Statistics and Its Interface Volume 15 (2022) 313–322

Estimation of conditional average treatment effect by covariates balance methods

Jun Wang and Changbiao Liu*

Conditional average treatment effects estimation is one of the crucial mainstays in observational studies. The conditional average treatment effect is defined as a functional parameter which is used to describe the variation of average treatment effect condition on some covariates. Based on the unconfoundedness assumption, we propose the covariates balance method to estimate the propensity score, and the estimated propensity score is applied to the non-parametric method to estimate the conditional average treatment effect. The proposed method is robust and superior to the parametric approach. The proposed method has a smaller RMSE than the true method when the propensity score model is correct specified. Meanwhile, compared with the kernel method, the proposed method is much more computationally efficient. The proposed estimator is consistent and asymptotic under some regularity conditions. Finally, we apply the proposed method to estimate the effect of maternal smoking on low birth weight infants given the age of mothers.

Keywords and phrases: Conditional average treatment effects, Covariates balance, Heterogeneity, Propensity score, Unconfoundedness.

1. INTRODUCTION

Although the individual treatment effect is heterogeneous in the population, the average treatment effect (ATE) in the population is still identifiable and estimable (see, Rosenbaum and Rubin 1983, Hahn 1998) when the treatment assignment is unconfounded given the observed covariates X. Given the heterogeneity of individual treatment effects, it may also be of interest to estimate the curve of the average treatment effect given some covariates, we call the value of average treatment effects given some covariates as conditional average treatment effects (CATE). For example, if the covariates is discrete (gender), the researchers may be separately concerned about estimating the average treatment effect given male or female, if the covariates is continuous (age), researchers are interested in estimating the average treatment effect at a particular age or specific age interval.

Many researchers define CATE as the average treatment effect at a certain point $X_1 = x_1$. In this case, CATE(x_1) represents the conditional mean of the treatment effect of any point $x_1 \in \mathcal{X}_1$, where $\mathcal{X}_1$ denotes the support of X_1, and we consider that the changes of CATE(x_1) as x_1 changes. In this paper, we mainly consider the case that X_1 is continuous.

Hahn(1998), Heckman, Ichimura and Todd (1997, 1998) introduced CATE(x) as the first stage of the ATE estimator. Robins and Ritov (1997), Angrist and Hahn (1999), Heckman and Vytlacil (2005) discussed the identification and estimation of CATE(x). Although Khan and Tamer (2010) explicitly mentions CATE(x), however, their focus was on ATE. Lee and Whang (2009) and Hsu (2012) considered estimating and testing hypotheses about CATE(x) when X is absolutely continuous, and gave the detailed asymptotic theory. MaCurdy, Chen and Hong (2011) also discussed the identification and estimation of CATE(x). Abrevaya, Hsu and Lieli (2015) extended the concept of CATE to a technically challenging situation that the conditioning covariates X_1 are continuous and form a strict subset of X. As the unconfoundedness assumption will not generally hold conditional on X_1 alone. The estimation process is to estimate CATE as a function of X, and then calculate the average of unwanted components by integrating as for the conditional distribution of $X_{(1)}$ given X_1, where $X_{(1)}$ denotes the remaining variable in X after removing X_1. However, this distribution is generally unknown and needs to be estimated.

Abrevaya Hsu and Lieli (2015) proposed to estimate the CATE by the following two steps. First, the propensity score that the probability of treatment conditional on X, is estimated either by a kernel-based regression method or by a parametric model method. In the second step the observed outcomes are analyzed based on the treatment status and the inverse of the estimated propensity score, and local averages are computed around points in the support X_1, using another set of kernel weights. (Intuitively, the second stage can be interpreted as integrating it with respect to a smoothed estimate of the conditional distribution of the inverse propensity weighted outcomes given X_1.) Particularly, CATE is the conditional average treatment effect given $X_1 = x_1$.

Kang and Schafer (2007), Smith and Todd (2005) showed that when the estimation of propensity score is even slightly biased, it may lead to a severe deviation in the estimation of average treatment effects. Therefore, the estimation of the propensity score model is wrong, then the CATE is biased.

*Corresponding author.

In this paper, we propose the estimation of the propensity score by the covariates balance methods (Imai and Ratkovic (2014), Wong and Chan (2018), Li *et al.*, (2018), Zhao (2019)). The method is robust to mild misspecification of the propensity score model. Under some regularity conditions, the proposed estimator is shown to be consistent and asymptotically normal. The simulation results show that the covariates balance method is superior to the parametric method when the propensity score model is misspecified. The covariates balance method has a smaller RMSE than the true method when the propensity score model is correctly specified, and has less computational cost than the kernel method, the proposed method is not affected from the dimensional changes of X.

The rest of the article is organized as follows. Section 2 introduces the CATE parameter and discusses its identification and estimation. The asymptotic properties of the proposed estimators are developed. In Section 3, simulation studies will be conducted to evaluate the performance of the proposed estimator. In Section 4, we apply our method to estimate the effect of maternal smoking during pregnancy on low birth weight infants given the age of mothers. Conclusions will be made in Section 5. Proofs of theorems will be presented in the Appendix.

2. THE PROPOSED ESTIMATOR

2.1 The formal framework and estimator

Let D be a binary treatment variable in a population of interest with $D = 1$ if the individual receives active treatment and $D = 0$ if the individual receives control treatment. $Y(1)$ and $Y(0)$ are potential outcomes given treatment status $D = 1$ and $D = 0$, respectively. Let X be k dimensional covariates with $k \geq 2$. Suppose that a random sample of n i.i.d. units (Y_i, X_i, D_i) $(i = 1, \cdots, n)$ are observed, where $Y_i = D_i Y_i(1) + (1 - D_i)Y_i(0)$. In order to identify CATE, we give the following assumption.

Assumption 1 (Unconfoundedness)**.** D is independent of $(Y(0), Y(1))$ conditional on X, i.e., $(Y(1), Y(0)) \perp D \mid X$.

Assumption 1 means that it rules out the existence of unobserved confounding factors that affect treatment variable and are also correlated with the potential outcomes, namely D is exogenous. It is a standard identification assumption in the treatment effect literature (Rosenbaum and Rubin 1983, Hirano *et al.*, 2003).

The propensity score is defined as the probability of treatment variable conditional on X, i.e.,

$$\pi(X) = P(D = 1 \mid X).$$

For example, a popular choice is the logistic model

$$\pi_{\boldsymbol{\beta}}(X) = \frac{\exp(X^T\boldsymbol{\beta})}{1 + \exp(X^T\boldsymbol{\beta})}. \tag{1}$$

where $\boldsymbol{\beta}$ is a k-dimensional vector of unknown parameters.

Assumption 2 (Overlap)**.** $\underline{p} \leq \pi(X) \leq \bar{p}$ almost surely, for some $\underline{p} > 0$ and $\bar{p} < 1$.

Assumption 2 implies that the population includes treated and untreated units for almost all values of X. Assumption 1 and 2 are often referred to as "strong ignorability" assumption. It means that adjusting for X is sufficient to eliminate all confounding factors.

