Computational Statistics and Data Analysis 203 (2025) 108079

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

A comparative analysis of different adjustment sets using propensity score based estimators

Shanshan Luo [a], Jiaqi Min [a], Wei Li [b], Xueli Wang [a,*], Zhi Geng [a]

[a] School of Mathematics and Statistics, Beijing Technology and Business University, 100048, Beijing, China
[b] Center for Applied Statistics and School of Statistics, Renmin University of China, 100872, Beijing, China

ARTICLE INFO

Keywords:
Confounders
Observational studies
Propensity score
Regression adjustment
Variable selection

ABSTRACT

Propensity score based estimators are commonly employed in observational studies to address baseline confounders, without explicitly modeling their association with the outcome. In this paper, to fully leverage these estimators, we consider a series of regression models for improving estimation efficiency. The proposed estimators rely solely on a properly modeled propensity score and do not require the correct specification of outcome models. In addition, we consider a comparative analysis by applying the proposed estimators to four different adjustment sets, each consisting of background covariates. The theoretical results imply that incorporating predictive covariates into both propensity score and regression model demonstrates the lowest asymptotic variance. However, including instrumental variables in the propensity score may decrease the estimation efficiency of the proposed estimators. To evaluate the performance of the proposed estimators, we conduct simulation studies and provide a real data example.

1. Introduction

Recent methodological advancements have greatly improved the evaluation of average treatment effects in observational studies, particularly when enriched covariates are available. Propensity score based estimators, initially formalized by Rosenbaum and Rubin (1983), are widely utilized for causal inference in observational studies. These methods involve estimating a propensity score model for each participant, representing the conditional probability of being in the treatment group given the covariates. The propensity score serves as a balancing score, facilitating the adjustment of covariate distributions between the treatment and control groups. Once the propensity score is estimated, various methods can be employed to estimate the average treatment effects, including stratification, matching, weighting, and covariate adjustment (Lunceford and Davidian, 2004; Vansteelandt and Daniel, 2014; Abadie and Imbens, 2016; Liu et al., 2016; Li et al., 2018). Among these, weighted estimators based on the propensity score, such as inverse probability weighted (IPW) estimators (Horowitz and Manski, 1995) and Hájek stabilized weighted estimators (Hájek, 1971), are the most widely used in practice.

In observational studies, covariates are typically classified into three types: true confounders, instrumental variables, and outcome predictors. Each of these covariate types plays a distinct role in practice, as detailed in Section 2. Proper covariate selection for the propensity score is important, as the inclusion of inadequate covariates can significantly harm the estimation efficiency of the average treatment effect (Brookhart et al., 2006; Franklin et al., 2015). For instance, when employing the IPW estimator, incorporating true

* Corresponding author.
E-mail address: xlwang@btbu.edu.cn (X. Wang).

https://doi.org/10.1016/j.csda.2024.108079
Received 31 May 2024; Received in revised form 2 October 2024; Accepted 19 October 2024

Available online 21 October 2024

confounders and instrumental variables in the propensity score can yield an unbiased estimation but may potentially reduce efficiency (Hahn, 2004). To address this issue, several methods for covariate selection have been proposed to mitigate efficiency loss (De Luna et al., 2011; Patrick et al., 2011). Conversely, incorporating true confounders and outcome predictors in the propensity score model for the weighted estimator produces an unbiased estimation and typically achieves optimal asymptotic performance (Craycroft et al., 2020).

Parametric methods are often preferred for estimating the propensity score, as nonparametric methods may suffer from the curse of dimensionality. However, the exploration of regression models using different types of covariates to improve efficiency remains limited. Regression adjustment strategies are typically considered model-assisted methods, offering valid inference with minimal statistical assumptions required for unadjusted estimation, as demonstrated in randomized experiments (Neyman, 1923; Fisher, 1966; Lin, 2013; Li and Ding, 2017; Bloniarz et al., 2016; Ting et al., 2023). Similarly, in observational studies, it is desirable that covariate adjustment does not introduce additional assumptions beyond those required by classical propensity score based estimators (Horvitz and Thompson, 1952; Hájek, 1971). When dealing with diverse pretreatment covariates, selecting appropriate adjustment sets for propensity score models or regression models becomes crucial. The selection process may involve adjusting for either true confounders alone or adjusting for both true confounders and additional outcome predictors. However, the question of which covariates should be adjusted to achieve optimal estimation, even with a known propensity score, remains unanswered (De Luna et al., 2011; Craycroft et al., 2020).

This paper introduces regression models to classical propensity score based estimators to address the aforementioned issues and further employs them for comparative analysis across four types of adjustment sets. These proposed estimators are linear modifications of classical weighted estimators and provide improved estimation without requiring correct outcome model specification. Inspired by the covariate adjustment methods commonly used in randomized experiments (Neyman, 1923; Fisher, 1966; Lin, 2013), the proposed estimator demonstrates optimal performance, exhibiting the smallest variance within its respective class and outperforming the benchmark weighted estimator across all adjustment sets. The findings suggest that when the propensity score is known, incorporating covariates other than instrumental variables in the proposed estimators yields the lowest variance. Conversely, including instrumental variables in propensity score models may increase estimation variance. Moreover, the paper explores the potential benefits of using estimated propensity scores, further enhancing estimation efficiency.

The rest of this paper is organized as follows. In Section 2, we introduce the notation and assumptions. Section 3 presents the theoretical results of the improved estimators with a known propensity score, while Section 4 covers the theoretical results of the improved estimators with an estimated propensity score. In Section 5, we evaluate the finite sample performance of the proposed estimators via simulations. In Section 6, we apply the proposed methods to estimate the causal effect of the number of children in a household on the household's payment. We conclude with a brief discussion in Section 7.

2. Notation and assumptions

Assuming that there are n individuals who are independently and identically sampled from a superpopulation of interest. For each individual i, let Z_i denote the treatment, Y_i denote the outcome variable, and X_i denote the baseline covariates. Without loss of generality, we omit the subscript i in the following text. We estimate the causal effect of a treatment Z on an outcome Y in the presence of confounding by the observed covariates X. To facilitate the discussion, we assume that the covariates have been centered around zero means, namely, $\mathbb{E}(X) = 0$. For non-centered covariates, it is recommended to center them by subtracting their mean. The potential outcomes under treatment and control are denoted as Y_1 and Y_0, respectively. The observed outcome Y can be expressed as a combination of these potential outcomes: $Y = ZY_1 + (1 - Z)Y_0$. We make the stable unit treatment value assumption (SUTVA), which states that there is only one version of potential outcomes for each individual, and there is no interference between units (Neyman, 1923; Rubin, 1974, 1990). Let τ_z denote the expectation of the potential outcome Y_z; we are interested in the average treatment effect (ATE) defined as $\tau = \mathbb{E}(Y_1 - Y_0)$.

A fundamental problem in causal inference is that we can never observe both potential outcomes for a unit simultaneously. To establish the identifiability of the ATE, the unconfoundedness or ignorability assumption is commonly employed (Rosenbaum and Rubin, 1983). This assumption states that the treatment assignment Z is independent of the potential outcomes (Y_0, Y_1) given the covariates X, thereby eliminating any potential unobserved confounding between the treatment and outcome. In a randomized experiment, the ignorability assumption naturally holds, as Z is independent of (Y_0, Y_1, X).

