European Economic Letters
ISSN 2323-5233
Vol 15, Issue 1 (2025)
http://eelet.org.uk

A Review of Propensity Score Matching Techniques: Estimation Of Income Effect with an Illustration Using Lalonde Dataset

Mrs. Nancy Purohit,

Research Scholar, Department of Business Administration, Sambalpur University, Jyoti Vihar, Burla, Odisha

Dr. Tushar Kanti Das,

Professor, Department of Business Administration, Samblapur University, Jyoti Vihar, Burla, Odisha

Dr. Kunal Mishra,

Assistant Professor, Department of Commerce, Rajendra University, Prajna Vihar, Balangir, Odisha

Abstract

For evaluating the success of social and economic programmes, impact evaluation of such is necessary. In randomization (Experimental studies) results of the estimates are free from bias, but randomization is not possible in every cases then we access the result of such programmes through observational study. Selection bias frequently makes it more difficult to estimate causal effects in observational studies because treatment assignment is not random. One of the most popular quasi-experimental methods to deal with this problem is Propensity Score Matching (PSM), which creates similar treatment and control groups based on observed characteristics through similar p-score. The conceptual framework of PSM, its underlying assumptions, matching algorithms, and estimation techniques are all covered in this paper. The estimation of treatment effects on income is explained with an example that makes use of fictitious calculations and dataset structure. This paper presents a methodological review of PSM and uses the Lalonde Training Program Dataset to illustrate its use. The advantages and limitation of PSM in empirical research are also discussed in the study. The results explains that PSM is a powerful and useful tool for assessing program impacts in situations where randomized experiments are impractical, though its efficacy is dependent on the appropriate model specification and availability of strong covariates.

Keywords: Observational studies, Propensity Score Matching, Selection Bias, Income Effect, Lalonde Dataset , Impact Evaluation.

Introduction

The main concern in public policy research and applied economics is the evaluation of effectiveness of economic and social welfare programs that aims at reducing poverty, enhancing employment opportunities and improving the welfare of household. Accessing whether these programs actually achieve their intended outcomes requires rigorous impact evaluation techniques.

In randomized experimental trials researcher often evaluate the casual effect of welfare programs by randomly assigning participants to control and treatment groups However, in case of real-world scenario, random assignment is not possible due to logistical, ethical or financial problems. As a result of which, researchers rely on observational data to evaluate program impacts. A key challenge in observational studies is *selection bias*, which occurs when individuals self select into program participation for which who participate in a program differ systematically from those who do not.

To address such an issue , several quasi-experimental techniques have been developed. Among these, Propensity Score Matching has become one of the most widely applied methods. PSM attempts to mimic randomized experiments by matching treated individuals with non-treated individuals who share similar observable characteristics.

Donald Rubin and Paul Rosenbaum made a contribution to the propensity score methodology in 1983. They showed that bias resulting from observable factors can be eliminated by matching on the probability of treatment conditional on observed covariates.
The Lalonde Training Program Dataset, a well-known dataset frequently utilized in the assessment of job training programs, is used in this paper to illustrate the application of PSM and give a methodological overview of it. The goal is to describe PSM's conceptual framework, implementation procedures, underlying presumptions, and problems in the context of income impact assessment.

.

Propensity Score Matching

Random experiment is not always feasible, in that case Propensity Score Matching is used to evaluate the causal effect of a treatment where control groups are framed statistically. It is a non experimental impact evaluation method where it uses the information from a pool of units that don't participate in the program to identify what would have happened to the participants in the absence of the intervention. This method pairs people with similar traits who have received treatment and those who have not, building a counterfactual outcome, is the fundamental concept of PSM. PSM evaluates this counterfactual by choosing comparable members of the control group. PSM simplifies the issue to a single index known as the propensity score, which is the likelihood of receiving treatment given observed covariates, rather than matching people on several characteristics at once. The technique guarantees that the distribution of observed characteristics is balanced between treatment and control groups by matching individuals with similar propensity scores by removing the problem of selection bias in observational studies.

