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:

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

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 untreated customers are comparable by ignoring the tails of the propensity score distributions for the two groups that do not overlap.

The ATTE^{ps} for an outcome variable k when using propensity score matching results in:

$$ATTE_k^{ps} = E_i\left(y_{i,k}^1\middle|d_i = 1, \hat{p}(z_i)\right) - E_j\left(y_{j,k}^0\middle|d_j = 0, \hat{p}\left(z_j\right)\right) \quad \forall k \in K, \tag{5}$$

where $\hat{p}()$ is the estimated propensity score, and $E_j\left(y_{j,k}^0\middle|d_j = 0, \hat{p}\left(z_j\right)\right)$ is the estimate for the counterfactual outcome.

However, propensity score matching does not guarantee that the matched treated and untreated customers are comparable with respect to their characteristics. *Hybrid matching* instead combines treated and untreated customers by using the propensity score and selected characteristics z^m that represent a subsample of z and appear to be important drivers of the self-selection process (Zhao, 2004):

$$ATTE_k^{hybrid} = E_i\left(y_{i,k}^1|d_i = 1, \hat{p}(z_i), z_i^m\right) - E_j\left(y_{j,k}^0|d_j = 0, \hat{p}\left(z_j\right), z_j^m\right) \quad \forall k \in K, \tag{6}$$

where $z_{i(j)}^m$ is the vector of selected characteristics for treated (untreated) customer i (j) ($z_{i(j)}^m \in z_{i(j)}$). By considering selected characteristics to build the matched samples, researchers can improve the accuracy of the estimated ATTE by ensuring that treated and untreated customers are similar not only with respect to the probability of participating but also in terms of the selected characteristics. In expressing the similarity between treated and untreated customers, Mahalanobis distance can ensure that the dimension of the conditioning does not increase (Rosenbaum & Rubin, 1985). Yet hybrid matching rarely appears in economics and econometrics, with the exception of studies by Zhao (2004), Rubin and Thomas (1996), and Rosenbaum and Rubin (1985). No marketing studies apply this approach to the best of the authors' knowledge.

In addition to the decision about which matching method to use, researchers must determine how many matching partners to consider for a particular treated customer. Popular algorithms are the one-nearest neighbor algorithm, which considers just one matching partner; the n-nearest neighbor algorithm that includes n matching partners; and the Gaussian kernel algorithm, which considers all untreated customers as matching partners (Caliendo & Kopeinig, 2008; Smith & Todd, 2005). The Gaussian kernel algorithm assigns a weight to every untreated customer on the basis of the distance to the treated customer; the smaller the distance, the larger the weight.

When considering more than one matching partner, bias in the estimated ATTE increases because of potentially poorer matches, but this method also decreases the variance (Caliendo & Kopeinig, 2008). Using more than one matching partner to construct the counterfactual outcome for each treated customer therefore is appropriate if many untreated customers are similar to the treated one. Although all algorithms asymptotically should yield the same results, their performance obviously depends on the data structure, for example, the similarity of treated and untreated customers (Zhao, 2004).

3.2. Instrumental variable method

Another method to control for self-selection effects is the instrumental variable (IV) method. The basis for this method is a regression model that explains customer h's (treated or untreated) outcome variable y$_{h,k}$:

$$y_{h,k} = \beta_{0,k} + \beta_{d,k} \cdot d_h + \beta_{x,k} \cdot x_h + \varepsilon_{h,k} \quad \forall h \in (I \cup J), k \in K, \tag{7}$$

where $\beta_{0,k}$ is the constant term for outcome variable k, $\beta_{d,k}$ equals the parameter for the effect of the binary treatment variable d for outcome variable k, $\beta_{x,k}$ is a vector of parameters for characteristics x for outcome variable k, x_h is a vector of customer h's characteristics, and $\varepsilon_{h,k}$ refers to the error term for customer h for outcome variable k.

Eq. (7) reflects the so-called outcome equation, and estimating this equation with ordinary least squares (OLS) leads to biased estimates; because the decision to participate in treatment d is endogenous when self-selection effects exist. The IV method controls for self-selection effects by modeling the decision to participate (here, to use online banking) as a function of customer characteristics through the selection equation:

$$d_h = \begin{cases} 1, if\ d_h^* = \gamma \cdot z_h + \eta_h > 0 \\ 0, otherwise. \end{cases} \quad \forall h \in (I \cup J), \tag{8}$$

where γ is a vector of parameters for customer characteristics z_h, and η_h is the error term for customer h.