In practice, it is of interest to classify the observed covariates X into three distinct types: (i) true confounders, denoted as C, which are associated with both the treatment and the outcome; (ii) instrumental variables, denoted as A, which are correlated with the treatment but not directly associated with the outcome, except through the treatment Z and true confounders C; and (iii) outcome predictors, denoted as W, which are correlated with the outcome but not directly associated with the treatment, except through the true confounders C. The aforementioned conditional independence can be expressed as follows:

Assumption 1. $(A, Z) \perp\!\!\!\perp (W, Y_z) \mid C$.

Assumption 1 states that the common covariates C sufficiently account for the potential confounding between (A, Z) and (W, Y_z). Fig. 1 illustrates Assumption 1 graphically. Assumption 1 suggests four possible adjustment sets in practice: (i) $V_c = C$, (ii) $V_{cw} = (C, W)$, (iii) $V_{ac} = (A, C)$, and (iv) $V_{acw} = (A, C, W)$. The following lemma emphasizes their ability to address potential confounding factors.

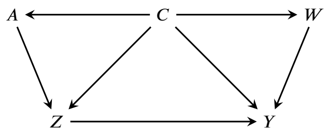

Fig. 1. Causal DAG with different types of covariates.

Lemma 1. *Given Assumption 1, for each adjustment set* $V_k \in \{V_c, V_{cw}, V_{ac}, V_{acw}\}$*, we have* $Z \perp\!\!\!\perp (Y_0, Y_1) \mid V_k$.

Lemma 1 implies that, under Assumption 1, controlling for the true confounders C is sufficient to address all potential confounding factors. Moreover, including instrumental variables A and outcome predictors W in any combination for additional adjustment would also satisfy the ignorability assumption (Craycroft et al., 2020).

The propensity score, denoted as $e(V_k) = \mathrm{pr}(Z = 1 \mid V_k)$, represents the conditional probability of treatment assignment given a subject's covariates, where V_k represents the adjustment set defined before Lemma 1. Moreover, we require adequate overlap between the treatment and control covariate distributions, quantified by the following positivity assumption.

Assumption 2. $0 < e(V_k) < 1$*, where* $V_k \in \{V_c, V_{cw}, V_{ac}, V_{acw}\}$.

The positivity assumption is crucial to avoid small quantities in the denominators of the weighted estimators (Rosenbaum and Rubin, 1983). In practice, to mitigate the curse of dimensionality, a common approach is to estimate the propensity score $e(V_k)$ by employing a parametric model $e(V_k; \gamma_k)$, where γ_k represents the corresponding nuisance parameter.

Adjusting for pretreatment covariates can improve estimation efficiency, particularly when at least one observed covariate predicts the outcome (Lin, 2013). The existence of multiple adjustment sets, as indicated in Lemma 1, provides an opportunity to further increase estimation efficiency. The nuisance parameters in the working model for the adjustment set V_k are defined as $\theta_k = (\theta_{k,0}^{\mathrm{T}}, \theta_{k,1}^{\mathrm{T}})^{\mathrm{T}}$, where $\theta_{k,z}$ represents the coefficient vector associated with the treatment assignment $Z = z$. Our goal is to estimate the ATE using two linear modified estimators, which can be considered as regression adjustment of the classic weighted estimators:

$$\tau_k^{\mathrm{ipw}}(\theta_k, \gamma_k) = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{Z_i(Y_i - V_{k,i}^{\mathrm{T}}\theta_{k,1})}{e(V_{k,i}; \gamma_k)}\right\} - \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{(1 - Z_i)(Y_i - V_{k,i}^{\mathrm{T}}\theta_{k,0})}{1 - e(V_{k,i}; \gamma_k)}\right\}, \tag{1}$$

$$\tau_k^{\mathrm{haj}}(\theta_k, \gamma_k) = \frac{1}{n_{k,1}^{\mathrm{haj}}}\sum_{i=1}^{n}\left\{\frac{Z_i(Y_i - V_{k,i}^{\mathrm{T}}\theta_{k,1})}{e(V_{k,i}; \gamma_k)}\right\} - \frac{1}{n_{k,0}^{\mathrm{haj}}}\sum_{i=1}^{n}\left\{\frac{(1 - Z_i)(Y_i - V_{k,i}^{\mathrm{T}}\theta_{k,0})}{1 - e(V_{k,i}; \gamma_k)}\right\}, \tag{2}$$

where $n_{k,1}^{\mathrm{haj}} = \sum_{i=1}^{n} Z_i / e(V_{k,i}; \gamma_k)$ and $n_{k,0}^{\mathrm{haj}} = \sum_{i=1}^{n}(1 - Z_i)/\{1 - e(V_{k,i}; \gamma_k)\}$. Both $n_{k,1}^{\mathrm{haj}}$ and $n_{k,0}^{\mathrm{haj}}$ depend on the nuisance parameter γ_k. The superscripts "ipw" and "hjk" indicate quantities associated with the IPW estimator and the Hájek estimator, respectively. The Hájek estimator in (2) replaces the denominator n in the IPW estimator (1) with the sum of the inverses of the sampling probabilities, which tends to reduce the variance of the estimator in practice (Horvitz and Thompson, 1952; Hájek, 1971). It is worth noting that centering covariates with zero means is crucial for ensuring the consistency of the two estimators discussed above. Additionally, when $\theta_k = 0$, both estimators (1) and (2) simplify to the classic IPW and Hájek estimators, respectively, which serve as benchmark estimators throughout this paper.

The proposed estimators (1) and (2) can be viewed as model-assisted estimators. They leverage the propensity score to derive more efficient estimators but do not require the linear working model to be correctly specified. When using estimators (1) and (2), we require the regression adjustment technique to be compatible with Assumption 1, but we do not impose any restrictions on the propensity score model. This is because, throughout this paper, we assume a properly modeled propensity score, which allows us to retain flexibility by including adjustments for all relevant covariates. This strategy is particularly reasonable in practical scenarios, especially in cases of stratified randomization designs where the treatment allocation mechanism is well-defined. Nonetheless, we avoid assuming any known model between the outcome and covariates. Using incorrect models may lead to extreme estimates in practice. Therefore, we utilize the available knowledge of the conditional independence from Assumption 1 whenever possible. In the outcome model, when both A and C are included in the set of covariates, we set the coefficients related to the instrument variables A to zero. Further details on the components of θ_k in other cases are provided in Sections S3.1 and S4 of the Supplementary Material.