PSM – Mathematical Framework

The propensity score (P-Score) is the estimated probability of being in the treatment group given the observable characteristics from a regression model (Logit / Probit) of participation,(Rosenbaum and Rubin 1983) . The P- Score is a conditional probability of receiving the treatment given the observed characteristics ($p(x_i) = P(D_i = 1 | x = x_i)$), where the outcome variable is binary or dichotomous, so the potential outcomes :

$$y_i = \begin{cases} y_{1i} & if\ D_i = 1 \\ y_{0i} & if\ D_i = 0 \end{cases}$$

where, y_i = Outcome of interest

D_i = Treatment variable D_i indicates whether individual i is treated

(1 = treated, 0 = Control)

P = Conditional Probability

p(xi) = Propensity score for individual *i*

xi = Observed characteristics (covariates)

The main parameter of interest in many program evaluations is the Average Treatment Effect on the Treated (ATT).

Treatment Effects:

- Individual level treatment effects: $(y_{1i} - y_{0i})$
- Average Treatment Effect (ATE):

$$\text{ATE} \equiv E(y_{1i} - y_{0i})$$

- Average Treatment Effect on the Treated (ATT, ATE):

$$\text{ATT} \equiv E(y_{1i} - y_{0i} \,|\, D_i = 1)$$

- Average Treatment Effect on the Untreated (ATU):

$$\text{ATU} \equiv E(y_{1i} - y_{0i} \,|\, D_i = 0)$$

Assumptions of Propensity Score Matching:

PSM requires 2 important assumptions to be followed strictly; else results of ATT will be biased.

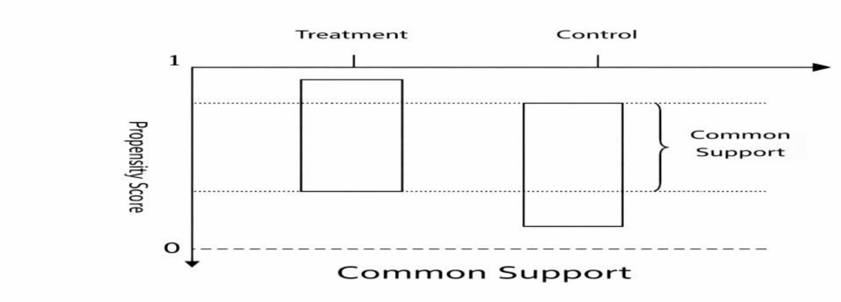
1) CIA (Conditional Independence Assumption) : After controlling for observable features, the outcome is independent whether the individuals participate in the program or not.
   Mathematically,

$$(y_1, y_0) \perp D|x$$

   Household differ in many ways in covariates (age, gender,caste, income, education, land size etc) ,so after controlling of these observed covariates the outcome is independent of treatment.

2) Common support/Overlap condition : For each set of observable features 'x' there must be a positive probability of being both treated and untreated.

Figure-1: Common Support of Treatment and Control Group



Mathematically,

$$0<P(Di=1|Xi=x)<1$$

This assumption ensures that there is an overlap between control and treatment group.

Matching methods:

Different matching algorithms are applied in PSM for estimating unbiased treatment effects.

1) Matching Nearest Neighbors (NN Match):

The control unit with the closest propensity score is paired with each treated unit. Earlier NN match used the Mahalanobis distance for matching which suffers the curse of dimensionality. Then Rosenbaum and Rubin (1983) uses propensity score to measure the distance between x_i and x_j .

2) K- nearest neighbor Matching :

K-nearest matching finds K matching units in the control group with closest p-score. If K =1, it's called "one to one matching" and for k> 1, it's called "one to many". One of the limitation of this is even the closest control unit is far away.

3) Caliper Matching:

This method removes the limitation of K-Nearest neighbor matching where matching all neighbors with an absolute distance of p-score :

$$|p_i - p_j| \leq \varepsilon$$

$$\varepsilon \leq 0.25\hat{\sigma}_{pscore}$$

where,

p(i) = Propensity score of treated individual i

p(j) = Propensity score of control individual j

ε is the caliper distance, usually set as a fraction (like 0.25) of the standard deviation of the propensity scores.