Let $X_1 \in \mathbb{R}^\ell$ be a strict subset of $X \in \mathbb{R}^k$, $1 \leq \ell < k$, X is absolutely continuous. The conditional average treatment effect (CATE) given $X_1 = x_1$ is defined as

$$\tau(x_1) = E[Y(1) - Y(0) \mid X_1 = x_1].$$

Under the assumption 1 and 2, $\tau(x_1)$ can be written as

$$\tau(x_1) = E\Big[E[Y \mid D = 1, X] - E[Y \mid D = 0, X] \mid X_1 = x_1\Big], \tag{2}$$

or

$$\tau(x_1) = E\left[\frac{DY}{\pi(X)} - \frac{(1 - D)Y}{1 - \pi(X)} \mid X_1 = x_1\right]. \tag{3}$$

The CATE can be identified and estimated by Equation (2) and (3). Two regression functions require being estimated in Equation (2) and the propensity score model need to be estimated in Equation (3). In this paper, we mainly consider the case (3), and the estimation of propensity score models may suffer from dimensionality disasters (nonparametric case) and model misspecification (parameter case). For example, Abrevaya, Hsu and Lieli (2015) proposed a nonparametric method (kernel estimation) or parameter models(eg, logit, probit model) to estimate propensity score models. We expand the result of Abrevaya Hsu and Lieli (2015). Kang and Schafer (2007), Smith and Todd (2005) indicated that the propensity score model plays an important role in estimating ATE, the ATE estimator would face a significant deviation for the incorrect estimation of propensity score models. Estimate the $\boldsymbol{\beta}$ by maximum likelihood estimator (MLE), and it implies the first order condition is

$$\frac{1}{n}\sum_{i=1}^{n} s_{\boldsymbol{\beta}}(D_i, X_i) = 0, \tag{4}$$

$$s_{\boldsymbol{\beta}}(D, X) = \frac{D\pi_\beta^{'}(X)}{\pi_\beta(X)} - \frac{(1 - D)\pi_\beta^{'}(X)}{1 - \pi_\beta(X)},$$

where $\pi_\beta^{'}(X) = \partial\pi_\beta(X)/\partial\boldsymbol{\beta}$. However, the main challenge of this standard method is that $\pi_{\boldsymbol{\beta}}(X)$ may be misspecified, which lead to the biased estimators for the parameters $\boldsymbol{\beta}$. Inspired by Imai and Ratkovic (2014), we operationalize the covariates balance method to estimate the propensity score to tackle the above-mentioned problem, namely,

$$E\left\{\frac{D\tilde{X}}{\pi_\beta(X)} - \frac{(1 - D)\tilde{X}}{1 - \pi_\beta(X)}\right\} = 0,$$

where $\tilde{X}_i = f(X_i)$ is an M-dimensional vector-valued measurable function of X_i specified by the researcher. Its sample form is

$$\frac{1}{n}\sum_{i=1}^{n} w_\beta(D_i, X_i)\tilde{X}_i = \frac{1}{n}\sum_{i=1}^{n} g_\beta(D_i, X_i), \tag{5}$$
$$w_\beta(D_i, X_i) = \frac{D_i - \pi_\beta(X_i)}{\pi_\beta(X_i)(1-\pi_\beta(X_i))}.$$

When the moment condition is exactly identifiable, the moment estimation can be used to estimate the parameter; if the moment condition is over-identifiable, the GMM methods (Hansen 1982) can be used to estimate the parameter, namely

$$\hat{\beta}_{GMM} = \arg\min_{\beta\in\Theta} \bar{g}_\beta(D,X)^T \Sigma_\beta^{-1}(D,X)\bar{g}_\beta(D,X), \tag{6}$$

where $\bar{g}_\beta(D,X)^T = \begin{pmatrix} s_{\boldsymbol{\beta}}(D,X) \\ w_\beta(D_i,X_i)\tilde{X}_i \end{pmatrix}$, $\Sigma_\beta(D,X)$ is the consistent covariance estimator of $\bar{g}_\beta(D_i,X_i)$,

$$\Sigma_\beta(D,X) = \frac{1}{n}\sum_{i=1}^{n} E\{\bar{g}_\beta(D_i,X_i)\bar{g}_\beta(D_i,X_i)^T \mid X_i\}.$$

The estimated propensity score model satisfies the following assumptions.

Assumption 3 (Estimated propensity score)**.** $\hat{\pi}(X_i)$ is estimated by the covariates balance methods, and satisfies $\sup_{x\in\mathcal{X}} |\hat{\pi}(x) - \pi(x)| = O_p(n^{-1/2})$ for all $x \in \mathcal{X}$.

Given the estimator of propensity score $\hat{\pi}(X)$, we estimate $\tau(x_1)$ by the following formal

$$\hat{\tau}(x_1) = \frac{\frac{1}{nh^\ell}\sum_{i=1}^{n}\left(\frac{D_iY_i}{\hat{\pi}(X_i)} - \frac{(1-D_i)Y_i}{1-\hat{\pi}(X_i)}\right)K\left(\frac{X_{1i}-x_1}{h}\right)}{\frac{1}{nh^\ell}\sum_{i=1}^{n} K\left(\frac{X_{1i}-x_1}{h}\right)}, \tag{7}$$

the kernel-based local average about x_1, where $K(\cdot)$ is a kernel function and h is a smoothing parameter (bandwidth).

2.2 Asymptotic properties of $\hat{\tau}(x_1)$

We estimate the propensity score model by the covariates balance methods. We derive the asymptotic properties of the CATE estimator based on the following regularity conditions.

Assumption 4 (i) $\mathcal{X}$ denotes the support of X, $\sup_{x\in\mathcal{X}} E[Y(d)^2 \mid X = x] < \infty$ for $d = 0, 1$; (ii) the density of X $f(x)$ is bound away from 0 on $\mathcal{X}$, the functions $m_d(x) = E[Y(d) \mid X = x]$ for $d = 0, 1$, $\pi(x)$ and $f(x)$ are $(s \geq k)$ times continuously differentiable on $\mathcal{X}$.

Assumption 5 (Kernels)**.** $K(u)$ is a kernel of order s, is symmetric around zero, and is s times continuously differentiable.

Assumption 6 (Bandwidths)**.** The bandwidths h satisfies the following conditions as $n \to \infty$:
(1) $h \to 0$ and $nh^\ell \to \infty$;
(2) $nh^\ell h^{2s} \to 0$.

Define the function $\psi(x, y, d)$ is

$$\psi(x,y,d) = \frac{dy}{\pi(x)} - \frac{(1-d)y}{1-\pi(x)}.$$

The following theorem states our main theoretical result.

Theorem 1. Suppose that Assumptions 1–6 are satisfied, and $\int |K(u)|^{2+\delta} du < \infty$ for some $\delta > 0$. For each point x_1 in the support of X_1, we have
(a),

$$\sqrt{nh^\ell}(\hat{\tau}(x_1) - \tau(x_1)) = \frac{1}{\sqrt{nh^\ell}}\frac{1}{f_1(x_1)} \times \sum_{i=1}^{n}\Big(\psi(X_i, Y_i, D_i) - \tau(x_1)\Big)K\left(\frac{X_{1i}-x_1}{h}\right) + o_p(1),$$

and (b),

$$\sqrt{nh^\ell}(\hat{\tau}(x_1) - \tau(x_1)) \xrightarrow{d} N\left(0, \frac{\|K\|_2^2 \sigma_\psi^2(x_1)}{f_1(x_1)}\right),$$

where $f_1(x_1)$ is the probability density function of x_1, $\|K\|_2 = \left(\int K(u)^2 du\right)^{1/2}$, and $\sigma_\psi^2(x_1) = E[(\psi(X,Y,D) - \tau(x_1))^2 \mid X_1 = x_1]$.