3. Comparative analysis with known propensity score

In this section, we utilize a fully known propensity score model $e(V_k; \gamma_k^{\star})$, which does not involve any nuisance parameters, to derive the optimal estimator for different adjustment sets (Pan and Zhao, 2021). In the next section, we will relax this restriction and consider the case where the nuisance parameter γ_k is estimated. Additionally, we will compare the asymptotic variances by specifically considering the optimal estimators within the classes (1) and (2). For simplicity, we denote $\tau_k^{\mathrm{ipw}}(\theta_k, \gamma_k^{\star})$ and $\tau_k^{\mathrm{haj}}(\theta_k, \gamma_k^{\star})$ as $\tau_k^{\mathrm{ipw}}(\theta_k)$ and $\tau_k^{\mathrm{haj}}(\theta_k)$, respectively. It is important to note that the choice of the parameter θ_k can significantly affect the performance of the proposed estimators. The asymptotic properties of these estimators can be derived using estimating equation theory (Stefanski and Boos, 2002), and we summarize these results in the following proposition:

Proposition 1. *Assuming that Assumptions 1-2 hold, the propensity score is known, and the regularity conditions in Van der Vaart (2000) hold, the proposed estimators are consistent and asymptotically normal for any θ_k. Specifically,*

$$\sqrt{n}\{\tau_k^{\mathrm{ipw}}(\theta_k)-\tau\}\xrightarrow{d} N(0,\Sigma_k^{\mathrm{ipw1}}(\theta_k)),\ \sqrt{n}\{\tau_k^{\mathrm{haj}}(\theta_k)-\tau\}\xrightarrow{d} N(0,\Sigma_k^{\mathrm{haj1}}(\theta_k)),$$

where

$$\Sigma_k^{\mathrm{ipw1}}(\theta_k)=\mathbb{E}\left\{\frac{(1-Z)(Y_0-V_k^{\mathrm{T}}\theta_{k,0})}{1-e(V_k;\gamma_k^\star)}\right\}^2+\mathbb{E}\left\{\frac{Z(Y_1-V_k^{\mathrm{T}}\theta_{k,1})}{e(V_k;\gamma_k^\star)}\right\}^2-\left(\tau_1-\tau_0\right)^2, \tag{3}$$

$$\Sigma_k^{\mathrm{haj1}}(\theta_k)=\mathbb{E}\left\{\frac{(1-Z)(Y_0-V_k^{\mathrm{T}}\theta_{k,0}-\tau_0)}{1-e(V_k;\gamma_k^\star)}\right\}^2+\mathbb{E}\left\{\frac{Z(Y_1-V_k^{\mathrm{T}}\theta_{k,1}-\tau_1)}{e(V_k;\gamma_k^\star)}\right\}^2. \tag{4}$$

If $\hat\theta_k-\theta_k=o_p(1)$, we have

$$\sqrt{n}\left\{\tau_k^{\mathrm{ipw}}(\hat\theta_k)-\tau_k^{\mathrm{ipw}}(\theta_k)\right\}=o_p(1),\ \sqrt{n}\left\{\tau_k^{\mathrm{haj}}(\hat\theta_k)-\tau_k^{\mathrm{haj}}(\theta_k)\right\}=o_p(1). \tag{5}$$

Proposition 1 provides closed form expressions for the asymptotic variances of the proposed estimators with a fixed coefficient θ_k or its consistent estimator $\hat\theta_k$. Additionally, it is worth mentioning that the verification of (5) also relies on the property of covariates having zero means; more details are provided in the Supplementary Material. In the special case where $\tau_1=\tau_0=0$, it can be observed that the two asymptotic variances, $\Sigma_k^{\mathrm{ipw1}}(\theta_k)$ and $\Sigma_k^{\mathrm{haj1}}(\theta_k)$, are equal.

Our goal is to find the optimal values of θ_k^{ipw1} and θ_k^{haj1} that minimize $\Sigma_k^{\mathrm{ipw1}}(\theta_k)$ and $\Sigma_k^{\mathrm{haj1}}(\theta_k)$, respectively. These optimal values lead to the smallest asymptotic variance among all proposed estimators in (1) and (2). Equation (5) demonstrates that, if $\hat\theta_k^{\mathrm{ipw1}}-\theta_k^{\mathrm{ipw1}}=o_p(1)$, the two estimators $\tau_k^{\mathrm{ipw}}(\hat\theta_k^{\mathrm{ipw1}})$ and $\tau_k^{\mathrm{ipw}}(\theta_k^{\mathrm{ipw1}})$ have the same asymptotic performance, and the difference between them is of the order $o_p(n^{-1/2})$. Therefore, treating $\hat\theta_k^{\mathrm{ipw1}}$ as if it is the true coefficient vector or pre-treatment vector does not affect its asymptotic performance.

In practice, the estimators $\hat\theta_k^{\mathrm{ipw1}}$ and $\hat\theta_k^{\mathrm{haj1}}$ can be obtained by minimizing the sample versions of the variances $\Sigma_k^{\mathrm{ipw1}}(\theta_k)$ and $\Sigma_k^{\mathrm{haj1}}(\theta_k)$, given by:

$$\hat\Sigma_k^{\mathrm{ipw1}}(\theta_k)=\frac{1}{n}\sum_{i=1}^n\left\{\frac{(1-Z_i)(Y_i-V_{k,i}^{\mathrm{T}}\theta_{k,0})}{1-e(V_{k,i};\gamma_k^\star)}\right\}^2+\frac{1}{n}\sum_{i=1}^n\left\{\frac{Z_i(Y_i-V_{k,i}^{\mathrm{T}}\theta_{k,1})}{e(V_{k,i};\gamma_k^\star)}\right\}^2-\left(\hat\tau_{k,1}^{\mathrm{ipw0}}-\hat\tau_{k,0}^{\mathrm{ipw0}}\right)^2, \tag{6}$$

$$\hat\Sigma_k^{\mathrm{haj1}}(\theta_k)=\frac{1}{n}\sum_{i=1}^n\left\{\frac{(1-Z_i)(Y_i-V_{k,i}^{\mathrm{T}}\theta_{k,0}-\hat\tau_{k,0}^{\mathrm{haj0}})}{1-e(V_k;\gamma_k^\star)}\right\}^2+\frac{1}{n}\sum_{i=1}^n\left\{\frac{Z_i(Y_i-V_{k,i}^{\mathrm{T}}\theta_{k,1}-\hat\tau_{k,1}^{\mathrm{haj0}})}{e(V_{k,i};\gamma_k^\star)}\right\}^2, \tag{7}$$

where

$$\hat\tau_{k,1}^{\mathrm{ipw0}}=\frac{1}{n}\sum_{i=1}^n\frac{Z_iY_i}{e(V_{k,i};\gamma_k^\star)},\quad \hat\tau_{k,0}^{\mathrm{ipw0}}=\frac{1}{n}\sum_{i=1}^n\frac{(1-Z_i)Y_i}{1-e(V_{k,i};\gamma_k^\star)},$$

$$\hat\tau_{k,1}^{\mathrm{haj0}}=\sum_{i=1}^n\frac{Z_iY_i}{e(V_{k,i};\gamma_k^\star)}\Big/\sum_{i=1}^n\frac{Z_i}{e(V_{k,i};\gamma_k^\star)},\quad \hat\tau_{k,0}^{\mathrm{haj0}}=\sum_{i=1}^n\frac{(1-Z_i)Y_i}{1-e(V_{k,i};\gamma_k^\star)}\Big/\sum_{i=1}^n\frac{1-Z_i}{1-e(V_{k,i};\gamma_k^\star)}.$$

It can be found that $\hat\tau_{k,1}^{\mathrm{ipw0}}-\hat\tau_{k,0}^{\mathrm{ipw0}}=\hat\tau_{k,z}^{\mathrm{ipw0}}(0,\gamma_k^\star)$ and $\hat\tau_{k,1}^{\mathrm{haj0}}-\hat\tau_{k,0}^{\mathrm{haj0}}=\hat\tau_{k,z}^{\mathrm{haj0}}(0,\gamma_k^\star)$, thus both differences are exactly the baseline IPW estimator and Hájek estimator.