The IV method also distinguishes between characteristics x that affect the outcome variable y and characteristics z that determine the decision to participate in treatment d. The characteristics x appear in z, but the number of characteristics in z that are not in x ($x \subset z$) must equal the number of endogenous variables: the exclusion restriction (Amemiya, 1985). This setting contains one endogenous variable, so the requirement is for at least one instrumental variable. The instrumental variables must satisfy two conditions: (i) be exogenous and uncorrelated with the error term ε in Eq. (7) and (ii) be highly correlated with the binary decision variable d in Eq. (8) (Stock, Wright, & Yogo, 2002). Stated differently, an instrumental variable must affect the probability of participating in the treatment but not the outcome variable.

The most common estimator of ATTE using the IV method is 2SLS, which first estimates the propensity score ($\hat{p}(z_h)$) on the basis of selection Eq. (8). In the second stage, the modeling of the outcome equation is a function of the customer characteristics and the estimated propensity score:

$$y_{h,k} = \beta_{0,k} + \beta_{d,k} \cdot \hat{p}(z_h) + \beta_{x,k} \cdot x_h + \varepsilon_{h,k} \quad \forall h \in (I \cup J),\ k \in K. \tag{9}$$

In line with the literature, ε and η have normal distributions with means of zero and a non-zero covariance (Amemiya, 1985). The parameter $\beta_{d,k}$ then equals the ATTE for an outcome variable k (Heckman & Navarro-Lozano, 2004).

3.3. Control function method

The control function method relies on Eqs. (7) and (8) but requires no exclusion restriction in the parametric implementation (Heckman & Navarro-Lozano, 2004), because this method does not replace the endogenous decision variable with the propensity score. This approach instead includes an estimate of the conditional mean of the error term in the outcome equation $\left(\hat{E}_h(\varepsilon_h|d_h, x_h, z_h\right)$ to control for self-selection effects (Heckman & Robb, 1985):

$$y_{h,k} = \beta_{0,k} + \beta_{d,k} \cdot d_h + \beta_{x,k} \cdot x_h + \hat{E}_{h,k}(\varepsilon_{h,k}|d_h, x_h, z_h) + \varepsilon_{h,k}^* \ \forall h \in (I \cup J), k \in K, \tag{10}$$

where $\varepsilon_{h,k}{}^* = \{\varepsilon_{h,k} - E_{h,k}(\varepsilon_{h,k}|d_h, x_h, z_h)\}$.

Assuming a joint normal distribution for the error terms and rewriting the model in the form of a switching regression provides the following expression for the conditional mean of the error term (Vella & Verbeek, 1999):

$$\hat{E}_{h,k}\left(\varepsilon_{h,k}|d_h, x_h, z_h\right) = \frac{Cov(\varepsilon_k, \eta_k)}{Var(\eta_k)} \cdot \frac{\phi(z_h \cdot \gamma)}{\Phi(z_h \cdot \gamma)} \cdot d_h + \frac{Cov(\varepsilon_k, \eta_k)}{Var(\eta_k)} \cdot \frac{-\phi(z_h \cdot \gamma)}{1 - \Phi(z_h \cdot \gamma)} \cdot (1 - d_h), \tag{11}$$

where $\phi(z_h \cdot \gamma)$ and $\Phi(z_h \cdot \gamma)$ are the density function and distribution function of the standard normal distribution respectively (Vella & Verbeek, 1999). The control function method explicitly examines the correlation of the two error terms; the size of the correlation indicates the size of the bias due to self-selection effects. A two-stage estimator can estimate the parameters. First, a probit model estimates the decision to participate (propensity score) according to the maximum likelihood. Second, the estimates of $\phi(z_h \cdot \gamma)$ and $\Phi(z_h \cdot \gamma)$ from the first-stage selection equation appear in Eq. (10), which can then be estimated by OLS, and yields $\frac{Cov(\varepsilon_k, \eta_k)}{Var(\eta_k)}$ (Vella & Verbeek, 1999). The ATTE for an outcome variable k equals $\beta_{d,k}$.