3. SIMULATION STUDY

In this section we present a simulation study aim at evaluating the finite sample properties of $\hat{\tau}(x_1)$ in Section 2.1.

First, the data generation process in the similar way as Imai and Ratkovic (2014) and Kang and Schafer (2007). Considering two cases: $k = \dim(X) \in \{2, 4\}$. In the first place, let $X = (X_1, X_2)$ be given by

$$X_1 = \epsilon_1, \quad X_2 = \epsilon_2$$

where $\epsilon_1 \sim$ unif$[-0.5, 0.5]$, $\epsilon_2 \sim N(0,1)$. The potential outcomes are generated as follow

$$Y_1 = X_1(1+2X_1)^2(-1+X_1)^2 + 2.5X_2 + v_1,$$
$$Y(0) = 2.5X_2 + v_0,$$

where $v_d \sim N(0, 0.25^2)$ $d = 0, 1$, and is independent with ϵ_1, ϵ_2. Therefore the CATE(x_1) is

$$\begin{aligned}\text{CATE}(x_1) &= E[Y(1) \mid X_1 = x_1] - E[Y(0) \mid X_1 = x_1] \\ &= x_1(1+2x_1)^2(-1+x_1)^2.\end{aligned}$$

The propensity score is generated as follows

$$P(D = 1 \mid X) = \frac{\exp(X_1 + X_2)}{1 + \exp(X_1 + X_2)}.$$

We estimate CATE(x_1), $x_1 \in \{-0.4, -0.2, 0, 0.2, 0.4\}$, for the samples size n = 500 and n = 2000. The following four scenarios are considered:
(i). The true propensity score is a logistic regression with X, which is denoted as True;
(ii). The incorrect propensity score model is a logistic regression with $X_i^* = (X_{1i}^*, X_{2i}^*)$ being the linear predictor; where $X_{1i}^* = \exp(X_{1i}/2)$, $X_{2i}^* = X_{2i}/(1+\exp(X_{1i})) + 10$, which is denoted as Misspecification;
(iii). The just-identified CBPS estimation with the covariates balancing moment conditions with respect to X_i^*, which is denoted as CBPS;
(iv). The just-identified CBPS estimation with the covariates balancing moment conditions with respect to X_i, which is denoted as CBPS1. The simulation result is reported in Table 1–2 for the number of Monte Carlo repetitions is 1000. As shown in Table 1–2, we find that the inverse probability weighted estimator encounter a large bias and RMSE when the propensity score model is misspecified. The CBPS estimator has a much lower RMSE for the true propensity score estimator, and the true propensity score model is standard logistic regression with X_i being the linear predictor, the CBPS method reduces RMSE at the expense of some increase in bias. Therefore the CBPS estimator is also comparable with the true propensity score estimator. The true propensity score estimator is a maximum likelihood estimator, while the CBPS estimator is a moment estimator. In general, the maximum likelihood estimator makes stricter assumptions than the moment estimator and is thus typically less robust. The covariates balance method is superior to the parametric method for the propensity score model is misspecified. Moreover, the covariates balance method is robust for the specification of propensity score models.

Next, Abrevaya *et al.* (2015) considered that CATE was estimated by estimating propensity scores by nonparametric methods, which it suffers from the curse of dimensionality when the dimension of X is large. To investigate the impact of the curse of dimensionality on the covariates balancing method, we also consider the following modification to the DGP: let $\dim(X) = 4$, namely, $X = (X_1, X_2, X_3, X_4)$

$$X_1 = \epsilon_1,\ X_2 = \epsilon_2,\ X_3 = \epsilon_3,\ X_4 = \epsilon_4,$$

where $\epsilon_1 \sim$ unif$[-0.5, 0.5]$, $\epsilon_i \sim N(0,1)$, $i = 2,3,4$. The potential outcome is

$$\begin{aligned} Y(1) &= X_1(1+2X_1)^2(-1+X_1)^2 + 1.5X_2 + 1.5X_3 \\ &\quad + 1.5X_4 + v_1, \\ Y(0) &= 1.5X_2 + 1.5X_3 + 1.5X_4 + v_0, \end{aligned}$$

where $v_d \sim N(0, 0.25^2)$, $d = 0, 1$ and is independent of $\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4$. The propensity score is

$$\pi(X) = \frac{\exp\Big(0.5(X_1 + X_2 + X_3 + X_4)\Big)}{1 + \exp\Big(0.5(X_1 + X_2 + X_3 + X_4)\Big)}.$$

We estimate CATE(x_1), $x_1 \in \{-0.4, -0.2, 0, 0.2, 0.4\}$, for the samples size n = 500 and n = 2000. The following four scenarios are considered:

Table 1. Estimation the CATE by the four methods for n = 500 and 1000 repetitions with dim(x) = 2

	Parameter	True			Misspecification			CPBS			CBPS1		
x_1		Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
					h $= 0.102n^{-1/9} = 0.044$								
0.4	0.46656	0.061	0.763	0.763	2.027	0.476	2.082	0.519	0.556	0.760	0.130	0.651	0.664
0.2	0.25088	0.002	0.666	0.667	2.039	0.469	2.093	0.403	0.569	0.697	0.189	0.616	0.644
0	0	0.087	0.671	0.670	2.021	0.467	2.074	0.193	0.608	0.638	0.166	0.618	0.639
-0.2	-0.10368	-0.023	0.880	0.879	2.026	0.489	2.084	-0.022	0.744	0.744	0.161	0.691	0.709
-0.4	-0.03136	0.064	0.746	0.747	2.008	0.489	2.066	-0.161	0.734	0.751	0.217	0.649	0.684
					h $= 0.067n^{-1/9} = 0.029$								
0.4	0.46656	-0.046	0.969	0.970	2.030	0.576	2.110	0.517	0.684	0.857	0.121	0.819	0.828
0.2	0.25088	0.034	0.846	0.847	2.030	0.587	2.113	0.398	0.721	0.823	0.182	0.788	0.808
0	0	0.013	0.854	0.853	2.019	0.585	2102	0.196	0.771	0.795	0.169	0.785	0.802
-0.2	-0.10368	-0.039	1.128	1.128	2.021	0.603	2.109	-0.040	0.967	0.967	0.144	0.889	0.900
-0.4	-0.03136	0.065	0.922	0.923	2.017	0.592	2.102	-0.150	0.935	0.946	0.230	0.811	0.843

Table 2. Estimation the CATE by the four methods for n = 2000 and 1000 repetitions with dim(x) = 2

