Computational Statistics and Data Analysis 115 (2017) 91–102

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

The finite sample performance of semi- and non-parametric estimators for treatment effects and policy evaluation

Markus Frölich [a], Martin Huber [b,*], Manuel Wiesenfarth [c]

[a] University of Mannheim, Department of Economics, L7, 3-5, D-68131 Mannheim, Germany
[b] University of Fribourg, Department of Economics, Bd. de Pérolles 90, CH-1700 Fribourg, Switzerland
[c] German Cancer Research Center, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany

ARTICLE INFO

Article history:
Received 22 March 2016
Received in revised form 5 May 2017
Accepted 20 May 2017
Available online 7 June 2017

Keywords:
Treatment effects
Policy evaluation
Simulation
Empirical Monte Carlo study
Propensity score
Semi- and non-parametric estimation

ABSTRACT

The finite sample performance of a comprehensive set of semi- and non-parametric estimators for treatment evaluation is investigated. The simulation design is based on Swiss labor market data and considers estimators based on parametric, semiparametric, and nonparametric propensity scores, as well as approaches directly controlling for covariates. Among the methods included are pair, radius, kernel, and genetic matching, inverse probability weighting, regression, doubly robust estimation, entropy balancing, and empirical likelihood estimation. The simulation designs vary w.r.t. sample size, selection into treatment, effect heterogeneity, and (non-)omission of a subset of the all in all 3 continuous and 11 binary confounders. Several nonparametric estimators outperform commonly used treatment estimators based on parametric propensity scores in terms of root mean squared error (RMSE), even though average RMSEs based on the 16 simulation designs considered are not statistically significantly different across the estimators investigated.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Estimators for the evaluation of binary treatments in observational studies under a 'selection-on-observables' assumption (see for instance Imbens, 2004 and Imbens and Wooldridge, 2009) are widely applied in empirical research. In most cases, the aim is to estimate the average causal effect of a treatment on an outcome by controlling for differences in observed covariates across treated and non-treated sample units. More generally, such estimators apply to any problem in which the means of an outcome variable in two subsamples should be purged from differences due to observed variables.

Most treatment effect estimators control for the treatment propensity score, i.e. the conditional probability to receive the treatment given the covariates, rather than directly for the covariates. Popular approaches include propensity score matching (see for instance Rosenbaum and Rubin, 1985, Heckman et al., 1998b, and Dehejia and Wahba, 1999) and inverse probability weighting (henceforth IPW, Horvitz and Thompson, 1952 and Hirano et al., 2003). Doubly robust estimation (henceforth DR, Robins et al., 1992, 1994, 1995, Robins and Rotnitzky, 1995, and Bang and Robins, 2005) relies on models for both the propensity score and the conditional mean outcome. In almost all applications, such estimators make use of parametric propensity score specifications. However, functional misspecification of the propensity score may entail an inconsistent treatment effect estimator, which is avoided in nonparametric modeling. In this light, investigating the finite sample behavior of fully nonparametric methods appears interesting, but is lacking in previous simulation studies, see Frölich (2004), Zhao

* Corresponding author.
E-mail addresses: froelich@uni-mannheim.de (M. Frölich), martin.huber@unifr.ch (M. Huber), m.wiesenfarth@dkfz-heidelberg.de (M. Wiesenfarth).

http://dx.doi.org/10.1016/j.csda.2017.05.007
0167-9473/© 2017 Elsevier B.V. All rights reserved.

(2004), Lunceford and Davidian (2004), Huber et al. (2013), Busso et al. (2014), and Bodory et al. (2016) (who in contrast to the previous papers investigate the performance of variance estimators rather than point estimators of treatment effects).

This paper aims at closing this gap by analyzing a comprehensive set of semi- and non-parametric estimators of the average treatment effect on the treated (ATET) in a simulation study based on labor market data from Behncke et al. (2010a, b). We vary several design features in the simulations – sample size, selection into treatment, effect heterogeneity, and (non-)omission of confounders – and include a range of propensity score methods: pair, kernel, and radius matching, IPW, and DR based on the efficient influence function and the targeted maximum likelihood (TMLE) approach of van der Laan and Rubin (2006), respectively. We also consider several propensity score estimators: probit estimation, the semiparametric estimator of Klein and Spady (1993), nonparametric kernel regression, and the 'covariate balancing propensity score' method (CBPS) of Imai and Ratkovic (2014). The latter is an empirical likelihood approach obtaining exact balancing of covariate moments in the sample and is thus related to inverse probability tilting (IPT) suggested in Graham et al. (2012, 2011), which is also implemented as a weighting estimator. Our analysis includes several nonparametric ATET estimators not relying on the propensity score: pair, radius, or one-to-many matching on the covariates, nonparametric regression, genetic matching (see Diamond and Sekhon, 2013), and entropy balancing (see Hainmueller, 2012). Finally, parametric regression among the nontreated is also considered.