3.4. Comparison of methods

All methods (except covariate matching) use the propensity score in some way to control for self-selection effects. However, differences mark the way each method controls for self-selection effects. The most crucial difference is whether a method relies on the assumption of selection on observable characteristics. Matching methods assume that the observed characteristics account for the self-selection effect, whereas the IV and control function methods assume that unobserved characteristics influence the self-selection effect. However, researchers cannot test whether self-selection effects are due to observable or unobservable characteristics.

Furthermore, the methods differ with respect to an exclusion restriction in that matching methods do not use one. Strictly speaking, the control function method does not require an exclusion restriction, but Puhani (2000) shows that failing to take an exclusion restriction into account can result in poor performance. The IV method relies on identifying appropriate instrumental variables to estimate the ATTE accurately, but no clear guidelines indicate how to do so. Further, when the use of instrumental variables does not satisfy the two conditions (exogenous and uncorrelated with the error term; highly correlated with the binary decision), then their use not only reduces the precision of the IV estimates but also leads to biases and inconsistencies that can be greater than the biases of the OLS estimates (e.g., Staiger & Stock, 1997; Woglom, 2001). Overall, the inability to test identifying assumptions creates difficulties for assessing theoretically which method is the most appropriate in a specific setting.

4. Empirical study

This empirical study compares the performance of the different methods (covariate matching, propensity score matching, hybrid matching, IV method, control functions method, and OLS) in a situation with only a limited number of available customer characteristics by assessing their predictive validity for determining the effects of using online banking on product usage.

4.1. Data

The transaction data comes from a large European retail bank and refers to 200,000 customers of whom 1,628 are active online customers. A customer qualifies as being active online if he or she makes more than one online transaction during the observation period: July through September 2003. Checking and savings account balances, brokerage and credit card account turnover, and the number of checking accounts, brokerage accounts, credit cards, and transactions (per channel) represent product usage. Information about these determinants of a customer's product usage is available on a monthly basis (the effective balance date is the last day of the month). To avoid biases due to individual fluctuations in product usage, this study uses the average monthly value in a quarter for the analyses and eliminates outliers according to the method proposed by Hadi (1994). Hadi's method detects outliers in multivariate samples by avoiding masking (i.e., not detecting an outlier because of the presence of others) and swamping (i.e., wrongly identifying an outlier due to the effect of some hidden outliers).

Furthermore, available information about individual-specific characteristics includes a customer's age, length of relationship with the bank, whether the checking account is a joint account, number of different product categories used, and number of savings accounts. The number of savings accounts is not considered as an outcome variable, because the retail bank does not charge fees and thus does not generate revenue from savings account ownership alone. The number of different product categories used indicates the strength of the customer's relationship with the retail bank; this variable thus also represents an individual-specific characteristic.

4.2. Estimating the propensity score

Propensity score matching, hybrid matching, the IV, and control function method all rely on the propensity score. The estimate of this score relies on the customer's age (AGE), length of relationship (LOR), whether the checking account is a joint account (JOINT), number of savings accounts (NOSAV), and the number of different product categories used (NOP). Prior studies show that online customers tend to be younger and have shorter bank relationships (e.g., Cortiñas, Chocarro, & Villanueva, 2010; Degeratu et al., 2000; Shankar et al., 2003). Further, customers using different bank services likely use online banking for convenience (Campbell & Frei, 2010). For the same reason, customers who have a joint account might be more likely to use online banking. Customers who use different savings accounts probably are less likely to use online banking, which is in line with a general preference for traditional, conservative types of investments. The estimated effects for the propensity score exhibit validity and are significant (see Table 2).

Additional characteristics can determine self-selection effects, such as the customer's income or attitude toward the Internet (e.g., Danaher, Wilson, & Davis, 2003), but such information is rarely available in transaction data and is not available in this study's dataset.