	Parameter	True			Misspecification			CPBS			CBPS1		
x_1		Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
					h $= 0.102n^{-1/9} = 0.044$								
0.4	0.46656	0.016	0.401	0.401	2.070	0.264	2.087	0.510	0.309	0.434	0.073	0.379	0.386
0.2	0.25088	-0.018	0.369	0.369	2.022	0.260	2.039	0.296	0.317	0.434	0.039	0.360	0.362
0	0	-0.003	0.390	0.390	2.019	0.247	2.034	0.095	0.369	0.458	0.054	0.380	0.384
-0.2	-0.10368	0.005	0.412	0.412	2.036	0.262	2.053	-0.134	0.438	0.458	0.064	0.398	0.403
-0.4	-0.03136	0.007	0.406	0.406	1.973	0.259	1.990	-0.374	0.464	0.596	0.066	0.382	0.387
					h $= 0.067n^{-1/9} = 0.029$								
0.4	0.46656	0.023	0.504	0.505	2.078	0.318	2.102	0.518	0.380	0.642	0.081	0.476	0.483
0.2	0.25088	-0.018	0.464	0.465	2.019	0.321	2.102	0.295	0.397	0.494	0.038	0.452	0.453
0	0	-0.015	0.490	0.490	2.011	0.303	2.034	0.084	0.461	0.469	0.043	0.475	0.477
-0.2	-0.10368	0.007	0.524	0.523	2.037	0.325	2.062	-0.131	0.558	0.573	0.066	0.504	0.509
-0.4	-0.03136	0.009	0.510	0.510	1.974	0.320	1.999	-0.372	0.594	0.701	0.068	0.481	0.486

(i). The true propensity score model is a logistic regression with X, which is denoted as True;
(ii). The incorrect propensity score model is a logistic regression with $X^* = (X_1^*, X_2^*, X_3^*, X_4^*)$ being the linear predictor, where $X_{1i}^* = \exp(X_{1i}/2)$, $X_{2i}^* = X_{2i}/(1+\exp(X_{1i}))+10$, $X_{3i}^* = (X_{1i}X_{3i}/25+0.6)^3$, $X_{4i}^* = (X_{1i}+X_{4i}+20)^2$, which is denoted as Misspecification;
(iii). The just-identified CBPS estimation with the covariates balancing moment conditions with respect to X_i^*, which is denoted as CBPS;
(iv). The just-identified CBPS estimation with the covariates balancing moment conditions with respect to X_i, which is denoted as CBPS1. The simulation result is reported in Table 3–4 for the number of Monte Carlo repetitions is 1000. We can observe that the misspecified estimator encounter a greater bias and RMSE than CBPS methods when the propensity score model is misspecified, the CBPS can yield robust estimator of conditional average treatment effects. The covariates balancing method does not encounter the curse of dimensionality from Table 3–4. In other words, the covariates balancing method can still be used to estimate the propensity score when the dimensional of X is large. The CBPS estimator has a much lower RMSE than the propensity score estimator that the propensity score is standard logistic regression with X_i being the linear predictor, and the CBPS reduces RMSE at the expense of some increase in bias.

Wong and Chan (2018) proposed a method to control the covariates functional balance over a reproducing-kernel Hilbert space (Aronszajn, 1950), which is unlike regression splines, smoothing splines, it did not require preselection of the number of knots and their locations. This method can estimate the individual effect $Y_i(1) - Y_i(0)$ by the covariates functional balance weight. Thus, the conditional average treatment effect can be estimated by (7). In the following simulation, we also consider the following the DGP which is similar to Abrevaya *et al.* (2015) and Imai and Ratkovic (2014):

$$X_1 = \epsilon_1,\ X_2 = \epsilon_2,\ X_3 = \epsilon_3,\ X_4 = \epsilon_4,$$

here $\epsilon_i \sim$ iid unif$[-0.5, 0.5], i = 1, 2, 3, 4$. The potential outcome is

$$\begin{aligned} Y(1) &= X_1(1+2X_1)^2(-1+X_1)^2 + 1.5X_2 + 1.5X_3 \\ &\quad + 1.5X_4 + v_1, \\ Y(0) &= 1.5X_2 + 1.5X_3 + 1.5X_4 + v_0, \end{aligned}$$

$v_d \sim N(0, 0.25^2)$, $d = 0, 1$ and is independent of $\epsilon_i, i = 1, 2, 3, 4$. The propensity score is

$$\pi(X) = \frac{\exp\Big(0.5(X_1+X_2+X_3+X_4)\Big)}{1+\exp\Big(0.5(X_1+X_2+X_3+X_4)\Big)}.$$

We estimate CATE(x_1), $x_1 \in \{-0.4, -0.2, 0, 0.2, 0.4\}$ by the following methods:
(i). The true propensity score model is a logistic regression with X, which is denoted as True;

Table 3. Estimation the CATE by the four methods for n = 500 and 1000 repetitions with dim(x) = 4

		True			Misspecification			CPBS			CBPS1		
x_1	Paremeter	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
h $= 0.102n^{-1/13} = 0.044$													
0.4	0.46656	-0.013	0.599	0.598	2.041	0.519	2.106	0.930	0.448	1.032	0.203	0.537	0.574
0.2	0.25088	0.001	0.601	0.601	1.666	0.470	1.730	0.908	0.457	1.016	0.222	0.528	0.573
0	0	0.009	0.587	0.587	1.356	0.444	1.427	0.873	0.457	0.985	0.232	0.527	0.575
-0.2	-0.10368	0.011	0.619	0.619	1.099	0.429	1.180	0.826	0.468	0.949	0.239	0.537	0.587
-0.4	-0.03136	0.010	0.660	0.660	0.840	0.465	0.960	0.779	0.502	0.927	0.235	0.583	0.628
h $= 0.067n^{-1/9} = 0.029$													
0.4	0.46656	-0.006	0.731	0.730	2.060	0.618	2.151	0.938	0.544	1.084	0.210	0.657	0.689
0.2	0.25088	-0.009	0.775	0.774	1.655	0.581	1.754	0.899	0.582	1.071	0.212	0.677	0.709
0	0	-0.009	0.750	0.750	1.336	0.555	1.446	0.855	0.578	1.032	0.214	0.677	0.709
-0.2	-0.10368	0.009	0.794	0.794	1.096	0.545	1.224	0.822	0.596	1.015	0.237	0.686	0.725
-0.4	-0.03136	0.014	0.799	0.798	0.834	0.564	1.006	0.780	0.606	0.988	0.239	0.701	0.740

Table 4. Estimation the CATE by the four methods for n = 2000 and 1000 repetitions with dim(x) = 4

		True			Misspecification			CPBS			CBPS1		
x_1	Paremeter	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
h $= 0.102n^{-1/13} = 0.044$													
0.4	0.46656	-0.018	0.319	0.319	2.091	0.270	2.108	0.880	0.242	0.913	0.073	0.304	0.312
0.2	0.25088	-0.004	0.319	0.319	1.670	0.247	1.688	0.838	0.252	0.875	0.087	0.305	0.317
0	0	0.013	0.317	0.317	1.364	0.231	1.383	0.796	0.252	0.835	0.105	0.304	0.322
-0.2	-0.10368	0.007	0.322	.0322	1.112	0.229	1.135	0.745	0.258	0.788	0.100	0.310	0.326
-0.4	-0.03136	0.002	0.341	0.340	0.841	0.241	0.874	0.682	0.276	0.735	0.095	0.323	0.337
h $= 0.067n^{-1/9} = 0.029$													
0.4	0.46656	-0.009	0.390	0.389	2.108	0.327	2.133	0.889	0.297	0.937	0.082	0.370	0.379
0.2	0.25088	-0.001	0.408	0.408	1.667	0.308	1.695	0.839	0.318	0.897	0.089	0.389	0.399
0	0	0.010	0.403	0.403	1.357	0.284	1.386	0.790	0.314	0.851	0.101	0.385	0.398
-0.2	-0.10368	0.007	0.407	0.407	1.111	0.289	1.148	0.744	0.328	0.813	0.100	0.392	0.405
-0.4	-0.03136	0.010	0.407	0.407	0.841	0.294	0.890	0.684	0.334	0.761	0.103	0.388	0.401

