Propensity Score Matching and Its Application to Risk Drivers Detection in Financial Setting

M. Karwański and U. Grzybowska*

Department of Informatics, Faculty of Applied Informatics and Mathematics, Warsaw University of Life Science,
Nowoursynowska 159, PL-02-776 Warsaw, Poland

In credit risk scoring models are used as a tool to evaluate the level of risk associated with applicants or customers. The aim of these models is not only to estimate the probability that the client will not be able to fulfill his financial commitments but also to identify and estimate the risk drivers i.e., client attributes that are responsible for risk occurrence. Unfortunately, scoring models are built based on historic data stored by bank over the clients. Selection of clients is not random. This leads to systematic errors. Therefore one seeks methods that allow for a model correction that enables application of statistical inference. Quasi-experimental designs are practical solutions to this dilemma. One of such methods is propensity score matching. Propensity score matching allows also for detecting risk drivers that are independent of borrowers attributes, e.g., triggered by various bank strategies. The aim of our research is to apply propensity score matching methodology to identify these risk drivers in credit risk that could not be detected e.g., by regression models.

1. Introduction

In 2011 in the paper "Risky Borrowers or Risky Mortgages: Disaggregating Effects Using Propensity Score Models", the authors applied propensity score matching (PSM) and found out that the default risk was about 70% lower with Community Reinvestment Act motivated loans than with a subprime mortgage [1]. Application of PSM revealed that mortgage default risk could not be attributed solely to borrower credit risk and the high default risk was strongly associated with the characteristics of the loan product. The PSM methodology was developed by Rosenbaum and Rubin in 1983 [2, 3]. Since then it has gained much attention and has been applied in a wide variety of disciplines for casual modeling. The PSM was designed to estimate casual effects in situations when it is impossible to design experiments or in other words it is not possible to implement random assignment due to ethical reasons, the cost or nature of the problem. Applications of PSM to numerous areas are reflected in rich literature in medicine, epidemiology, social sciences, etc. The method has been used to evaluate medical treatment effects, impact of educational programs, influence on labor market of various activity actions [4–6]. However, to our knowledge, there are only very few examples of PSM application as a tool for detecting risk drivers in credit risk, see e.g. [7]. One of the reasons is accessibility of data. The aim of our research is to apply PSM methodology to identify risk drivers in credit risk. Namely, following [1] we will apply PSM to show that certain bank strategies have causal effect on default rate. We will

present results of our research conducted on real life data. The data concerns customers signing credit agreements to finance the purchase of a car dedicated to small and medium enterprises (SME). We will conduct calculations with help of logistic regression models and data mining approach involving families of classifiers: random forest and gradient boosting.

The paper is organized as follows. In Sect. 2 we briefly describe PSM methodology. Section 3 is devoted to data description. In Sect. 4 we present a detailed description of the procedure applied to the data. We also present and discuss obtained results. In the last section we will draw conclusions. The calculations were done in SAS Enterprise Miner ver. 13.2.

2. Propensity score matching

In situations when one is interested in measuring the effects of an action e.g., some treatment on individuals we observe the outcome of those who received the treatment and those who did not receive the treatment. To be able to estimate the effect of the action on a participating individual we must compare the observable outcome with the outcome that would have resulted had the individual not been subject to the action. However, only one outcome is observed for one individual [8]. Moreover, due to the selection procedure the individuals in both groups may differ. The key idea to overcome differences between the treated and untreated groups was to match similar individuals representing both groups based on their characteristics. As it is impossible to match based on separate attributes characterizing individuals, each participant and nonparticipant is assigned a score, a propensity score, i.e., the probability of being treated. Implementing PSM requires the following steps [8]:

———————
*corresponding author; e-mail: urszula_grzybowska@sggw.pl

1. Calculation of a propensity score using e.g., logistic regression.

2. Choice of an appropriate algorithm to perform matching using propensity score.

3. Estimation of the average causal effect by comparing matched groups on the outcome.

The steps of PSM are shown in Fig. 1. The propensity score can be calculated using a binary choice model (logit or probit) or families of classifiers. Next, provided common support of propensity scores in both groups exists, matching procedure is performed. For each treated object within the common support its nonparticipating counterpart with similar propensity score is selected as shown in Fig. 2.
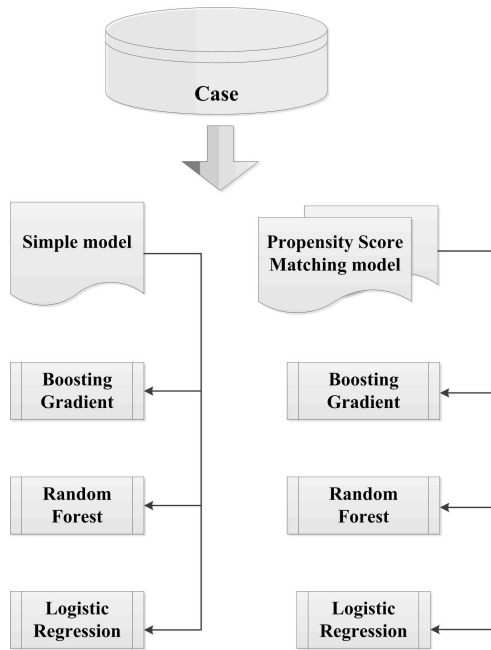

Fig. 1. Diagram representing steps of PSM.