To find the estimator $\hat\theta_k^{\mathrm{ipw1}}$, we minimize $\hat\Sigma_k^{\mathrm{ipw1}}(\theta_k)$ in (6) with respect to $\theta_k=(\theta_{k,1},\theta_{k,0})$ by performing sample weighted least squares regression, using known weights $1/e^2(V_k;\gamma_k^\star)$. For example, for the treatment group $Z=1$, we differentiate $\hat\Sigma_k^{\mathrm{ipw1}}(\theta_k)$ with respect to $\theta_{k,1}$ and set the derivative to zero. This yields the following estimating equation:

$$\frac{1}{n}\sum_{i=1}^n\frac{Z_iV_{k,i}(Y_i-V_{k,i}^{\mathrm{T}}\theta_{k,1})}{e^2(V_{k,i};\gamma_k^\star)}=0\quad\Rightarrow\quad\hat\theta_{k,1}^{\mathrm{ipw1}}=\left\{\frac{1}{n}\sum_{i=1}^n\frac{Z_iV_{k,i}V_{k,i}^{\mathrm{T}}}{e^2(V_{k,i};\gamma_k^\star)}\right\}^{-1}\left\{\frac{1}{n}\sum_{i=1}^n\frac{Z_iV_{k,i}Y_i}{e^2(V_{k,i};\gamma_k^\star)}\right\}. \tag{8}$$

To minimize $\hat\Sigma_k^{\mathrm{haj1}}(\theta_k)$ in (7), we first differentiate it with respect to $\theta_{k,1}$ and follow a similar procedure as in (8):

$$\frac{1}{n}\sum_{i=1}^n\frac{Z_iV_{k,i}(Y_i-\hat\tau_{k,1}^{\mathrm{haj0}}-V_{k,i}^{\mathrm{T}}\theta_{k,1})}{e^2(V_{k,i};\gamma_k^\star)}=0\quad\Rightarrow\quad\hat\theta_{k,1}^{\mathrm{haj1}}=\left\{\frac{1}{n}\sum_{i=1}^n\frac{Z_iV_{k,i}V_{k,i}^{\mathrm{T}}}{e^2(V_{k,i};\gamma_k^\star)}\right\}^{-1}\left\{\frac{1}{n}\sum_{i=1}^n\frac{Z_iV_{k,i}(Y_i-\hat\tau_{k,1}^{\mathrm{haj0}})}{e^2(V_{k,i};\gamma_k^\star)}\right\}. \tag{9}$$

Similarly, we can obtain the optimal parameter vectors $\hat{\theta}_k^{\mathrm{ipw1}} = (\hat{\theta}_{k,0}^{\mathrm{ipw1}}, \hat{\theta}_{k,1}^{\mathrm{ipw1}})$ and $\hat{\theta}_k^{\mathrm{haj1}} = (\hat{\theta}_{k,0}^{\mathrm{haj1}}, \hat{\theta}_{k,1}^{\mathrm{haj1}})$ by repeating the same estimation process for the control group $Z = 0$. Compared to (8), the solution to (9) includes an additional constant shift $\hat{\tau}_{k,1}^{\mathrm{haj0}}$, which can be interpreted as a population-weighted least squares estimator with a fixed intercept. Since $(\hat{\tau}_{k,0}^{\mathrm{haj0}}, \hat{\tau}_{k,1}^{\mathrm{haj0}})$ are consistent estimators for (τ_0, τ_1), we can conclude that $\hat{\theta}_k^{\mathrm{haj1}} - \theta_k^{\mathrm{haj1}} = o_p(1)$, based on (9). As implied by (5), it follows that $\tau_k^{\mathrm{haj}}(\hat{\theta}_k^{\mathrm{haj1}})$ is the optimal estimator within the class of Hájek estimators in (2).

In Section S4 of the Supplementary Material, we provide explicit expressions for $\hat{\theta}_k^{\mathrm{ipw1}}$ and $\hat{\theta}_k^{\mathrm{haj1}}$ within each adjustment set V_k. Among the optimal estimators based on various adjustment sets, an important question arises: although each of the four types of adjustment sets provides an optimal estimator within its respective set, which type is the most advantageous or superior in terms of achieving the smallest asymptotic variance? The following theorem demonstrates the ordering of the asymptotic variances among the four types of optimal estimators.

Theorem 1. *Assuming that Assumptions 1-2 hold, the propensity score is known, and the regularity conditions in Van der Vaart (2000) hold, the optimal estimators exhibit asymptotic variances in the following order:*

$$\begin{aligned}
\Sigma_{cw}^{\mathrm{ipw1}}(\theta_{cw}^{\mathrm{ipw1}}) &\le \left\{ \Sigma_{c}^{\mathrm{ipw1}}(\theta_{c}^{\mathrm{ipw1}}),\ \Sigma_{acw}^{\mathrm{ipw1}}(\theta_{acw}^{\mathrm{ipw1}}) \right\} \le \Sigma_{ac}^{\mathrm{ipw1}}(\theta_{ac}^{\mathrm{ipw1}}),\\
\Sigma_{cw}^{\mathrm{haj1}}(\theta_{cw}^{\mathrm{haj1}}) &\le \left\{ \Sigma_{c}^{\mathrm{haj1}}(\theta_{c}^{\mathrm{haj1}}),\ \Sigma_{acw}^{\mathrm{haj1}}(\theta_{acw}^{\mathrm{haj1}}) \right\} \le \Sigma_{ac}^{\mathrm{haj1}}(\theta_{ac}^{\mathrm{haj1}}).
\end{aligned} \tag{10}$$

Theorem 1 establishes that utilizing the true confounders C and outcome predictors W in both the propensity score and regression model results in the smallest variance. However, introducing instrumental variables A into the propensity score may reduce the efficiency of ATE estimation. Our findings are consistent with previous conclusions drawn in other contexts (Patrick et al., 2011; Franklin et al., 2015). For example, Hahn (2004) demonstrated that incorporating instrumental variables in the propensity score model can increase the semi-parametric efficiency bound. Additionally, Craycroft et al. (2020) showed that including true confounders and outcome predictors in IPW estimators generally improves the precision of treatment effect estimates. Theorem 1 complements existing literature by showing that even in the context of regression adjustment, incorporating outcome predictors W, which are not directly related to the treatment variable Z, into both the propensity score and regression model can still enhance estimation efficiency.

4. Comparative analysis with estimated propensity score

In observational studies, the true propensity score is often unknown and needs to be estimated. To estimate it, a parametric model $e(V_k;\gamma_k)$ is assumed, which involves the nuisance parameter γ_k. Let $\hat{\gamma}_k$ be the maximum likelihood estimator of γ_k based on $\{(Z_i, V_{k,i})\}_{i=1}^n$. The goal is to find $\tau_k^{\mathrm{ipw}}(\theta_k, \hat{\gamma}_k)$ and $\tau_k^{\mathrm{haj}}(\theta_k, \hat{\gamma}_k)$ that have the smallest asymptotic variance among the class of estimators (1) and (2) when the propensity score is correctly specified. In order to establish the large sample properties of the proposed estimators, it becomes essential to construct a vector of estimating equations while accounting for the influence of estimating γ_k. We begin by considering the following log-likelihood function for the propensity score model,

$$\log f_k(Z, V_k;\gamma_k) = Z \log\left\{ e(V_k;\gamma_k) \right\} + (1-Z)\log\left\{ 1 - e(V_k;\gamma_k) \right\},$$

where $f_k(Z, V_k;\gamma_k)$ is the observed likelihood with respect to Z and V_k. Let $S_k(Z, V_k;\gamma_k) = \partial \log f_k(Z, V_k;\gamma_k)/\partial\gamma_k$ denote the score function. We then obtain the maximum likelihood estimator (MLE) $\hat{\gamma}_k$ by solving the following estimating equations,

$$\frac{1}{n}\sum\nolimits_{i=1}^n S_k(Z_i, V_{k,i};\gamma_k) = \frac{1}{n}\sum\nolimits_{i=1}^n \left\{ \frac{Z_i e_k(V_{k,i};\gamma_k)}{e(V_{k,i};\gamma_k)} - \frac{\left(1 - Z_i\right) e_k(V_{k,i};\gamma_k)}{e(V_{k,i};\gamma_k)} \right\} = 0, \tag{11}$$

where $e_k(V_k;\gamma_k) = \partial e(V_k;\gamma_k)/\partial \gamma_k^{\mathrm{T}}$. The following theorem establishes that both estimators $\tau_k^{\mathrm{ipw}}(\theta_k, \hat{\gamma}_k)$ and $\tau_k^{\mathrm{haj}}(\theta_k, \hat{\gamma}_k)$ are asymptotically normal and provides closed form expressions for their asymptotic variances.