Table 5. Estimation the CATE by the five methods for $n = 1000$ and 1000 repetitions with dim(x) = 4

	Parameter	True			Misspecification			CPBS			CBPS1			Kernel		
x_1		Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
							h $= 0.102n^{-1/13} = 0.060$									
0.4	0.46656	-0.018	0.440	0.440	2.076	0.396	2.113	0.904	0.336	0.965	0.129	0.403	0.423	-0.042	0.163	0.168
0.2	0.25088	-0.030	0.444	0.445	1.668	0.326	1.700	0.862	0.328	0.922	0.122	0.397	0.415	0.030	0.169	0.172
0	0	0.009	0.418	0.418	1.366	0.318	1.403	0.830	0.345	0.899	0.156	0.399	0.428	0.533	0.347	0.635
-0.2	-0.10368	0.024	0.428	0.428	1.119	0.315	1.163	0.796	0.348	0.869	0.172	0.405	0.440	0.055	0.199	0.206
-0.4	-0.03136	-0.011	0.481	0.481	0.840	0.332	0.903	0.718	0.358	0.803	0.142	0.436	0.458	-0.036	0.210	0.213
							h $= 0.067n^{-1/13} = 0.039$									
0.4	0.46656	-0.011	0.537	0.537	2.092	0.468	2.144	0.911	0.409	0.923	0.135	0.494	0.512	-0.055	0.307	0.311
0.2	0.25088	-0.044	0.583	0.584	1.656	0.405	1.705	0.852	0.417	0.861	0.110	0.510	0.521	-0.012	0.334	0.334
0	0	0.002	0.539	0.539	1.356	0.397	1.413	0.820	0.439	0.837	0.148	0.512	0.533	0.598	0.482	0.768
-0.2	-0.10368	0.017	0.555	0.555	1.111	0.395	1.180	0.789	0.439	0.813	0.164	0.522	0.547	0.013	0.377	0.377
-0.4	-0.03136	0.024	0.155	0.157	0.091	0.235	0.252	0.082	0.215	0.215	0.037	0.144	0.148	0.020	0.122	0.123

(ii). The incorrect propensity score model is a logistic regression with $X^* = (X_1^*, X_2^*, X_3^*, X_4^*)$ being the linear predictor; where $X_{1i}^* = \exp(X_{1i}/2)$, $X_{2i}^* = X_{2i}/(1 + \exp(X_{1i})) + 10$, $X_{3i}^* = (X_{1i}X_{3i}/25 + 0.6)^3$, $X_{4i}^* = (X_{1i} + X_{4i} + 20)^2$, which is denoted as Misspecification;

(iii). The just-identified CBPS estimation with the covariates balancing moment conditions with respect to X_i^*, which is denoted as CBPS;

(iv). The just-identified CBPS estimation with the covariates balancing moment conditions with respect to X_i, which is denoted as CBPS1;

(v). The method is proposed by Raymond *et al.* (2018), which is denoted as Kernel. The simulation result is reported in Table 5 for the number of Monte Carlo repetitions is 1000 and $n = 1000$. We find that the covariates balance method performs well than the parameter method when the propensity score model is misspecified, and the RMSE is small for the kernel method, however the kernel method reduces RMSE at the cost of calculation, where the CBPS and kernel method take 0.4 and 14 seconds to run once, respectively. As far as the computation is considered parametric methods are computationally faster than the non-parametric methods. In general, the covariates balance method is superior to the parametric method when the propensity score model is misspecified, computationally faster than the kernel methods, and is robust for the propensity score models.

4. EMPIRICAL STUDIES

Many studies showed that smoking and drinking is generally regarded as one of the major modifiable risk factors for low birth weight, such as Almond *et al.*, (2005), Misra *et al.*, (2005), Jackson *et al.* (2007), Walker *et al.*, (2009). This dataset is considered by Cattaneo (2010) and Wang *et al.* (2019), and they considered how to estimate the effect of maternal smoking during pregnancy on low birth weight.

Many researchers attempted to estimate the effect of maternal smoking on low birth weight (Currie and Almond, 2011, Cattaneo, 2010, Wang *et al.*, 2019). In this paper, our goal is to consider how to explore the heterogeneity of this effect across subpopulations by using the values of some continuous covariates. In particular, X_1 denotes mother's age,

Table 6. Baseline characteristics of the all, smoker and nonsmoker

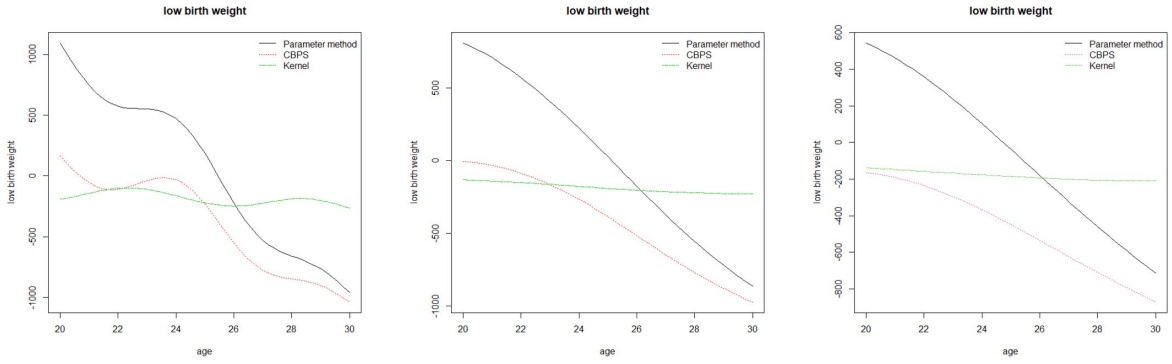
Characteristics	All (n = 4642)	Smoker ($n_1 = 864$)	Nonsmoker ($n_0 = 3778$)
lbweight	0.060 ± 0.238	0.110 ± 0.313	0.050 ± 0.216
alcohol	0.032 ± 0.177	0.091 ± 0.288	0.019 ± 0.136
mage	$26.505 \pm +5.619$	25.167 ± 5.301	26.810 ± 5.646
medu	12.690 ± 2.521	11.639 ± 2.168	12.930 ± 2.534
nprenatal	10.758 ± 3.681	9.862 ± 4.208	10.963 ± 3.518
prenatal	1.202 ± 0.508	1.308 ± 0.630	1.176 ± 0.473

that is, we are interested in estimating how the expected smoking effect changes with age, while averaging out all other confounders.

