Matching for Quantile Treatment Effect on the treated using different propensity score estimation methods

Hoang Thanh Lam Nguyen, supervised by Wei Huang

2024/2025 Vacation Scholarship Program, the University of Melbourne

Introduction

Causal inference estimates treatment effects, which measure the difference in the probability distributions of potential outcomes for an individual under treatment and non-treatment. Since outcomes are influenced by multiple factors, methods to reduce covariate bias are essential. This study focuses on quantile treatment effects on the treated (QTT) using double-score matching method to reduce biases in the estimator, comparing QTT estimates with propensity scores from logistic and single-index models.

Matching Method

The propensity score and prognostic score are used for matching. Notations are defined as following:

- Z: treatment assignment
- Y: observed outcome
- $Y(0)$: potential outcome without receiving treatment
- $Y(1)$: potential outcome under treatment
- X: observed covariates
- $e(X)$: propensity score
- $\Psi(X)$: prognostic score

The propensity score is defined as [1]:

$$e(X) = P(Z = 1|X)$$

The prognostic score is a sufficient statistic for $Y(a)$ so that [2]:

$$Y(z) \perp\!\!\!\perp |\Psi_z(X) \text{ for z = 0, 1}$$

As a result, we have $E(Y(0)|X)$ and $E(Y(1)|X)$ as the prognostic scores [3].
To estimate QTT, we matched every treated unit with a control unit with the nearest distance based on the two scores. This method, double score matching (DSM), helps reduce imbalance in covariate and disease risk [3].
Specifically, the propensity score is estimated using the logistic and single-index models [4]. The PSID-1 model includes variables such as age, education, marital status, high school degree status, black and hispanic, pre-treatment earnings for 1974 and 1975 (with squared terms), and an interaction between unemployment in 1974 and race. The CPS-1 model adds age cubed, unemployment in 1975, and education-earnings interaction. The prognostic model uses the same variables as the propensity score model but excludes interaction terms.

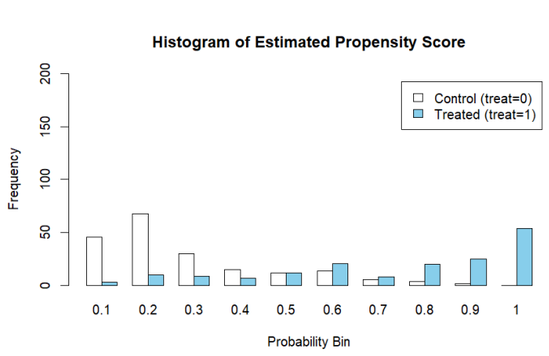

Fig. 1: Single-Index Model

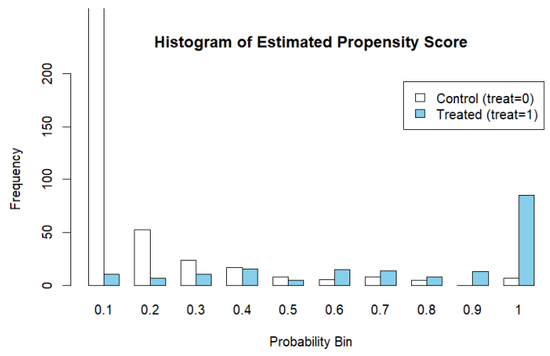

Fig. 2: Logistic Regression Model

Figure 1: Propensity Score Distribution after discarding controls

Result

All datasets have excluded control units with propensity scores outside the range of treated units. The tables below present the 0.25, 0.5, and 0.75 QTT estimates for three datasets using two propensity score estimation methods.

Data	Quantiles		
	0.25	0.5	0.75
NSW	485	1099	2351
PSID-1	-11041	-16456	-19912
CPS-1	-5182	-12189	-15922

Table 1: Quantile Treatment Effects for NSW, PSID-1, and CPS-1 without controlling covariates

Data	Method	Quantiles		
		0.25	0.5	0.75
PSID-1	Logistic	485.23	3214.16	3732.09
	Single-index	485.23	4232.31	5948.68
CPS-1	Logistic	485.23	2333.43	2099.21
	Single-index	485.23	4232.31	9609.01

Table 2: Comparison of Logistic and Single-Index Methods Across Quantiles

Dataset

This study uses the Lalonde dataset, as subsampled by Dehejia and Wahba [4]. The NSW Demonstration aimed to provide work experience to individuals facing economic and social challenges [4]. Pre-intervention data were collected from surveys and Social Security records. Lalonde's nonexperimental treatment effect estimates use two comparison groups: the Panel Study of Income Dynamics (PSID-1) and the Matched Current Population Survey-Social Security Administration File (CPS-1). Overall, these include 185 treated units, 260 controls, 15,992 CPS-1 observations, and 2,490 PSID-1 observations.