In our simulations, we find that several (but not all) nonparametric ATET estimators – namely regression, DR based on the influence function as considered in Cattaneo (2010) and Rothe and Firpo (2013), and IPW – by and large outperform ATET estimators based on a parametric propensity score in terms of the root mean squared error (RMSE). This finding holds across various simulation features. For instance, our results suggest that these nonparametric methods can be quite competitive even in moderate samples with 750 observations. However, as a word of caution it needs to be pointed out that the dimension of covariates considered is moderate (3 continuous and 11 dummy variables) and that the relative performance of the nonparametric methods likely deteriorates with higher dimensions. Furthermore, average RMSEs based on the 16 simulation designs considered are not statistically significantly different across the methods investigated.

The remainder of this paper is organized as follows. Section 2 introduces the ATET and propensity score estimators. Section 3 discusses the data and simulation design. Section 4 presents the main results on the performance of the estimators. Section 5 concludes. Further results are provided in the separate online appendix.

2. Estimators

2.1. Overview

Let D denote the binary treatment indicator, Y the outcome, and X a vector of observed covariates. We would like to compare the mean outcome of the treated group ($D = 1$) to that of the non-treated group ($D = 0$), after making the latter group comparable to the former in terms of X. Formally, the parameter of interest is

$$\Delta = E[Y|D = 1] - E[E[Y|D = 0, X]|D = 1]. \tag{1}$$

Δ corresponds to the average treatment effect on the treated (ATET) if the 'selection on observables' or 'conditional independence' assumption is invoked (see for instance Imbens, 2004), which rules out confounders that jointly influence D and Y conditional on X. This parameter has received much attention in the program evaluation literature, even though it may also be considered in non-causal contexts such as wage gap decompositions.

Estimators of (1) either control for X directly or for the conditional treatment probability $\Pr(D = 1|X)$, henceforth referred to as propensity score. The latter possesses the so-called 'balancing property', see Rosenbaum and Rubin, 1983, such that conditioning on $\Pr(D = 1|X)$ equalizes the distribution of X across D. In Section 2.4 we present estimators that directly control for X, while Section 2.3 introduces methods that make use of the propensity score and therefore depend on plug-in estimators of the propensity score. The various propensity score estimators are discussed in Section 2.2.

2.2. Estimation of propensity score

We define the propensity score as a real valued function of x as $p(x) = \Pr(D = 1|X = x)$ and consider the following estimators: probit regression, the empirical likelihood-based method of Imai and Ratkovic (2014), semiparametric estimation as suggested in Klein and Spady (1993), and nonparametric kernel regression. In all settings we assume an i.i.d. sample of size N containing $\{Y_i, D_i, X_i\}$ for each observation i.

The probit estimator of $p(X_i)$ is given by $\hat{p}(X_i) = p(X_i; \hat{\beta}) = \Phi((1, X_i)'\hat{\beta})$, with Φ being the standard normal cumulative distribution function (cdf) and $\hat{\beta}$ the coefficients from maximum likelihood (ML) estimation. However, if the propensity score model is misspecified, $\hat{p}(X)$ might not (fully) balance X. As a second approach, we therefore consider the 'covariate balancing propensity score' (CBPS) procedure of Imai and Ratkovic (2014), an empirical likelihood (EL) method which models treatment assignment and at the same time optimizes covariate balance. The CBPS algorithm starts out from a logit model for the propensity score and then adjusts the coefficients of the initial model such that specific moments of the covariates are fully balanced in the sample. The ML score condition of the propensity score model may serve as additional moment condition in the estimation. In the simulations, we consider a just identified version of CBPS method for almost all propensity score based

Download English Version:

https://daneshyari.com/en/article/4949239

Download Persian Version:

https://daneshyari.com/article/4949239

Daneshyari.com