0_{i,k}|d_i = 1\right) = E_i\left(y^1_{i,k} - y^0_{i,k}|d_i = 1\right) \;\forall k \in K, \tag{1}$$

where $E_i(y^1_{i,k}|d_i = 1)$ is the expected value of all treated customers i for (observed) outcome variable k (e.g., checking account balance), and $E_i(y^0_{i,k}|d_i = 1)$ is the expected value of all treated customers i for (unobserved) outcome variable k if they were not treated (Table 1). The latter is the missing (counterfactual) outcome in Eq. (1).

As stated previously, using the expected outcome for untreated customers j ($j \neq i$) ($E_j(y^0_{j,k}|d_j = 0)$) is only a valid estimate for the counterfactual outcome in Eq. (1) if no self-selection effects exist; for example, when people receive random assignments to the treatment in an experiment and the characteristics of the treated and untreated customers are comparable. If this condition does not hold, then researchers must apply methods to control for self-selection effects; otherwise, the estimated ATTE will be biased (Heckman & Navarro-Lozano, 2004). This bias corresponds to the average self-selection effect (SE):

$$SE_k = \left[E_i\left(y^1_{i,k}\middle|d_i = 1\right) - E_j\left(y^0_{j,k}\middle|d_j = 0\right)\right] - E_i\left(y^1_{i,k} - y^0_{i,k}\middle|d_i = 1\right) \quad \forall k \in K. \tag{2}$$

The first term on the right-hand side of Eq. (2) equals the expected difference between treated and untreated customers (observed mean difference), and the second term represents the average treatment on the treated effect (ATTE). Rewriting Eq. (2) leads to:

$$\left[E_i\left(y^1_{i,k}\middle|d_i = 1\right) - E_j\left(y^0_{j,k}\middle|d_j = 0\right)\right] = ATTE_k + SE_k \quad \forall k \in K. \tag{3}$$

Eq. (3) demonstrates that the mean difference between treated and untreated customers can be larger or smaller than ATTE depending on the size of the self-selection effect.

Table 1
Notation for observed and unobserved outcomes.

	Observed outcome	Unobserved outcome
Treatment: treated customers	$E_i(y^1_{i,k}\|d_i = 1)$	$E_i(y^0_{i,k}\|d_i = 1)$
Control: untreated customers	$E_j(y^0_{j,k}\|d_j = 0)$	$E_j(y^1_{j,k}\|d_j = 0)$

3. Methods to control for self-selection effects

3.1. Matching methods

Matching methods attempt to eliminate self-selection effects by comparing customers with similar observed characteristics. Thus, these methods rebuild the design of an experimental study by pairing treated and untreated customers who have comparable characteristics but not treatments. The outcome from matching untreated customers provides an estimate of the counterfactual outcome and, hence, the average difference between the matched customers provides an estimate of the treatment effect (Caliendo & Kopeinig, 2008).

The observed characteristics must be informative enough that controlling for them is sufficient to remove any self-selection effect (selection on observables). This so-called conditional independence assumption implies that the outcome variables must be independent of the treatment and conditional on the characteristics (Rosenbaum & Rubin, 1983). Theory and previous research should guide the selection of appropriate characteristics, because researchers can not formally test the assumption (Smith & Todd, 2005).

Covariate matching pairs treated and untreated customers who are similar with respect to individual characteristics and therefore is an intuitive approach to control for self-selection effects (for applications in marketing see, e.g., Hitt & Frei, 2002; Degeratu, Rangaswamy, & Wu, 2000; Shankar et al., 2003). For matching on individual-specific characteristics, the $ATTE^{cov}$ for every outcome variable k is (Dehejia & Wahba, 2002):

$$ATTE^{cov}_k = E_i\left(y^1_{i,k}|d_i = 1, z_i\right) - E_j\left(y^0_{j,k}|d_j = 0, z_j\right) \quad \forall k \in K, \tag{4}$$

where $z_{i(j)}$ is a vector of observed characteristics for treated (untreated) customer i (j), and $E_j(y^0_{j,k}|d_j = 0, z_j)$ is the average value of the outcome variable k for the matched untreated customers, which represents the estimate for the counterfactual outcome.

A wealth of characteristics might make it impractical to match directly on multiple characteristics, because the consideration of many different characteristics increases the difficulty of finding treated and untreated customers who have the same characteristics. In this case, mapping the multiple characteristics onto a single number through a metric such as the Mahalanobis distance is useful (e.g., Zhao, 2004).

Another way to reduce the number of characteristics is *propensity score matching*, which represents the state-of-the-art method in economics and econometrics (e.g., Caliendo & Kopeinig, 2008; Dehejia & Wahba, 2002). However, few studies in marketing apply this method (Campbell & Frei, 2010; Mithas, Krishnan, & Fornell, 2005; Von Wangenheim & Bayon, 2007). Propensity score matching uses the conditional probability that a customer with particular observed characteristics participates in the treatment. The propensity score $\hat{p}(z)$ is a function of the observed characteristics where the conditional distribution of z, given the propensity score, is the same for the treated and untreated groups (Rosenbaum & Rubin, 1983). In this study's setting, the conditional probability involves whether a customer with particular observed characteristics uses online banking. A logit or probit model estimates the propensity score. Implementing a common support restriction further ensures that treated and