Proposition 2. *Assuming that Assumptions 1-2 hold, the propensity score is correctly specified, and the regularity conditions in Van der Vaart (2000) hold, the proposed estimators are consistent and asymptotically normal for any $\theta_k = (\theta_{k,0}, \theta_{k,1})$. Specifically, we have*

$$\sqrt{n}\{\tau_k^{\mathrm{ipw}}(\theta_k, \hat{\gamma}_k) - \tau\} \xrightarrow{d} N(0, \Sigma_k^{\mathrm{ipw2}}(\theta_k)),\ \ \sqrt{n}\{\tau_k^{\mathrm{haj}}(\theta_k, \hat{\gamma}_k) - \tau\} \xrightarrow{d} N(0, \Sigma_k^{\mathrm{haj2}}(\theta_k)),$$

where for each $s \in \{\mathrm{ipw}, \mathrm{haj}\}$,

$$\begin{aligned}
&\Sigma_k^{\mathrm{ipw2}}(\theta_k) = \Sigma_k^{\mathrm{ipw1}}(\theta_k) - \psi_k^{\mathrm{ipw}}(\theta_k),\quad \Sigma_k^{\mathrm{haj2}}(\theta_k) = \Sigma_k^{\mathrm{haj1}}(\theta_k) - \psi_k^{\mathrm{haj}}(\theta_k),\\
&\psi_k^{s}(\theta_k) = \iota W_k^{s}(\theta_k) M_k^{-1} W_k^{s}(\theta_k)^{\mathrm{T}} \iota^{\mathrm{T}},\quad W_k^{s}(\theta_k) = \mathbb{E}\{G_k^{s}(Z, Y, V_k;\theta_k, \tau_0, \tau_1) S_k(Z, V_k;\gamma_k^{\star})\},
\end{aligned} \tag{12}$$

$$G_k^{\mathrm{ipw}}\left(Z,Y,V_k;\theta_k,\tau_0,\tau_1\right)=\left\{\frac{(1-Z)\left(Y-V_k^{\mathrm{T}}\theta_{k,0}\right)}{1-e(V_k;\gamma_k^{\star})}-\tau_0,\ \frac{Z(Y-V_k^{\mathrm{T}}\theta_{k,1})}{e(V_k;\gamma_k^{\star})}-\tau_1\right\}^{\mathrm{T}},$$

$$G_k^{\mathrm{haj}}\left(Z,Y,V_k;\theta_k,\tau_0,\tau_1\right)=\left\{\frac{(1-Z)\left(Y-V_k^{\mathrm{T}}\theta_{k,0}-\tau_0\right)}{1-e(V_k;\gamma_k^{\star})},\ \frac{Z(Y-V_k^{\mathrm{T}}\theta_{k,1}-\tau_1)}{e(V_k;\gamma_k^{\star})}\right\}^{\mathrm{T}},$$

$\imath=(-1,1)$ *and* $M_k=\mathbb{E}\{S_k(Z,V_k;\gamma_k^{\star})S_k^{\mathrm{T}}(Z,V_k;\gamma_k^{\star})\}$. *If* $\hat{\theta}_k-\theta_k=o_p(1)$, *we have*

$$\sqrt{n}\{\tau_k^{\mathrm{ipw}}(\hat{\theta}_k,\hat{\gamma}_k)-\tau_k^{\mathrm{ipw}}(\theta_k,\hat{\gamma}_k)\}=o_p(1),\ \sqrt{n}\{\tau_k^{\mathrm{haj}}(\hat{\theta}_k,\hat{\gamma}_k)-\tau_k^{\mathrm{haj}}(\theta_k,\hat{\gamma}_k)\}=o_p(1). \tag{13}$$

Proposition 2 establishes the consistency and asymptotic normality of the proposed estimators when the propensity score is correctly specified. In comparison to (3) and (4), both the asymptotic variances in (12) are reduced by an additional term that results from estimating the parameter γ_k. For each $s\in\{\mathrm{ipw},\mathrm{haj}\}$, the positive semi-definiteness of the matrix $\psi_k^s(\theta_k)$ in (12) implies that the asymptotic variance of $\tau_k^s(\theta_k,\hat{\gamma}_k)$ when the propensity score is estimated is not greater than when the propensity score is known. When $\theta_{k,1}=\theta_{k,0}=0$, the proposed estimators $\tau_k^s(0,\hat{\gamma}_k)$ reduce to the benchmark weighted estimators. This finding is consistent with well-known results in causal inference that even when the propensity score is known, using the estimated propensity score can improve the estimation efficiency (Hirano et al., 2003). Therefore, in practice, we recommend employing $\tau_k^s(\theta_k,\hat{\gamma}_k)$ to improve estimation efficiency rather than $\tau_k^s(\theta_k,\gamma_k^{\star})$.

To illustrate the process of finding the optimal estimator with an estimated propensity score, we focus on the IPW estimator as a specific example. For each adjustment set V_k, the goal is to minimize $\Sigma_k^{\mathrm{ipw2}}(\theta_k)$ in (12) and obtain the optimal coefficients, denoted by θ_k^{ipw2}. Specifically, we can obtain $\hat{\theta}_k^{\mathrm{ipw2}}$ by solving the sample versions of the following estimating equations:

$$\frac{\partial}{\partial\theta_{k,1}}\Sigma_k^{\mathrm{ipw2}}(\theta_k)=\mathbb{E}\left\{\frac{ZV_k(Y-V_k^{\mathrm{T}}\theta_{k,1})}{e^2(V_k;\gamma_k^{\star})}\right\}-\frac{\partial}{\partial\theta_{k,1}}\psi_k^{\mathrm{ipw}}(\theta_k)=0,$$

$$\frac{\partial}{\partial\theta_{k,0}}\Sigma_k^{\mathrm{ipw2}}(\theta_k)=\mathbb{E}\left[\frac{(1-Z)V_k(Y-V_k^{\mathrm{T}}\theta_{k,0})}{\{1-e(V_k;\gamma_k^{\star})\}^2}\right]-\frac{\partial}{\partial\theta_{k,0}}\psi_k^{\mathrm{ipw}}(\theta_k)=0.$$

For simplicity, we omit the details of the estimation procedure for the Hájek estimator $\tau_k^{\mathrm{haj}}(\theta_k,\hat{\gamma}_k)$. Corollary 1 compares the optimal asymptotic variances of different estimators when using the known propensity score versus the estimated propensity score.