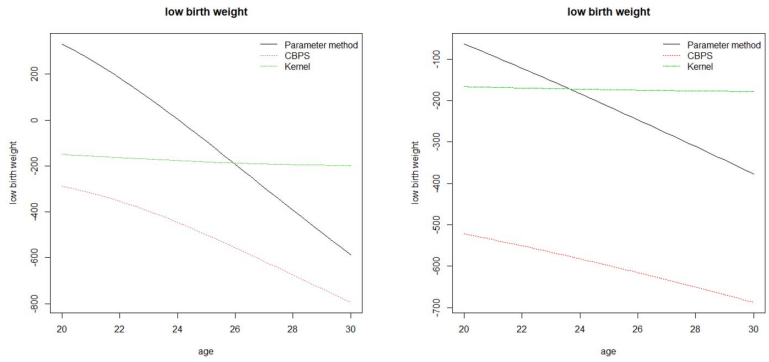
The variables are included in this analysis: lbweight (the baby's weight), mbsmoke (1 if a mother smoked while pregnant; = 0 otherwise), alcohol (= 1 if alcohol is consumed during pregnant; = 0 otherwise), mage (mother's age), medu (mother's education attainment), nprenatal (number of prenatal care visits) and prenatal (trimester of first prenatal care visit). Among the n = 4642 subjects, $n_1 = 864$ mothers smoked during pregnancy (i.e., mbsmoke = 1), while $n_0 = 3778$ mothers did not smoke during pregnancy (i.e., mbsmoke = 0). The baseline characteristics of the smoker and nonsmoker are summarized in Table 6, we can get a straightforward conclusion that the covariates are unbalanced between the treatment groups form Table 6. We estimate the CATE function over a grid of between the 20–30 years of the age distribution. We report five different CATE estimators corresponding to bandwidths $h_1 = 0.25\hat{\sigma}, 0.5\hat{\sigma}, 0.8\hat{\sigma}, 1\hat{\sigma}, 2\hat{\sigma}$ for the three methods, where $\hat{\sigma}$ is the sample standard deviation of X_1 (mother's age).

The results based on various methods are shown in figure 1(a)–1(e) for five different bandwidths.

From the figure 1(a)–1(e), we obtained that the propensity score of parameters methods, with the increase of age, and the effect of smoking on low birth weight changes from positive to negative. The reason may be that the propensity score model is not a logistic model with linear prediction of observed covariates, but the parametric method is a logistic model with linear prediction of observed covariates.



(a) h=025σ (b) h=0.5σ (c) h=0.75σ



(d) h=σ (e) h=2σ

Figure 1. The effect of mothers' smoking on infant birth weight at a given age, with h = 0.25σ, 0.5σ, 0.75σ, σ, 2σ.

Therefore, it leads to inconsistent results with the existing literature. The CBPS and Kernel methods with the increase in age, and the influence of smoking on low birth children are very significant. Therefore, it indicates that maternal smoking has a negative influence on infant birth weight, and that this effect increased with maternal age, this impact becomes more prevalent. Mothers are not advised to smoke while pregnant.

5. CONCLUSION

In this paper, we consider the estimation of the conditional average treatment effect, a functional parameter which is considered to describe the variation of average treatment effect conditional on some covariates. We propose the covariates balance method to estimate the propensity score model, then the nonparametric method is used to estimate the conditional average treatment effect. The covariates balance method performs well than the parametric method when the propensity score model is misspecified. That's because when the propensity score model is misspecified, the covariates balance method still considers the propensity score of covariates balance property. The parametric method that the propensity score model is estimated by maximum likelihood method, the maximum likelihood method does not satisfy covariates balance property. Therefore, the covariates balance method utilizes more information than MLE method. The covariates balance method has a smaller RMSE than the parametric method when the propensity score model is correctly specified, and has less computational cost than the kernel method. The covariates balance method is robust for the misspecification of propensity score models. The true propensity score estimator is a maximum likelihood method estimator, while the CBPS estimator is a moment estimator. In general, the maximum likelihood method estimator makes stricter assumptions than the moment estimator, and and the proposed method is not affected from the dimensional changes of X. The proposed estimator is consistent and asymptotically normal under some regularity conditions. Finally, we applied the proposed method to estimate the effect of maternal smoking on low birth weight infants given the age of mothers. The results showed that maternal smoking has a negative influence on infant birth weight, and that this effect increased with maternal age.

In the future, we will consider that the covariates balance methods to estimate propensity score models was applied to multiple treatments, panel data and the presence of unobserved confounding frameworks.

APPENDIX: PROOF

Let $\{W_i = (X_i, Y_i, D_i)\}_{i=1}^n$ are independent and identically distributed. We consider the case $\ell = 1$, and we can write the proposed CATE estimator as

$$\sqrt{nh}(\hat{\tau}(x_1) - \tau(x_1)) = \frac{\frac{1}{\sqrt{nh}} \sum_{i=1}^n K\Big(\frac{X_{1i}-x_1}{h}\Big)}{\frac{1}{nh} \sum_{i=1}^n K\Big(\frac{X_{1i}-x_1}{h}\Big)} \times \Big[\psi(W_i, \hat{\pi}(X_i)) - \tau(x_1)\Big].$$

As

$$\frac{1}{nh} \sum_{i=1}^n K\Big(\frac{X_{1i}-x_1}{h}\Big) \xrightarrow{p} f_1(x_1),$$

under the assumptions, we have

$$\begin{aligned}
&\frac{1}{\sqrt{nh}} \sum_{i=1}^n K\Big(\frac{X_{1i}-x_1}{h}\Big)\Big[\psi(W_i, \hat{\pi}(X_i)) - \tau(x_1)\Big] \\
=&\frac{1}{\sqrt{nh}} \sum_{i=1}^n K\Big(\frac{X_{1i}-x_1}{h}\Big)\Big[\psi(W_i, \pi(X_i)) - \tau(x_1)\Big] \\
=&\frac{1}{\sqrt{nh}} \sum_{i=1}^n K\Big(\frac{X_{1i}-x_1}{h}\Big)\Big[\psi_\pi(W_i, \pi^*(X_i))\Big] \\
&\times (\hat{\pi}(X_i) - \pi(X_i))\Big),
\end{aligned}$$

where $\pi^*(X_i)$ is between $\hat{\pi}(X_i)$ and $\pi(X_i)$ and

$$\psi_\pi(W_i, \pi^*(X_i)) = -\frac{D_i Y_i}{\pi^{*2}(X_i)} - \frac{(1-D_i)Y_i}{(1-\pi^*(X_i))^2}.$$