4.3. Estimating the effect of using online banking on product usage

For *covariate matching*, the estimation of the effect of using online banking on product usage uses Eq. (4) and the customer characteristics mentioned in the previous subsection. The Mahalanobis distance facilitates the matching process that determines the similarity of online and offline customers (Zhao, 2004). The *propensity score matching* relies on Eq. (5) by using the propensity score estimates from the model described in Section 4.2. The basis for the *hybrid matching* is the propensity score and a customer's age and the length of the relationship, because several studies show that these two characteristics strongly influence online use (e.g., Hitt & Frei, 2002). The Mahalanobis distance again determines the similarity of online and offline customers.

Table 2
Estimated parameters of customer characteristics in propensity score model (binary probit model).

	Parameter	z-value
Age (years)	−0.03	−10.56
Length of relationship (month)	−0.01	−19.85
Joint account (dummy variable)	0.14	2.60
Number of product categories used	0.36	12.11
Number of savings accounts	−0.37	−11.62
Intercept	−2.50	−19.56
Log-Likelihood = −5393.06, Pseudo R^2 = .14, N = 86,754		

Different algorithms can estimate the counterfactual outcomes for the different matching approaches. For this study, these are the one-nearest neighbor algorithm, four-nearest neighbors algorithm (Abadie, Drukker, Herr Leber, & Imbens, 2001), and a Gaussian kernel algorithm. This study uses a procedure with replacement because matching with replacement can improve the average quality of matching and minimize bias (Caliendo & Kopeinig, 2008). Also, a common support restriction for propensity score matching and hybrid matching helps ensure appropriate matching partners.

The IV method requires an exclusion restriction; that is, at least one instrumental variable that affects the use of online banking but not the outcome variables. Previous research shows that age (AGE) and the length of the relationship (LOR) are important drivers of customers' online use (e.g., Hitt & Frei, 2002; Shankar et al., 2003). Thus AGE and LOR (separately) serve as instrumental variables for online banking use with the recognition that both characteristics correlate substantially with the outcome variables (e.g., balance checking account). The outcome equations then become:

$$\begin{aligned} y_{h,k} &= \beta_{0,k} + \beta_{d,k}\cdot\hat{p}(z_h) + \beta_{1,k}\cdot\delta\cdot AGE_h + \beta_{2,k}\cdot(1-\delta)\cdot LOR_h \\ &\quad + \beta_{3,k}\cdot JOINT_h + \beta_{4,k}\cdot NOP_h + \beta_{5,k}\cdot NOSAV_h + \varepsilon_{h,k} \\ \text{with}\,\delta &= \begin{cases} 1, \textit{if AGE is instrumental}\ \text{variable}, \\ 0, \textit{otherwise}. \end{cases} \qquad \forall h \in (I \cup J), k \in K, \end{aligned} \tag{12}$$

where $\hat{p}(z_h)$ is the propensity score.

The control function method does not require an exclusion restriction. The models therefore exclude an instrumental variable [Eq. (13)] or include AGE or LOR as the instrumental variable [Eq. (14)], respectively:

$$\begin{aligned} y_{h,k} &= \beta_{0,k} + \beta_{d,k}\cdot d_h + \beta_{1,k}\cdot AGE_h + \beta_{2,k}\cdot LOR_h + \beta_{3,k}\cdot JOINT_h \\ &\quad + \beta_{4,k}\cdot NOP_h + \beta_{5,k}\cdot NOSAV_h + \frac{Cov(\varepsilon_k,\eta_k)}{Var(\eta_k)}\cdot\frac{\phi(z_h\cdot\gamma)}{\Phi(z_h\cdot\gamma)}\cdot d_h \\ &\quad + \frac{Cov(\varepsilon_k,\eta_k)}{Var(\eta_k)}\cdot\frac{-\phi(z_h\cdot\gamma)}{1-\Phi(z_h\cdot\gamma)}\cdot(1-d_h) + \varepsilon^*_{h,k}\forall h \in (I \cup J), k \in K, \end{aligned} \tag{13}$$