4) Global Matching:

Each treated unit is matched with all control units by assigning weights to the control unit. Closest control unit will get higher weight and vice versa.

5) Kernel Matching:

Similar to global matching but uses a kernel function for weights.

$$w(i,j) = \frac{K\left(\frac{p(x_j) - p(x_i)}{h}\right)}{\sum_{k:D_k=0} K\left(\frac{p(x_k) - p(x_i)}{h}\right)}$$

Mathematically,

where,

h: Bandwidth parameter.

K(·): Kernel function that gives higher weight to observations with closer propensity scores

Dk: Treatment indicator variable (1 = treated, 0 = control).

w(i,j)– Weight assigned to control unit j when matching with treated unit i.

Example using the Lalonde Dataset

The Lalonde dataset is commonly employed to validate the efficiency of the Propensity Score Matching (PSM). This dataset assesses the impact of the National Supported Work (NSW) job training program on the earnings of the participants.

The dataset consists of two categories of individuals:

Treatment group (D = 1): People who have undergone the job training program.

Control group (D = 0): People who have not undergone the job training program.

For each individual, several background information are provided, such as their age, education, race, marital status, and previous earnings.

Using these covariates, a logistic regression model (Logit/Probit) is used to estimate the p-score for each individual. Once the p-score are calculated, individuals in the treatment group are matched with individuals in the control group with similar propensity scores. This makes the groups comparable and reduces selection bias.

For example,

Individuals	Treatment (D)	P -score	Income
A	1	0.74	11,000
B	0	0.70	9000

where individual "A" is the treated unit matched with individual "B" is the control unit as they have similar p-score.

Treatment Effect = 11,000 – 9000 = 2000

This means that participation in the program increased income by 2000 units for this matched pair.

In practice, PSM estimates the Average Treatment Effect on the Treated (ATT) by averaging the income differences between all matched treated and control individuals in the dataset.

Lalonde Dataset : Variables :

Lalonde dataset contains earnings variables and demographic features to estimate p-scores. It contains information on individuals who participated (Treated) and not participated (Control) in the National Supported Work (NSW) job training program .

Some of the important variables are:

treat : Treatment indicator **age** – Age of the individual in years.

educ : Number of years of schooling completed.

race : Categorical variable indicating race (usually categorized as Black, Hispanic, or White).

married : Marital status (1 = married, 0 = not married).

no degree : Indicates whether the individual has a high school degree (1 = no degree, 0 = has degree).

re74 : Real earnings of the individual in the year 1974 (before the program).

re75 : Real earnings in the year 1975 (before the program).

re78 : Real earnings in the year 1978 (after the program), often used as the outcome variable.

The calculated p-score are used to match treated and control individuals with similar characteristics. In PSM, the Covariates such as race, education, previous earnings ,age etc are used to estimate the probability of participating in the training program.. After matching, the difference in post income i.e. re78 between the matched individuals is used to estimate the Average Treatment Effect on the Treated (ATT).

Estimation of PSM using STATA:

PSM analysis is usually estimated using a well known statistical software "STATA".

Important steps include:

1. Select covariates "x" such as age, gender, education level, land size etc (Before matching)
2. Estimate p-score using regression (logit/probit)
3. Matching of treated and control unit as per their p-score
4. Covariate balance (After matching)
5. Estimate treatment effect (ATT/ATE)

Propensity Score Matching: Merits

In impact evaluation application of PSM offers several advantages such as,

- Minimizes selection bias issue through matching of observable variables/ covariates which is done before matching.
- Causal inference can be calculated using observational/survey data
- Widely accepted in empirical economic research
- Provides transparent matching procedures

Propensity Score Matching: Demerits

Although it is widely accepted in empirical economic research still faces PSM several limitations:

- It does not count for unobserved variables such as motivation ,hidden talent ,risk taking ability etc which may give biased result if is ignored.
- Correct model specification is important for unbiased estimates.
- Poor overlap between groups can reduce accuracy
- large sample size needed for reliable matching

Conclusion:

The study aimed to explore the methodological importance of Propensity Score Matching (PSM) as a technique to assess the effectiveness of welfare policy interventions on income outcomes. The research highlighted how, in the absence of experimental designs, program evaluations often encounter a challenge of selection bias, which might cause program evaluations to yield misleading results. The study demonstrated how PSM offers a rigorous technique to address the issue of selection bias by creating a counterfactual comparison group that is comparable to the treatment group on a set of observable characteristics.