The second term is considered as

$$\begin{aligned}
&\frac{1}{\sqrt{nh}}\Bigg|\sum_{i=1}^n K\Big(\frac{X_{1i}-x_1}{h}\Big)\Big[\psi_\pi(W_i, \pi^*(X_i))\Big] \\
&\quad\times (\hat{\pi}(X_i) - \pi(X_i))\Big)\Bigg| \\
&\quad\le \sqrt{nh} \sup_{x\in\mathcal{X}} |\hat{\pi}(x) - \pi(x)| \cdot \\
&\quad\frac{1}{\sqrt{nh}}\Bigg|\sum_{i=1}^n K\Big(\frac{X_{1i}-x_1}{h}\Big)\Big[\psi_\pi(W_i, \pi^*(X_i))\Big]\Bigg|
\end{aligned}$$

where the first factor is $o_p(1)$ by the Assumption 3 and the second term is $O_p(1)$ by the Assumption 4, 5, 6. Therefore,

$$\begin{aligned}
&\frac{1}{\sqrt{nh}} \sum_{i=1}^n K\Big(\frac{X_{1i}-x_1}{h}\Big)\Big[\psi(W_i, \hat{\pi}(X_i)) - \tau(x_1)\Big] \\
&\quad= \frac{1}{\sqrt{nh}} \sum_{i=1}^n K\Big(\frac{X_{1i}-x_1}{h}\Big)\Big[\psi(W_i, \pi(X_i)) - \tau(x_1)\Big] \\
&\quad+ o_p(1).
\end{aligned}$$

Proof (b), we rewrite

$$\begin{aligned}
&\sqrt{nh}\Big(\hat{\tau}(x_1) - \tau(x_1)\Big) \\
=&\frac{1}{\sqrt{nh}} \sum_{i=1}^n K\Big(\frac{X_{1i}-x_1}{h}\Big)\Big[\psi(W_i, \pi(X_i)) - \tau(x_1)\Big] + o_p(1) \\
&\quad (8) \\
=&\frac{1}{\sqrt{nh}} \frac{1}{f_1(x_1)} \sum_{i=1}^n \Big[\psi(W_i, \pi(X_i)) - \tau(X_{1i})\Big] K\Big(\frac{X_{1i}-x_1}{h}\Big) \\
&\quad (9) \\
&+ \frac{1}{\sqrt{nh}} \frac{1}{f_1(x_1)} \sum_{i=1}^n \Big[\tau(X_{1i}) - \tau(x_1)\Big] K\Big(\frac{X_{1i}-x_1}{h}\Big) \\
&+ o_p(1)
\end{aligned}$$

It can be easily proved that

$$E\Big[\Big(\psi(W_i, \pi(X_i)) - \tau(X_{1i})\Big) K_{in}\Big] = 0,$$

where we write $K_{in} = K\Big(\frac{X_{1i}-x_1}{h}\Big)$ to make it explicit that this quantity depends on n through h. For each n, the random variables $\Big\{\Big[\psi(W_i, \pi(X_i)) - \tau(X_{1i})\Big] K_{in}\Big\}_{i=1}^n$ are independent and one can apply Lyapunov's CLT for triangular arrays to (8) to obtain the asymptotic distribution (Chung 1974, p. 209), namely

$$\begin{aligned}
&\frac{1}{\sqrt{nh}} \frac{1}{f_1(x_1)} \sum_{i=1}^n \Big[\psi(W_i, \pi(X_i)) - \tau(X_{1i})\Big] K\Big(\frac{X_{1i}-x_1}{h}\Big) \\
&\qquad\xrightarrow{d} N(0, \sigma^2),
\end{aligned}$$

where

$$\sigma^2 = \lim_{n\to\infty} E\Bigg[h^{-1}\Big(\psi(W_i, \pi(X_i)) - \tau(X_{1i})\Big)^2 \frac{K^2\Big(\frac{X_{1i}-x_1}{h}\Big)}{f_1^2(x_1)}\Bigg].$$

Let $\sigma^2(u) = E\Big[\Big(\psi(W_i, \pi(X_i)) - \tau(X_{1i})\Big)^2 \mid X_1 = u\Big]$,

$$\begin{aligned}
\sigma^2 &= \lim_{n\to\infty} E\Bigg[h^{-1}\Big(\psi(W_i, \pi(X_i)) - \tau(X_{1i})\Big)^2 \frac{K^2\Big(\frac{X_{1i}-x_1}{h}\Big)}{f_1^2(x_1)}\Bigg] \\
&= \lim_{n\to\infty} \int E\Bigg[h^{-1}\Big(\psi(W_i, \pi(X_i)) - \tau(X_{1i})\Big)^2 \\
&\quad\times K^2\Big(\frac{X_{1i}-x_1}{h}\Big)/f_1^2(x_1) \mid X_1 = u\Bigg] f_1(u)du \\
&= \lim_{n\to\infty} \int h^{-1}\sigma^2(u) K^2\Big(\frac{u-x_1}{h}\Big)/f_1^2(x_1) f_1(u)du \\
&= \lim_{n\to\infty} \int \sigma^2(x_1 + hu) K^2(u)/f_1^2(x_1) f_1(x_1 + uh)du \\
&= \sigma^2(x_1) \int K^2(u)du f_1(x_1)/f_1^2(x_1) = \frac{\sigma^2(x_1)\int K^2(u)du}{f_1(x_1)}.
\end{aligned}$$

Next, the bias term (9) can be written as

$$\begin{aligned} B_n &= \frac{1}{\sqrt{nh}}\frac{1}{f_1(x_1)}\sum_{i=1}^{n}\Big[\tau(X_{1i})-\tau(x_1)\Big]K\Big(\frac{X_{1i}-x_1}{h}\Big) \\ &= \frac{1}{f_1(x_1)}\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\Big[\tau(X_{1i})-\tau(x_1)\Big]K\Big(\frac{X_{1i}-x_1}{h}\Big) \\ &= \frac{1}{f_1(x_1)}\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\epsilon_i \\ &= \frac{1}{f_1(x_1)}b_n. \end{aligned}$$

where $\epsilon_i = \Big(\tau(X_{1i})-\tau(x_1)\Big)K\Big(\frac{X_{1i}-x_1}{h}\Big)$, $b_n = \frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\epsilon_i$. We should show that b_n converges to zero in probability. For this purpose, we need to show that $E[b_n]\to 0$, $\mathrm{Var}b_n \to 0$ as $n\to\infty$.

First,

$$\begin{aligned} E[b_n] =& (nh)^{-1/2}\sum_{i=1}^{n}E[\epsilon_i] = (nh)^{1/2}E[h^{-1}\epsilon] \\ =& (nh)^{1/2}\int\Big(\tau(x_1+hu)-\tau(x_1)\Big) \\ & \times K(u)f_1(x_1+hu)du \\ =& (nh)^{1/2}\Big(h^2\mu_2\Big[\tau^{(1)}f_1^{(1)}+\tau^{(2)}f_1/2\Big]\Big)+o(h^4) \\ & \to 0, \text{ for } (nh)^{1/2}h^2\to 0. \end{aligned}$$

where $\tau^{(i)}$ and $f_1^{(i)}$ denote i-th derivative, $\mu_2 = \int u^2K(u)du$. Because the $\{\epsilon_i\}_{i=1}^n$ are i.i.d. we have

$$\begin{aligned} \mathrm{Var}(b_n) &= E[b-n^2]-E^2[b_n] \\ &= \frac{1}{nh}\sum_{i=1}^{n}E[\epsilon_i^2]-\frac{1}{nh}(\sum_{i=1}^{n}E[\epsilon_i]) \\ &= \frac{E[\epsilon^2]}{h}-\frac{n}{h}E^2[\epsilon]. \end{aligned}$$