$$\text{with}\ z_h = (AGE_h, LOR_h, JOINT_h, NOSAV_h, NOP_h)$$

$$\begin{aligned} y_{h,k} &= \beta_{0,k} + \beta_{d,k}\cdot d_h + \beta_{1,k}\cdot\delta\cdot AGE_h + \beta_{2,k}\cdot(1-\delta)\cdot LOR_h + \beta_{3,k}\cdot JOINT_h \\ &\quad + \beta_{4,k}\cdot NOP_h + \beta_{5,k}\cdot NOSAV_h + \frac{Cov(\varepsilon_k,\eta_k)}{Var(\eta_k)}\cdot\frac{\phi(z_h\cdot\gamma)}{\Phi(z_h\cdot\gamma)}\cdot d_h \\ &\quad + \frac{Cov(\varepsilon_k,\eta_k)}{Var(\eta_k)}\cdot\frac{-\phi(z_h\cdot\gamma)}{1-\Phi(z_h\cdot\gamma)}\cdot(1-d_h) + \varepsilon^*_{h,k} \\ \text{with}\ \ \delta &= \begin{cases} 1, \textit{if AGE is instrumental}\,\text{variable}, \\ 0, \textit{otherwise}. \end{cases} \qquad \forall h \in (I \cup J), k \in K. \end{aligned} \tag{14}$$

Estimates from the OLS provide a benchmark for the IV and control function estimates, because biases in OLS estimates due to self-selection effects can be less severe than biases in the IV and control function estimates when weak instrumental variables exist (Staiger & Stock, 1997; Woglom, 2001).

4.4. Assessing the predictive validity of different methods

The test of the predictive validity of the different methods relies on information about the 1,628 online customers for an earlier three-month period (October–December 2002). In this period, 108 of the 1,628 online customers are offline, so they start using the online channel after December 2002. The absolute percentage error (APE) discloses the predictive validity of the different methods:

$$APE_k = \frac{\left|E_i\left(y^1_{i,k,2003}\middle|d_i = 1\right) - \left[E_i\left(y^0_{i,k,2002}\middle|d_i = 1\right) + ATTE_k\right]\right|}{E_i\left(y^1_{i,k,2003}\middle|d_i = 1\right)} \qquad \forall k \in K. \tag{15}$$

The critical difference is between the actual $(E_i(y^1_{i,\ k,\ 2003}|d_i = 1))$ and the estimated $(E_i(y^0_{i,k,\ 2002}|d_i = 1) + ATTE_k)$ values of the outcome variables for customers who go online after December 2002. No effects other than using online banking should play a major role (e.g., increase in checking account balance due to increase in income), because the time span for the difference is less than one year. The low product usage rate limits the computation of APE$_k$ to the checking account balances, the number of checking accounts, and the number of transactions.

5. Empirical study results

5.1. Predictive validity of different methods

The first assessment pertains to which matching method and matching algorithm lead to the highest predictive validity where the APE$_k$ is the dependent variable in an OLS and the matching method (hybrid, propensity score, covariate matching), and the matching algorithm (one-nearest neighbor, 4-nearest neighbors, Gaussian kernel algorithm) are the independent variables. The three different outcome variables serve as controls. The second step involves a comparison of the predictive validity of the best performing matching method with the predictive validity of the IV, control function, and OLS methods.

Hybrid matching tends to decrease the APE (β = −0.03), though the effect is not significant (t = −1.01). See Table 3. Using more than one matching partner generally increases the predictive validity, and the Gaussian kernel algorithm improves the predictive validity significantly (β = −0.05, t = −3.05). Hence, many offline customers are similar to an online customer; otherwise, the performance of the four-nearest neighbors and one-nearest neighbor algorithms would be comparable. Overall, hybrid matching in combination with the Gaussian kernel algorithm increases predictive validity in this empirical study.

The next comparison involves the predictive validity of the hybrid matching method with the IV, control function, and OLS methods. The hybrid matching method in combination with the Gaussian kernel algorithm always leads to the lowest APE for all three outcome

Table 3
Effect of matching method and matching algorithm on predictive validity of estimated treatment effects (ATTE).

Independent variables	Coefficient	Standardized coefficient	t value
Matching method			
Hybrid matching	−0.03	−0.13	−1.01
Propensity score matching	−0.01	−0.10	−0.80
Covariate matching[a]	0.00		
Matching algorithm			
Gaussian kernel algorithm	−0.05	−0.39	−3.05
4-nearest neighbors	−0.04	−0.32	−2.55
1-nearest-neighbor [a]	0.00		
Control variables			
Number of transactions	0.11	0.87	6.87
Checking account balances	0.03	0.23	1.81
Number of checking accounts [a]	0.00		
Constant	0.08		4.69
N = 27, df = 20, R^2 = 0.76, F value = 10.41 (p value = 0.00)			

Validity (i.e., dependent variable) measured by absolute percentage error (APE) based on hold-out sample.