Finally, average treatment effect is calculated e.g., by calculating the difference of outcomes in matched groups.

3. Data

Our study concerned car loans, more precisely car lease for individual clients. The data included car lease agreements signed in the years 2008–2010. Deals with at least 60-day delay in payment in the first year of the contract, were regarded as defaulted. Due to some business procedure some of the clients were selected to receive easy credit treatment. The others were given standard treatment. The assignment to both groups has not been random and it was based on a new just introduced classification rule. The question posed by business was whether bank procedure could influence the default risk. In our application of PSM the bank procedure easy credit or
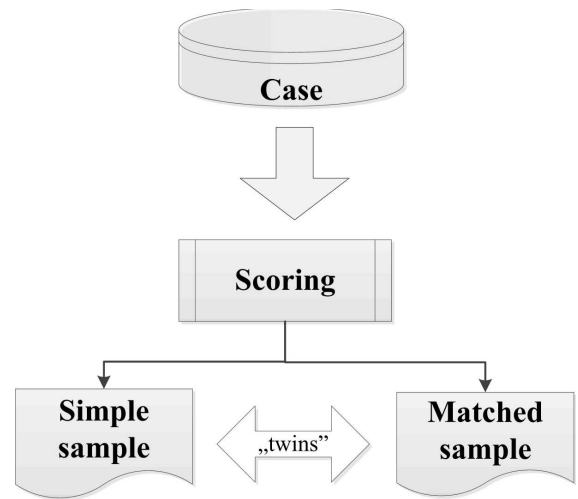

Fig. 2. Propensity score matching.

standard is the counterpart of a treatment or action while the default is the effect of interest. In both groups (see Table I) the default rate was different and it was equal to 3.08% and 9.14%, respectively. The odds ratio is 3.17 which means that it is over three times more probable that a client under strategy standard will default than an easy credit client [9]. It is however not obvious whether that depends on differences in clients selected to both groups or can be attributed rather to bank procedure.

TABLE I

Frequencies of applied strategies.

defaulted		percentiles of default		
		strategy		
		Easy Credit	standard	total
no=0	frequency	1134	159	1293
	row percent	87.7	12.3	
	column percent	96.92	90.86	
yes=1	frequency	36	16	52
	row percent	69.23	30.77	
	column percent	3.08	9.14	
	total	1170	175	1345

TABLE II

Advance price of the car.

price	frequency	percent	cum. frequency	cum. percent
0–33	427	28.68	427	28.68
34–36	382	25.65	809	54.33
37–51	680	45.67	1489	100.00

The aim of the research was to examine whether bank strategy can influence the default rate so it can be regarded as a risk driver. Clients were described by attributes collected in their application forms, among others information about their age, occupation, martial status. The data was analyzed and cleaned. As a result

TABLE III

Martial status.

status	frequency	percent	cum.frequency	cum. percent
couple	1162	78.04	1162	78.04
single	327	21.96	1489	100.00

some columns were removed. The data observation window was prepared in such a way that for the remaining observations it was possible to observe lease for at least 10 months. Apart from variables sourced from lease application forms, three behavioral factors were included: the number of previous agreements of the client, the number of settled agreements, the value of previous commitments. All variables were categorized using established bank practice. Exemplary attributes are shown in Table II and Table III.

The final data set used in calculations contained 1489 observations, among them 57 (3.83%) were identified as default. In the first step the selection of variables was performed using best subset method that compares all possible models [10]. As a result 12 variables were selected for PSM application. The variables influence both the probability of default and the membership into strategy groups. Selected variables are shown in Table IV.

TABLE IV

Set of final variables used in the analysis.

strategy identifier
defaulted
advance price of car
age of business
age of owner
age of vehicle
district
ownership form
industry sector
length of contract
number of correctly closed previous contracts
number of previous contracts of the consumer
type
value of objects in previous contracts

4. Calculations and results

We have first calculated propensity scores i.e., the probabilities of being assigned to one of two strategy groups based on selected variables for all clients using logistic regression, gradient boosting and random forests (compare [11]).