The study also explored the conceptual basis, mathematical formulation, and assumptions of PSM. The study further demonstrated how to apply Average Treatment Effect on the Treated (ATT) to measure the impact of a policy intervention on income outcomes. The study also provided an example based on a hypothetical scenario and an example based on a well-known data set, Lalonde Data Set. The study further demonstrated how to apply STATA to estimate ATT.

The results of the methodological discussion suggest that PSM is an important quasi-experimental method for the evaluation of the impact of public policies when the experimental data are lacking. The method enables the production of better estimates of the results of public policies by increasing the comparability of the treated and untreated groups. Nevertheless, the results of the PSM method depend on the quality of the available data and the inclusion of the relevant variables, since the method cannot adjust for the unobservable variables.

Overall, the paper under review suggests the importance of the PSM method to the process of public policy analysis, especially with regard to the income effects of social welfare programs.

References:

1. Abadie, A., & Imbens, G. W. (2016). Matching on the estimated propensity score. *Econometrica, 84*(2), 781–807. https://doi.org/10.3982/ECTA11293
2. Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*(3), 399–424.
3. Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys, 22*(1), 31–72.
4. Dehejia, R. H., & Wahba, S. (1999). Causal effects in non experimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association, 94*(448), 1053–1062.
5. Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for non experimental causal studies. *Review of Economics and Statistics, 84*(1), 151–161.

6. D'Agostino, R. B. (1998). Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine, 17*(19), 2265–2281.
7. Heckman, J. J., Ichimura, H., & Todd, P. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies, 64*(4), 605–654.
8. Heckman, J. J., Ichimura, H., Smith, J., & Todd, P. (1998). Characterizing selection bias using experimental data. *Econometrica, 66*(5), 1017–1098.
9. Hirano, K., Imbens, G. W., & Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica, 71*(4), 1161–1189.
10. Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity. *Review of Economics and Statistics, 86*(1), 4–29.
11. Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature, 47*(1), 5–86.
12. Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis, 20*(1), 1–24.
13. King, G., & Nielsen, R. (2019). Why propensity scores should not be used for matching. *Political Analysis, 27*(4), 435–454.
14. LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review, 76*(4), 604–620.
15. Lechner, M. (2002). Some practical issues in the evaluation of heterogeneous labour market programmes. *Journal of the Royal Statistical Society: Series A, 165*(1), 59–82.
16. Lunceford, J. K., & Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects. *Statistics in Medicine, 23*(19), 2937–2960.
17. Morgan, S. L., & Winship, C. (2015). *Counterfactuals and causal inference: Methods and principles for social research* (2nd ed.). Cambridge University Press.
18. Pearl, J. (2009). *Causality: Models, reasoning and inference* (2nd ed.). Cambridge University Press.
19. Rosenbaum, P. R. (2002). *Observational studies* (2nd ed.). Springer.
20. Rosenbaum, P. R. (2010). *Design of observational studies*. Springer.
21. Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55.
22. Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using sub classification on the propensity score. *Journal of the American Statistical Association, 79*(387), 516–524.
23. Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician, 39*(1), 33–38.
24. Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66*(5), 688–701.
25. Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*(3), 581–592.
26. Rubin, D. B. (2008). For objective causal inference, design trumps analysis. *Annals of Applied Statistics, 2*(3), 808–840.
27. Smith, J. A., & Todd, P. E. (2005). Does matching overcome Lalonde's critique of non experimental estimators? *Journal of Econometrics, 125*(1–2), 305–353.

28. Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science, 25*(1), 1–21.

29. Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). MIT Press.

30. Wooldridge, J. M. (2013). Introductory econometrics: A modern approach (5th ed.). Cengage Learning.