Corollary 1. *Assuming that Assumptions 1-2 hold, the propensity score is correctly specified, and the regularity conditions in Van der Vaart (2000) hold, we have* $\Sigma_k^{\mathrm{ipw2}}(\theta_k^{\mathrm{ipw2}})\le\Sigma_k^{\mathrm{ipw1}}(\theta_k^{\mathrm{ipw1}})$ *and* $\Sigma_k^{\mathrm{haj2}}(\theta_k^{\mathrm{haj2}})\le\Sigma_k^{\mathrm{haj1}}(\theta_k^{\mathrm{haj1}})$.

Corollary 1 highlights the advantage of using the estimated propensity score over the known propensity score. However, when the propensity score is estimated, it becomes challenging to provide a theoretical result similar to Theorem 1 due to the complex expression of $\psi_k^s(\theta_k)$ in (12). This expression involves Fisher information matrices defined through the propensity score, making it difficult to directly compare the asymptotic variances under different adjustment sets. However, the simulation results in the next section demonstrate that, even when using the estimated propensity score, results of a similar order typically hold in practice.

5. Experiments

5.1. Simulation studies with appropriate propensity scores

In this section, we perform simulation studies to evaluate the finite sample performance of the proposed estimators. Specifically, we compare two classes of improved estimators with benchmark estimators, both under the true propensity score and the estimated propensity score. We consider the following data-generating mechanism:

(a) We generate the common covariate, instrumental variable, and outcome predictor from the following distributions: $C\sim N(0,0.5^2)$, $A=0.6C+\epsilon_A$, where $\epsilon_A\sim N(0,0.5^2)$, and $W=0.6C+\epsilon_W$, where $\epsilon_W\sim N(0,0.5^2)$.

(b) We generate the treatment variable Z from a Bernoulli distribution:

$$\mathrm{pr}(Z=1\mid A,C,W)=\Phi(\eta-0.5A+0.25C) \tag{14}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution and the parameter η characterizes the degree of deviation from the positivity Assumption 2. We will use the parameter η to study the robustness of solutions to (8) and (9) in the next subsection.

(c) For the potential outcomes, we consider the following two cases:
Case 1: $Y_0=2-3W-3C+\epsilon_0$ and $Y_1=5+5W+6C+\epsilon_1$;
Case 2: $Y_0=2-3C-3\exp(W)+\epsilon_0$ and $Y_1=5+5W+6C^2+\epsilon_1$;
where $\epsilon_0\sim N(0,1)$ and $\epsilon_1\sim N(0,1)$. The observed outcome is generated from $Y=ZY_1+(1-Z)Y_0$.

Table 1
Asymptotic performance among various estimators for Case 1. Bias are scaled by sample size 10^2.

	V_c				V_{cw}				V_{ac}				V_{acw}			
IPW estimators	Bias	Evar	Avar	CP	Bias	Evar	Avar	CP	Bias	Evar	Avar	CP	Bias	Evar	Avar	CP
$\tau_k^{\mathrm{ipw}}(0,\gamma_k^\star)$	-0.39	121.08	118.19	0.96	-0.39	121.08	118.19	0.96	-0.98	126.30	123.29	0.95	-0.98	126.30	123.29	0.95
$\tau_k^{\mathrm{ipw}}(0,\hat{\gamma}_k)$	0.57	67.07	68.04	0.96	0.70	67.07	67.59	0.96	-0.07	70.94	71.01	0.95	0.06	71.27	70.56	0.95
$\tau_k^{\mathrm{ipw}}(\hat{\theta}_k^{\mathrm{ipw1}},\gamma_k^\star)$	-0.58	69.07	66.22	0.95	-0.47	51.83	49.83	0.96	-0.82	72.73	69.12	0.94	-0.60	54.18	52.05	0.95
$\tau_k^{\mathrm{ipw}}(\hat{\theta}_k^{\mathrm{ipw2}},\hat{\gamma}_k)$	-0.04	20.58	20.44	0.94	-0.01	3.85	4.07	0.95	-0.24	22.30	21.34	0.94	-0.05	4.22	4.34	0.95
	V_c				V_{cw}				V_{ac}				V_{acw}			
Hájek estimators	Bias	Evar	Avar	CP	Bias	Evar	Avar	CP	Bias	Evar	Avar	CP	Bias	Evar	Avar	CP
$\tau_k^{\mathrm{haj}}(0,\gamma_k^\star)$	0.28	69.91	68.03	0.95	0.28	69.91	68.03	0.95	-0.28	73.54	70.74	0.95	-0.28	73.54	70.29	0.95
$\tau_k^{\mathrm{haj}}(0,\hat{\gamma}_k)$	0.56	67.07	70.37	0.96	0.70	67.05	70.37	0.96	-0.01	70.61	73.15	0.96	0.11	70.92	73.15	0.95
$\tau_k^{\mathrm{haj}}(\hat{\theta}_k^{\mathrm{haj1}},\gamma_k^\star)$	-0.05	20.56	20.36	0.94	-0.01	3.84	4.00	0.96	-0.24	22.23	21.18	0.93	-0.08	4.10	4.17	0.95
$\tau_k^{\mathrm{haj}}(\hat{\theta}_k^{\mathrm{haj2}},\hat{\gamma}_k)$	-0.04	20.57	20.44	0.94	-0.01	3.85	4.07	0.95	-0.22	22.18	21.24	0.94	-0.05	4.14	4.23	0.95

It is straightforward to verify that the true propensity score models used in this simulation study are all probit models. Tables 1-2 present the simulation results for Cases 1-2, respectively, with a sample size of $n = 1,000$. The bias (Bias), empirical variance (Evar), average estimated variance (Avar), and 95% coverage probabilities (CP) are computed using 1,000 replications. The specific formulas for Evar and Avar can be found in Section S7.1 of the Supplementary Material. In summary, Evar represents a nonparametric estimate of the variance (scaled by the sample size), while Avar corresponds to the sample variance estimate obtained from our proposed formulas (also scaled by the sample size, as shown in (3), (4), and (12)). The results for these two variance estimates are expected to be similar.

We first evaluate the finite sample performance of the estimators from (1) and (2) with $\eta = 0.1$ in (14). In the next subsection, we will assess the stability of the estimators by varying different values of η. Each estimator is applied with distinct combinations of γ_k and θ_k. Bias values are multiplied by 100, and metrics for evaluating asymptotic variances are scaled by the sample size as mentioned earlier. The conclusions from Table 1 are summarized as follows:

(1) The proposed estimators typically yield unbiased estimates when the propensity score is correctly modeled or known, regardless of whether $\gamma_k^\star$ or $\hat{\gamma}_k$ is used. Additionally, the results for Evar and Avar are very close, indicating the validity of our variance expressions in (3), (4), and (12). Furthermore, the 95% coverage probabilities are close to the nominal level in most cases.
(2) When known propensity scores are employed, the simulation performance of the proposed estimators, defined on four adjustment sets, aligns with the findings presented in (10). Specifically, the proposed estimators utilizing V_{cw} are expected to achieve the smallest variance, while those utilizing V_{ac} are expected to exhibit the worst performance. For instance, from the seventh row of Table 1, we observe that the average estimated variance of $\tau_{cw}^{\mathrm{haj}}(\hat{\theta}_{cw}^{\mathrm{haj1}},\gamma_{cw}^\star)$ is 4.00, while $\tau_{ac}^{\mathrm{haj}}(\hat{\theta}_{ac}^{\mathrm{haj1}},\gamma_{ac}^\star)$ is 21.18.
(3) The proposed estimators based on the estimated propensity score significantly reduce both the asymptotic variance and the estimated variance, providing strong support for the validity of (12). For example, from the third and fourth rows of Table 1, we observe that the average estimated variance of $\tau_c^{\mathrm{ipw}}(\hat{\theta}_c^{\mathrm{ipw2}},\hat{\gamma}_c)$ is 20.44, while $\tau_c^{\mathrm{ipw}}(\hat{\theta}_c^{\mathrm{ipw1}},\gamma_c^\star)$ is 66.22.
(4) Compared to the benchmark IPW estimator $\tau_k^{\mathrm{ipw}}(0,\hat{\gamma}_k)$, we find that the proposed estimator $\tau_k^{\mathrm{ipw}}(\hat{\theta}_k^{\mathrm{ipw1}},\gamma_k^\star)$ has smaller variances, and the Hájek-type benchmark estimator $\tau_k^{\mathrm{haj}}(0,\hat{\gamma}_k)$ also exhibits larger variances compared to $\tau_k^{\mathrm{haj}}(\hat{\theta}_k^{\mathrm{haj1}},\gamma_k^\star)$. For example, from the second and fourth rows of Table 1, we see that the average estimated variance of $\tau_{cw}^{\mathrm{ipw}}(\hat{\theta}_{cw}^{\mathrm{ipw1}},\hat{\gamma}_{cw})$ is 67.59, while $\tau_{ac}^{\mathrm{ipw}}(\hat{\theta}_{ac}^{\mathrm{ipw1}},\hat{\gamma}_{ac})$ is 4.07.
(5) The empirical evidence consistently shows that the Hájek estimators generally exhibit smaller variances compared to the IPW estimators across various cases. For instance, from the first and fifth rows of Table 1, we observe that the average estimated variance of $\tau_{cw}^{\mathrm{ipw}}(0,\gamma_{cw}^\star)$ is 118.19, whereas $\tau_{cw}^{\mathrm{haj}}(0,\gamma_{cw}^\star)$ is notably lower at 68.03.
(6) Despite the lack of theoretical guarantees, the empirical results suggest that the asymptotic variance of the proposed estimators based on the estimated propensity score exhibits a similar order of performance to that of the true propensity score, as given in (10). Therefore, considering both asymptotic efficiency and finite sample performance, we recommend the use of the estimator $\tau_{cw}^{\mathrm{haj}}(\hat{\theta}_{cw}^{\mathrm{haj2}},\hat{\gamma}_{cw})$ in practice.

Simulation results for Case 2 are similar to those of Case 1 and are presented in Table 2, which we omit for brevity.

5.2. Simulation studies with potentially extreme propensity scores

As shown in Proposition 1 and Proposition 2, the inverse of the propensity score influences the variance. For instance, $1/e^2(V_{k,i};\gamma_k^\star)$ affects the solutions of (8) and (9). In this subsection, we vary the parameter η in (14) to investigate the finite sample performance of the proposed estimators under potentially extreme propensity scores.

The results for $\tau_k^{\mathrm{ipw}}(0,\hat{\gamma}_k)$ and $\tau_k^{\mathrm{haj}}(0,\hat{\gamma}_k)$ for Case 1 are shown in Fig. 2, with the true value represented by the horizontal dashed line. The results for $\tau_k^{\mathrm{ipw}}(\hat{\theta}_k^{\mathrm{ipw2}},\hat{\gamma}_k)$ and $\tau_k^{\mathrm{haj}}(\hat{\theta}_k^{\mathrm{haj2}},\hat{\gamma}_k)$ for Case 1 are shown in Fig. 3, with the true value again marked by the horizontal

Table 2
Asymptotic performance among various estimators for Case 2. Bias are scaled by sample size 10^2.

	V_c				V_{cw}				V_{ac}				V_{acw}			
IPW estimators	Bias	Evar	Avar	CP	Bias	Evar	Avar	CP	Bias	Evar	Avar	CP	Bias	Evar	Avar	CP
$\tau_k^{\mathrm{ipw}}(0,\gamma_k^\star)$	-0.14	69.60	70.08	0.96	-0.14	69.60	70.08	0.96	-0.53	75.28	75.31	0.96	-0.53	75.28	75.31	0.96
$\tau_k^{\mathrm{ipw}}(0,\hat{\gamma}_k)$	0.07	46.55	49.42	0.96	0.17	46.31	49.24	0.96	-0.47	48.78	51.59	0.96	-0.38	48.72	51.41	0.96
$\tau_k^{\mathrm{ipw}}(\hat{\theta}_k^{\mathrm{ipw1}},\gamma_k^\star)$	-1.26	51.57	50.89	0.95	-1.40	33.56	32.57	0.94	-1.58	56.67	55.12	0.95	-1.73	37.28	36.06	0.94
$\tau_k^{\mathrm{ipw}}(\hat{\theta}_k^{\mathrm{ipw2}},\hat{\gamma}_k)$	-1.33	28.98	30.08	0.95	-1.73	12.47	12.64	0.95	-1.32	33.14	33.54	0.95	-1.41	14.42	14.46	0.94
	V_c				V_{cw}				V_{ac}				V_{acw}			
Hájek estimators	Bias	Evar	Avar	CP	Bias	Evar	Avar	CP	Bias	Evar	Avar	CP	Bias	Evar	Avar	CP
$\tau_k^{\mathrm{haj}}(0,\gamma_k^\star)$	0.29	48.10	49.39	0.95	0.29	48.10	49.39	0.95	-0.20	50.05	51.47	0.96	-0.20	50.05	51.29	0.96
$\tau_k^{\mathrm{haj}}(0,\hat{\gamma}_k)$	0.07	46.53	51.09	0.96	0.17	46.29	51.09	0.96	-0.43	48.55	53.30	0.96	-0.35	48.48	53.30	0.96
$\tau_k^{\mathrm{haj}}(\hat{\theta}_k^{\mathrm{haj1}},\gamma_k^\star)$	-0.96	31.19	32.25	0.96	-1.14	13.60	13.93	0.95	-1.25	33.02	33.52	0.96	-1.31	14.07	14.44	0.95
$\tau_k^{\mathrm{haj}}(\hat{\theta}_k^{\mathrm{haj2}},\hat{\gamma}_k)$	-1.35	29.04	30.06	0.95	-1.69	12.43	12.60	0.95	-1.28	32.98	33.48	0.95	-1.37	14.28	14.40	0.95

Table 3
Covariate description for the SIPP data.