The distributions of propensity scores in both groups overlap as shown in Fig. 3 but they are different. Afterward, for all clients in the treated group (strategy=1, easy credit) their non-treated counterparts (strategy=0, standard) with similar propensity score were found using

logistic regression, gradient boosting and random forests. Matching was performed using nearest neighbors method with $n = 4$ and caliper= 0.1. As a result two matched groups were distinguished.

The distribution of propensity scores in both groups are similar as shown in Fig. 4 for the case of random forest approach. The Kolmogorov–Smirnov distance measure D between the distributions is $D = 0.4946$ for raw data and $D = 0.2252$ for matched groups. We can presume, following the main idea of PSM, that in selected groups the strategy does not depend on the client's attributes. The clients in both groups are "similar" and the outcome (the default rate) in both selected groups does not depend on clients' attributes. The difference in default rate in selected groups can be attributed to the strategy and not to clients' characteristic.
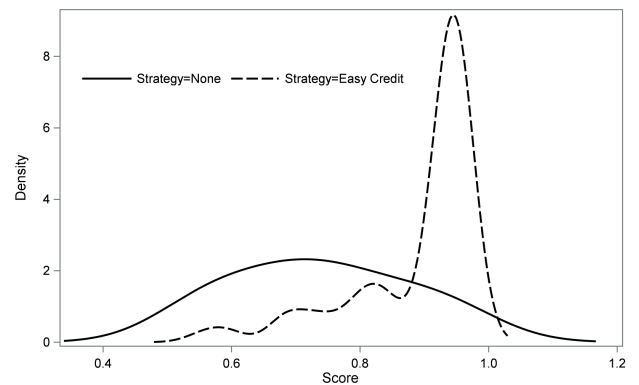

Fig. 3. Distribution of propensity scores in both groups before matching.
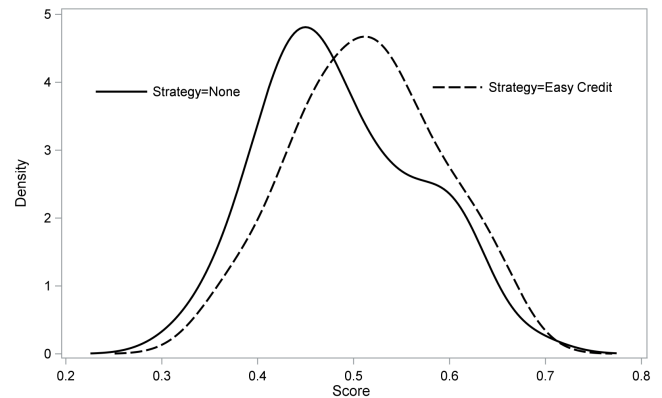

Fig. 4. Propensity score distribution in matched groups with scores obtained by Random Forests.

We compare the outcome (default rate) in both matched groups: easy credit and standard. The ratios are 4.89% in the first group and 7.45% in the second (see Table V). The ratios are different but the difference is much smaller than in initial groups. The odds ratio is 1.56 and it means that it is 1.56 times more probable

that a standard strategy client will default than an easy credit client with the same propensity score.

TABLE V

Number of defaulted by strategy in matched groups.

defaulted		percentiles of default		
		strategy		
		Easy Credit	standard	total
no=0	frequency	447	436	882
	row percent	50.68	49.32	
	column percent	95.11	92.55	
yes=1	frequency	23	35	58
	row percent	39.66	60.34	
	column percent	4.89	7.45	
	total	470	470	940

TABLE VI

ROC analysis for RDM and PSM model.

model	ROC		95% Wald C.L.		correlations		
	AUC	std. err.	lower	upper	som.'d	gamma	tau
RDM	0.8019	0.0506	0.7027	0.9012	0.6039	0.6089	0.0361
PSM	0.7323	0.0618	0.6113	0.8533	0.4646	0.4684	0.0270

ROC — receiver operating characteristic, AUC — area under the ROC curve, Wald C.L. — Wald confidence limits, Som.'D — Somers'D.

TABLE VII

Odds ratio of different PSM models versus RDM.

probability of default	RDM			PSM		
	OR	LCL	UCL	OR	LCL	U. CL
random forests						
strategy 0/1	1.6911	0.8214	3.4816	1.868	1.036	3.366
gradient boosting						
strategy 0/1	1.6911	0.8214	3.4816	1.34	0.70	2.55
logistic regression						
strategy 0/1	1.6911	0.8214	3.4816	1.999	1.094	3.654

LCL — lower confidence limit, UCL — upper confidence limit.