[a] Reference category (dummy coding).

Table 4
Predictive validity of different methods based on estimated treatment effect (ATTE).

	Absolute percentage error (%)			
Method	Checking account balance	Number of checking accounts	Number of transactions	Mean APE
Hybrid matching method with Gaussian kernel	3.48	2.01	30.63	12.04
OLS	13.05	3.72	37.00	17.92
Control function method (instrument: AGE)	33.83	165.92	333.68	177.81
Control function method (instrument: LOR)	245.02	109.42	303.75	219.40
Control function method (no instrument)	282.39	156.87	303.75	247.67
Instrumental variable method (instrument: AGE)	295.71	439.06	804.75	513.18
Instrumental variable method (instrument: LOR)	444.34	266.24	582.84	413.14

variables: 3.48% for checking account balance, 2.01% for number of checking accounts, and 30.63% for the number of transactions (Table 4). This approach also leads to a substantially lower mean APE across the three outcome variables (MAPE = 12.04%) compared with all other methods. However, the performance of the OLS method is surprisingly good (MAPE = 17.92%). The IV and control function methods do not perform well, probably because of weak instrumental variables, the estimated effects are heavily biased. When age is the instrumental variable, the mean APE for the control function and IV method are 177.81% and 513.18% respectively. If the length of the relationship is the instrumental variable, the mean APE values for the control function and IV method are equal to 219.40% and 413.14% respectively. Thus, the use of the IV and control function methods, when no strong and valid instrumental variables appear, is not advisable (Table 4).

5.2. Size of self-selection effects

To determine the importance of controlling for self-selection effects, this study next uses hybrid matching in combination with the Gaussian kernel algorithm and Eq. (3). Table 5 discloses the results of the corresponding decomposition of the mean difference between online and offline customers into the treatment (ATTE) and self-selection (SE) effect for all outcome variables and reports the relative size of the different effects.

Substantial self-selection effects emerge for almost all outcome variables. For example, for the savings account balances, 71.5% of the observed mean difference is due to self-selection. The size of the self-selection effect for the number of checking accounts (+72.2%) is similar. The observed mean difference for the number of credit cards is only due to self-selection effects (+100.0%) for this study. Assessing the effect of online use on product usage without controlling for self-selection effects causes the effect to be biased upward for these outcome variables. For example, retail banks earn money by charging fees for checking accounts. A comparison of online and offline customers indicates that online customers own more checking accounts than offline customers. Thus, managers might conclude that stimulating customer channel migration to the online channel is profitable because it increases the number of checking accounts. However, most of this difference reflects self-selection; the customers who have multiple checking accounts are more likely to use online banking, probably because they take banking more seriously.

For checking account balances and the number of brokerage accounts, the self-selection effect is greater than 100%, which implies that using online banking and self-selection have contrary influences on the outcome variables. For both outcome variables, the estimated effect of using online banking is positive, whereas the self-selection effect is negative; online customers in general have less money at their disposal.

Overall, this study shows that substantial self-selection effects exist, which confirms the importance of controlling for these effects. A failure to control for self-selection effects can lead to biased estimates of the influence of online banking on product usage, and ultimately to inappropriate managerial decisions.

6. Conclusion and limitations

This study outlines several methods to control for self-selection effects in situations characterized by cross-sectional transaction data with a limited number of customer characteristics: five in this study. The results differ across methods. Among the various matching methods, hybrid matching in combination with a Gaussian kernel algorithm offers the highest predictive validity. However, covariate matching and propensity score matching also perform well. The hybrid matching method has a much higher predictive validity than the IV and control function methods. The results further indicate that IV and control function methods are not appropriate for observations with only weak instrumental variables that are not exogenous or highly correlated with the error term. This scenario is likely in settings that offer cross-sectional transaction data with limited additional information about customers (e.g., no attitudinal data). Applying hybrid matching in such settings is feasible to control for self-selection effects. Researchers and managers interested in the effects of certain treatments, for example online use or use of loyalty programs, can therefore rely on hybrid matching to control for self-selection.