As $n\to\infty$,

$$\begin{aligned} \frac{E[\epsilon^2]}{h} =& h^{-1}E\Big[K^2\Big(\frac{X_{1i}-x_1}{h}\Big)(\tau(X_{1i})-\tau(x_1))\Big] \\ =& \int\Big(\tau(x_1+hu)-\tau(x_1)\Big)^2K^2(u)f_1(x_1+hu)du\to 0, \end{aligned}$$

and

$$\begin{aligned} \frac{nE^2[\epsilon]}{h} =& nh\left[\int\Big(\tau(x_1+hu)-\tau(x_1)\Big)K(u)f_1(x_1+hu)du\right]^2 \\ & \to 0, \end{aligned}$$

Thus $\mathrm{Var}(b_n)\to 0$, $b_n \xrightarrow{p} E[b_n]\to 0$, it implying that $b_n = 0_p(1)$, Namely

$$\frac{1}{\sqrt{nh}}\frac{1}{f_1(x_1)}\sum_{i=1}^{n}\Big[\tau(X_{1i})-\tau(x_1)\Big]K\Big(\frac{X_{1i}-x_1}{h}\Big) = o_p(1).$$

Therefore has no bearing on the limit distribution, namely

$$\sqrt{nh}(\hat{\tau}(x_1)-\tau(x_1)) \xrightarrow{d} N\left(0, \frac{\|K\|_2^2\sigma_\psi^2(x_1)}{f_1(x_1)}\right).$$

Pagan and Ullah (1999) gave the discussion about higher order kernels, the proof of the case $\ell > 1$ is omitted here.

ACKNOWLEDGMENT

The authors thank referees, editors, and associate editors for their constructive comments and suggestions, which improve the manuscript significantly. The work is supported by the National Natural Science Foundation of China (12101545), the scientific research project of colleges and universities in Inner Mongolia Autonomous Region (No:NJSY21569) and the startup fund project of Introducing High-level Talents in Inner Mongolia Normal University (NO:2020YJRC055).

Received 16 October 2020

REFERENCES

[1] Abrevaya, J., Hsu, Y. C., and Lieli, R. P. (2015). Estimating Conditional Average Treatment Effects. *Journal of Business and Economic Statistics* **33**(4) 485–505. MR3416596

[2] Almond, D., Chay, K. Y., and Lee, D. S. (2005). The costs of low birth weight. *Q. J. Econ.* **120** 1031–1083.

[3] Angrist, J. D. and Hahn, J. Y. (1999). When to Control for Covariates? Panel-Asymptotic Results for Estimates of Treatment Effects. *NBER technical working paper no. 241.*

[4] Aronszajn, N. (1950). Theory of reproducing kernels. *Trans. Am. Math. Soc.* **68** 337–404. MR0051437

[5] Cattaneo, M. D. (2010). Efficient semiparametric estimation of multi-valued treatment effects under ignorability. *J. Econometrics* **155**(2) 138–154. MR2607191

[6] Chung, K. L. (1974). *A Course in Probability Theory*, 2nd ed. Wiley, New York. MR0346858

[7] Currie, J. and Almond, D. (2011). Human capital development before age five. In *Handbook of labor economics. 4*, 1315–1486.

[8] Hahn, J. (1998). On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects. *Econometrica* **66** 315–331. MR1612242

[9] Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50** 1029–1054. MR0666123

[10] Heckman, J., Ichimura, H., and Todd, P. (1997). Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program. *Review of Economic Studies* **64** 605–654. MR1623713

[11] Heckman, J., Ichimura, H., and Todd, P. (1998). Matching as an Econometric Evaluations Estimator. *Review of Economic Studies* **65** 261–294. MR1623713

[12] Heckman, J. and Vytlacil, E. (2005). Structural Equations, Treatment Effects, and Econometric Policy Evaluation. *Econometrica* **73** 669–738. MR2135141

[13] Hirano, K., Imbens, G. W., and Ridder, G. (2003). Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score. *Econometrica* **71** 1161–1189. MR1995826
[14] Hsu, Y. C. (2012). Consistent Tests for Conditional Treatment Effects. *The Econometrics Journal* **20**(1) 1–22. MR3636959
[15] Imai, K. and Ratkovic, M. (2014). Covariate Balancing Propensity Score. *Journal of The Royal Statistical Society Series B-Statistical Methodology* **76**(1) 243–263. MR3153941
[16] Jackson, D., Batiste, E., and Rendallmkosi, K. M. (2007). Effect of smoking and alcohol use during pregnancy on the occurrence of low birthweight in a farming region in South Africa. *Paediatric and Perinatal Epidemiology* **21**(5) 432–440.
[17] Kang, J. D. and Schafer, J. L. (2007). Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data (with discussions). *Statistical Science* **22** 523–539. MR2420458
[18] Khan, S. and Tamer, E. (2010). Irregular Identification, Support Conditions, and Inverse Weight Estimation. *Econometrica* **6** 2021–2042. MR2768989
[19] Lee, S. and Whang, J.Y. (2009). Nonparametric Tests of Conditional Treatment Effects. *Cowles Foundation Discussion Papers 1740, Cowles Foundation, Yale Universit.*
[20] Li, F., Morgan, K. L., and Zaslavsky, A. M. (2018). Balancing Covariates via Propensity Score Weighting. *Journal of the American Statistical Association* **113**(521) 390–400. MR3803473
[21] MaCurdy, T., Chen, X., and Hong, H. (2011). Flexible Estimation of Treatment Effect Parameters. *American Economic Review* **16** 544–551.
[22] Misra, D. P., Astone, N., and Lynch, C. D. (2005). Maternal smoking and birth weight: interaction with parity and mother's own in utero exposure to smoking. *Epidemiology* **16**(3) 288–293.
[23] Pagan, A. and Ullah, A. (1999). *Nonparametric Econometrics*. Cambridge University Press, Cambridge, UK. MR1699703
[24] Robins, J. M. and Ritov, Y. A. (1997). Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semiparametric models. *Statistics in Medicine* **16**(3) 285–319.
[25] Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**(1) 41–55. MR0742974
[26] Smith, A. J. and Todd, E. P. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics* **125** 305–353. MR2143379
[27] Walker, M. B., Tekin, E., and Wallace, S. (2009). Teen Smoking and Birth Outcomes. *Southern Economic Journal* **75** 892–907.
[28] Wang, J., Gao, W., and Tang, M. (2019). Estimation of treatment effects for heterogeneous matched pairs data with probit models. *Scandinavian Journal of Statistics* **46**(2) 575–594. MR3948568
[29] Wong, R. K. and Chan, K. C. G. (2018). Kernel-based covariate functional balancing for observational studies. *Biometrika* **105** 199–213. MR3768874
[30] Zhao, Q. (2019). Covariate balancing propensity score by tailored loss functions *Annals of Statistics* **47**(2) 965–993. MR3909957

Jun Wang
Yunnan Key Laboratory of Statistical Modeling and Data Analysis
Yunnan University
Kunming 650091
People's Republic of China
E-mail address: wangj198@nenu.edu.cn

Changbiao Liu
College of Mathematical Sciences
Inner Mongolia Normal University
Hohhot 010022
People's Republic of China
E-mail address: liuchang6707@163.com