To compare the effect of the strategy on clients we have built a scoring model on the set of matched groups. The scoring model was built using random forests. The PSM model was then compared with a model built on the whole initial set of clients (raw data). The results presented in Table VI contain comparison of model obtained on raw data (RDM) and PSM random forest model. The obtained statistics indicate that both models have strong discrimination power.

The AUC measures the quality of the model. In both cases, raw data model and PSM model, it is relatively high. Because the target variable is binary, we use the odds ratio to compare the differences in outcomes i.e., to investigate the influence of strategy on the probability of default. Table VII presents results for PSM obtained with help of random forest, gradient boosting and logistic regression compared to the raw data model together with

relevant confidence limits for calculated odds ratios. All results confirm that the default rate decreased for the strategy easy credit or in other words for the strategy standard default risk for borrowers with the same attributes is much higher. The odds ratio for raw data is 1.6911 meaning that the risk of default is 1.6911 times higher for strategy standard than it is for easy credit. After the influence of attributes has been removed and PSM was performed with random forests, the odds ratio was 1.8678 meaning that the risk of default increased by almost 10.4%. Similar conclusion can be drawn for logistic regression. On the other hand, the odds ratio estimated with gradient boosting is lower than for raw data model. It also indicates that strategy easy credit influences the risk of default but the influence is not so strong as it is shown by random forests and logistic regression.

5. Concluding remarks

In order to adopt correctly effective policy of intervention it is very important to use proper causal evidence based on true quantity values. In observational studies the effect of treatment is often disturbed by systematic differences in subjects under treatment. PSM is one of low cost estimation methodologies that can be utilized to overcome these difficulties and obtain the real effect of treatment. The central problem in PSM is calculating balancing scores. In order to obtain balancing scores the common practice is the usage of logistic regression. Algorithmic approach involving random forests and gradient boosting appears here very promising, yet not fully recognized. The calculations presented in the article conducted with random forests, gradient boosting and logistic regression revealed that bank strategy easy credit decreases the risk of default. It was also observed that obtained results depend significantly on the applied method. The results obtained with gradient boosting, a method regarded as one of the best machine learning methods, differ from these obtained with logistic regression and random forests as shown in Fig. 5.
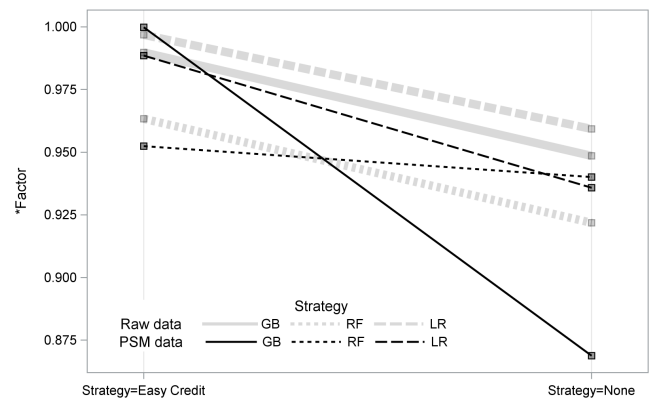

Fig. 5. Odds ratios for all applied methods.

Logistic regression and random forests give similar results. All methods indicate that the risk of default is

higher with standard strategy and this cannot be attributed to client's characteristic alone. These findings have important implications for business and may allow to control the financial aspects, both losses due to default and the costs of actually wrong strategy.

References

[1] L. Ding, R. Quercia, J. Ratcliffe, *J. Real Estate Res.* **33**, 245 (2011).

[2] P.R.Rosenbaum, D.B. Rubin, *Biometrika* **70**, 41 (1983).

[3] P.R. Rosenbaum, D.B. Rubin, *J. Am. Stat. Assoc.* **79**, 516 (1983).

[4] A. Bryson, R. Dorsett, S. Purdon, *The use of propensity score matching in the evaluation of active labour market policies*, 2002.

[5] R. Trzciński, *The Use of Propensity Score Matching Technique in Evaluation Research*, Polska Agencja Rozwoju Przedsiębiorczości, Warszawa 2009 (in Polish).

[6] M. Li, *Org. Res. Meth.* **39**, 1 (2012).

[7] A. Saunders, S. Steffen, *Rev. Fin. Stud.* **24**, 4091 (2011).

[8] C. Heinrich, A. Maffioll, G. Vazquez, *A Primer for Applying Propensity Score Matching Impact-Evaluation Guidelines*, Technical Notes No. IDB-TN-161 2010.

[9] D. Larose, *Data Mining Methods and Models*, Wiley, New Jersey 2006.

[10] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning. Data Mining, Inference and Prediction*, 2nd ed., Springer, New York 2009.

[11] J. Luellen, W. Shadish, M. Clark, *Evaluat. Rev.* **29**, 530 (2005).