This research further demonstrates how to decompose the difference in product usage between online and offline customers into self-selection and treatment effects. The empirical results show that self-selection effects can be larger than treatment effects and that not taking self-selection effects into account can lead to wrong managerial decisions. For example, online customers have, on average, more credit cards than offline customers. Managers might conclude that it is profitable to move offline customers to the online channel to improve credit card usage. Yet, this conclusion is only correct if

Table 5
Relative effect sizes based on the hybrid matching method with Gaussian kernel algorithm: effect of using online banking and the self-selection effect (*p* value in brackets).

Outcome variable	Number of online customers	Observed mean difference [online–offline customers]	Relative size of effect of online banking (ATTE) [in %]	Relative size of self-selection effect (SE) [in %]
Checking account balances	1,614	−201.38 (0.02)	−9.2 (0.58)	+109.2
Savings account balances	414	−883.33 (0.00)	+28.5 (0.06)	+71.5
Brokerage account turnovers	39	−125.74 (0.63)	+89.2 (0.09)	+10.8
Credit card turnovers	57	−50.53 (0.25)	+123.4 (0.38)	−23.4
Number of checking accounts	1,628	0.18 (0.00)	+27.8 (0.00)	+72.2
Number of brokerage accounts	1,628	−0.02 (0.00)	−50.0 (0.16)	+150.0
Number of credit cards	1,628	0.02 (0.00)	+0.0 (0.26)	+100.0
Number of transactions	1,628	3.68 (0.00)	+72.8 (0.00)	+27.2

Note: Calculation of relative effect size from Eq. (3): [ATTE/observed mean difference] and [SE/observed mean difference].

using the online channel results in an increased use of credit cards, which requires a positive treatment effect. This conclusion is wrong if customers who use credit cards intensively decide to use the online channel; but the online channel itself has no effect on the usage of credit cards, as is the case in this study. In this situation, the treatment effect is zero and the self-selection of customers is responsible for the entire difference in the number of credit cards. Therefore, motivating offline customers to use the online channel would not yield an increase in credit card usage.

While investigating the effects of channel usage in a multi-channel setting is highly relevant (e.g., Gensler, Dekimpe, & Skiera, 2007; Noble, Griffith, & Weinberger, 2005), the methodological discussion is not restricted to this setting. Managers often face potential self-selection effects, such as in assessing the effectiveness of loyalty programs or social media activities (Hennig-Thurau et al., 2010). Thus, this study helps extend the literature by advancing state-of-the-art knowledge about how to control for self-selection effects in situations with a limited number of customer characteristics and underlines the vast potential of matching methods.

Yet, this study suffers from some limitations that offer opportunities for further research. The empirical dataset provides only limited information about the individual customer. Additional customer characteristics also could determine online channel use, including attitude toward the Internet. Further research could increase the set of customer characteristics to control for self-selection effects. This study further observes customer behavior within only one firm and thus cannot consider competitive behavior. Researchers should collect more information about customers across all firm relationships to shed more light on customer behavior. Because this study only uses cross-sectional data, the long-term effects of online use cannot be examined. Future research might also want to investigate such effects. Finally, this study does not investigate whether product characteristics moderate the effect of online use. Further research could examine how differences in the effects of online use on product usage can be explained by product characteristics.

References

Abadie, A., Drukker, D., Herr Leber, J., & Imbens, G. (2001). Implementing matching estimators for average treatment effects in Stata. *The Stata Journal*, *1*(1), 1–18.

Amemiya, T. (1985). *Advanced econometrics.* Boston: Harvard University Press.

Blundell, R., Dearden, L., & Sianesi, B. (2005). Evaluating the effect of education on earnings: Models, methods and results from the National Child Development Survey. *Journal of Research in Statistics & Sociology A*, *168*(3), 473–512.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Survey*, *22*(1), 31–72.

Campbell, D., & Frei, F. X. (2010). Cost structure, customer profitability, and retention implications of self-service distribution channels: Evidence from customer behavior in an online banking channel. *Management Science*, *56*(1), 4–24.

Cortiñas, M., Chocarro, R., & Villanueva, M. L. (2010). Understanding multi-channel banking customers. *Journal of Business Research*, *63*(11), 1215–1221.