QTT Estimators for one-to-one matching [3]

The quantile treatment effect on the treated (QTT) is defined as the difference in population quantiles of the potential outcome distributions for the treated group [3]. After matching treated units with comparable control units based on X, under standard positivity, ignorability assumptions and overlap condition, we estimate the quantiles as follows [3]:

1. For the quantiles of the $Y(1)$ distribution, $q_{1,\xi|Z=1} = \inf_q \{\mathbb{P}(Y \leq q \mid Z = 1) \geq \xi\}$ is estimated by:

$$\widehat{q}_{1,\xi|Z=1} = \inf_q \{\widehat{F}_1(q \mid Z = 1) \geq \xi\}$$

where: $\widehat{F}_1(q \mid Z = 1) = n_1^{-1} \sum_{i=1}^n Z_i \mathbf{1}(Y_i \leq q)$ and n_1 is the number of treated units $(Z = 1)$.

2. For the quantiles of the $Y(0)$ distribution, we want to estimate:

$$q_{0,\xi|Z=1} = \inf_q \{\mathbb{E}[\mathbb{P}\{Y \leq q \mid Z = 0, e(X)\} \mid Z = 1] \geq \xi\}$$

$$= \inf_q \{\mathbb{E}[\mathbb{P}\{Y \leq q \mid Z = 0, \Psi_0(X)\} \mid Z = 1] \geq \xi\}$$

The DSM estimators are given by:

$$\widehat{q}_{0,\xi|Z=1} = \inf_q \{\widehat{F}_{0,\text{dsm}}(q \mid Z = 1) \geq \xi\}$$

where:

- $\widehat{F}_{0,\text{dsm}}(q \mid Z = 1) = \widehat{F}_0^{(0)}(q \mid Z = 1) - n_1^{-1/2} \widehat{B}_{0,n}(q)$
- $\widehat{F}_0^{(0)}(q \mid Z = 1) = n_1^{-1} \sum_{i=1}^n (1 - Z_i) \mathbf{1}(\hat{Y}_i(0) \leq q)$
- $\hat{Y}_i(0)$ represents the matched control unit's predicted outcome based on the nearest distance

In addition, the bias term $\widehat{B}_{0,n}(q)$ is defined as:

$$\widehat{B}_{0,n}(q) = -n_1^{-1/2} \sum_{i=1}^n Z_i \{\widehat{F}_0(q; \{e(X_i), \Psi_0(X_i)\}) - \widehat{F}_0(q; \{e(X_i^{\text{match}}), \Psi_0(X_i^{\text{match}})\})\}$$

where:

- $F_0(q; \{e(X_i), \Psi_0(X_i)\}) = \mathbb{P}\{Y(0) \leq q \mid \{e(X_i), \Psi_0(X_i)\}\}$.
- X_i^{match} refers to the covariates of the matched control unit $\hat{Y}_i(0)$ for treated unit i.

3. Finally, the Quantile Treatment Effect for the treated $\Delta_{QTT,\text{dsm}} = q_{1,\xi|Z=1} - q_{0,\xi|Z=1,\text{dsm}}$ is estimated by:

$$\widehat{\Delta}_{QTT,\text{dsm}} = \widehat{q}_{1,\xi|Z=1} - \widehat{q}_{0,\xi|Z=1}$$

Discussion

- Propensity scores estimated by the single-index model result in more values extremely close to or exactly 1 and 0 compared to the logistic model before discarding some control units. As a result, the number of discard control observations using single-index models increases, potentially affecting the matching quality, as we have fewer control units.
- Because we do not estimate the variance, we cannot say if our results lead to the same significance inference as the experimental QTT estimates. Still, the difference between experimental and non-experimental QTT estimates were significantly reduced, as the result of the matching has narrowed the bias and balanced the covariates.
- However, the estimators, particularly under the single-index model, tend to overestimate QTT values, especially at higher quantiles. This overestimation might reflect the instability of the single-index model. The single-index model involves nonparametric part, hence, leads to higher variance than the logistic regression model and gives more extreme estimated values.

References

[1] Peng Ding. *A First Course in Causal Inference*. 2023. DOI: 10.48550/arXiv.2305.18793. URL: https://arxiv.org/abs/2305.18793.

[2] Ben B. Hansen. "The prognostic analogue of the propensity score". In: *Biometrika* 95.2 (June 2008), pp. 481–488. DOI: 10.1093/biomet/asn004. URL: https://doi.org/10.1093/biomet/asn004.

[3] S. Yang and Y. Zhang. "Multiply robust matching estimators of average and quantile treatment effects". In: *Scandinavian Journal of Statistics* 50.1 (2023), pp. 235–265. DOI: 10.1111/sjos.12585. URL: https://doi.org/10.1111/sjos.12585.

[4] Rajeev H. Dehejia and Sadek Wahba. "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs". In: *Journal of the American Statistical Association* 94.448 (1999). Accessed: 16-12-2024 02:15 UTC, pp. 1053–1062. URL: https://www.jstor.org/stable/2669919.