Danaher, P., Wilson, I., & Davis, R. (2003). A comparison of online and offline consumer brand loyalty. *Marketing Science*, *22*(4), 461–476.

Degeratu, A., Rangaswamy, A., & Wu, J. (2000). Consumer choice behavior in online and traditional supermarkets: The effects of brand name, price and other search attributes. *International Journal of Research in Marketing*, *17*(1), 55–78.

Dehejia, R., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review Economic Statistics*, *84*(1), 151–161.

Gensler, S., Dekimpe, M., & Skiera, B. (2007). Evaluating channel performance in multi-channel environments. *Journal of Retailing and Consumer Services*, *14*(1), 17–23.

Hadi, A. (1994). A modification of a method for the detection of outliers in multivariate samples. *Journal of Research in Statistics & Sociology B*, *56*(2), 393–396.

Heckman, J., & Navarro-Lozano, S. (2004). Using matching, instrumental variables, and control functions to estimate economic choice models. *Review of Economic Statistics*, *86*(1), 30–57.

Heckman, J., & Robb, R. (1985). Alternative methods for evaluating the impact of interventions. *Journal of Econometrics*, *30*(1/2), 239–267.

Hennig-Thurau, T., Malthouse, E., Friege, C., Gensler, S., Lobschat, L., Rangaswamy, A., et al. (2010). The impact of new media on customer relationships: From bowling to pinball. *Journal of Service Research*, *13*(3), 311–330.

Hitt, L. M., & Frei, F. X. (2002). Do better customers utilize electronic distribution channels? The case of PC banking. *Management Science*, *48*(6), 732–748.

Leenheer, J., Van Heerde, H., Bijmolt, T., & Smidts, A. (2007). Do loyalty programs really enhance behavioral loyalty? An empirical analysis accounting for self-selecting members. *International Journal of Research in Marketing*, *24*(1), 31–47.

Mithas, S., Krishnan, M. S., & Fornell, C. (2005). Why do customer relationship management applications affect customer satisfaction? *Journal of Marketing*, *69*(4), 201–209.

Noble, S., Griffith, D., & Weinberger, M. (2005). Consumer derived utilitarian value and channel utilization in a multi-channel retail context. *Journal of Business Research*, *58*(12), 1643–1651.

Puhani, P. (2000). The Heckman correction for sample selection and its critique. *Journal of Economic Survey*, *14*(1), 53–68.

Rosenbaum, P., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55.

Rosenbaum, P., & Rubin, D. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *American Statistics*, *39*(1), 33–38.

Rubin, D., & Thomas, N. (1996). Matching using estimated propensity scores: Relating theory to practice. *Biometrics*, *52*(1), 249–264.

Shankar, V., Smith, A. K., & Rangaswamy, A. (2003). Customer satisfaction and loyalty in online and offline environments. *International Journal of Research in Marketing*, *20*(2), 153–175.

Smith, J., & Todd, P. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics*, *125*(1/2), 305–353.

Staiger, D., & Stock, J. (1997). Instrumental variables regression with weak instruments. *Econometrica*, *65*(3), 557–586.

Stock, J., Wright, J., & Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business Economic Statistics*, *20*(4), 518–529.

Tedeschi, B. (2005). (available on their web sites to their ATM's. New York Times, URL). *E-commerce report—To attract more internet customers, some banks are adding services* http://query.nytimes.com/gst/fullpage.html?res=9806E5DD1E3DF934A35750C0A9639C8B63&sec=&spon=&pagewanted=all (last update: September 11, 2011)

Vella, F., & Verbeek, M. (1999). Estimating and interpreting models with endogenous treatment effects. *Journal of Business Economic Statistics*, *17*(4), 473–478.

Von Wangenheim, F., & Bayon, T. (2007). Behavioral consequences of overbooking service capacity. *Journal of Marketing*, *71*(4), 36–47.

Woglom, G. (2001). More results on the exact small sample properties of the instrumental variable estimator. *Econometrica*, *69*(5), 1381–1389.

Zhao, Z. (2004). Using matching to estimate treatment effects: Data requirements, matching metrics, and Monte Carlo evidence. *Review of Economic Statistics*, *86*(1), 91–107.