Name	Description	Covariate Type
Gender	Gender	C
Age	Age	W
Race	Race	C
Education	Education levels	C
Marital	Marital status	W
Citizenship	Hispanic, Hispanic or Latino or not	A

dashed line. Due to space constraints, additional results for Case 1 are provided in Section S7.2 of the Supplementary Material, and results for Case 2 are also presented in Section S7.3 of the Supplementary Material. All the findings are summarized as follows:

(1) We find that Fig. 3 shows significantly smaller estimated variances compared to Fig. 2, indicating that for all values of η, the proposed estimators $\tau_k^{\mathrm{ipw}}(\hat{\theta}_k^{\mathrm{ipw2}},\hat{\gamma}_k)$ and $\tau_k^{\mathrm{haj}}(\hat{\theta}_k^{\mathrm{haj2}},\hat{\gamma}_k)$ perform better than the benchmark propensity score based estimators $\tau_k^{\mathrm{ipw}}(0,\hat{\gamma}_k)$ and $\tau_k^{\mathrm{haj}}(0,\hat{\gamma}_k)$. For example, the 95% confidence intervals in the first row, first column of Fig. 3 are narrower than those in the first row, first column of Fig. 2.
(2) In both Figs. 2 and 3, we observe that including the additional covariate W results in narrower 95% confidence intervals, making the estimates in the second column more stable compared to those in the first column. For instance, across all values of η, the 95% confidence intervals in the second row, second column of Fig. 2 are narrower than those in the second row, first column.
(3) Finally, we find that our simulation results are quite robust to changes in η. The point estimates deviate from the true values only in some extreme cases, such as when $\eta = -2$. However, most of the 95% confidence intervals still cover the true value, except when using the V_{cw} adjustment set. This issue may arise because the estimated variance for V_{cw} is too small and the estimated bias is relatively large, leading to confidence intervals that do not cover the true value.

6. Application

In this section, we use real data from the Survey of Income and Program Participation (SIPP) to illustrate our methods. The survey collects information on child and dependent care, employment status, hours worked, labor force, income from work, and demographic characteristics of household members. Our focus is on estimating the causal effect of the number of children in a household on the household's monthly child payment, using hourly wages in U.S. dollars as a measure of pay.

The binary treatment variable Z takes on the value of 0 if a family has only one child and 1 if a family has more than one child. Due to the reported payments in the original dataset ranging from 1 to 25,000, with an empirical variance of 6,030, we take the square root of the reported payment and denote it as Y. In addition to the treatment and outcome variables, there are six demographic covariates included. Table 3 provides a brief description of these covariates. For all covariates, both continuous and categorical, we center them by subtracting the corresponding sample means. We compile a dataset from the SIPP 2021 data, which includes 7956 individuals with no missing values.

Before applying our method, we need to classify the confounding factors. We use similar model parameterizations for the propensity score model as in the simulation studies, using probit models for four adjustment sets. For simplicity, we establish a linear model for the outcome variable to classify the types of covariates. We label the covariates that appear in both models as C, the covariates that only appear in the propensity score model as A, and the covariates that only appear in the outcome model as W. The results are shown in the last column of Table 3. In addition, we also consider other nonparametric methods such as classification and regression trees (Breiman, 2017), Bayesian additive regression trees (Chipman et al., 2010), and Causal Tree (Athey and Imbens, 2016).

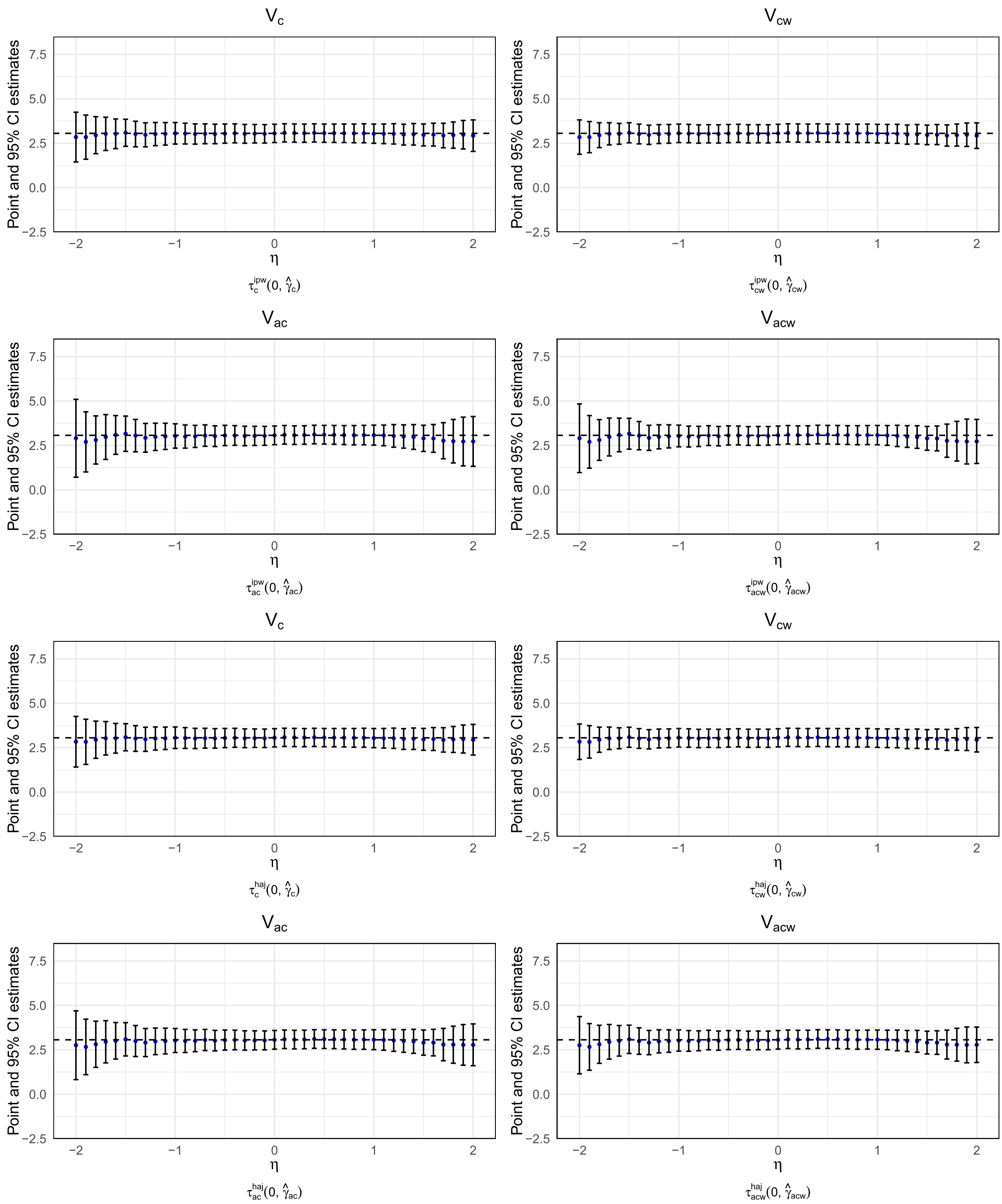

Fig. 2. The point and 95% confidence interval (CI) estimates of $\tau_k^{\text{ipw}}(0, \hat{\gamma}_k)$ and $\tau_k^{\text{haj}}(0, \hat{\gamma}_k)$ with different η in propensity score (14) for Case 1.

Importantly, all these methods yield similar variable selection results as shown in Table 3. This indicates that the classification of background covariates is robust and not sensitive to the specific selection models used.

Table 4 presents the point estimates, average estimated variance (Avar), and 95% confidence interval (CI) of our analysis. As the true propensity score is unknown in practice, we utilize the estimated propensity score for real data analysis. All the candidate estimators yield similar results, indicating that the number of children has a significant causal effect on household payment. Specifically, the cost of one child in a family is about 20 less than that of families with more than two children on the square scale. Moreover, when using the estimated propensity score, the proposed estimator based on the adjustment set V_{cw} consistently provides the estimate with

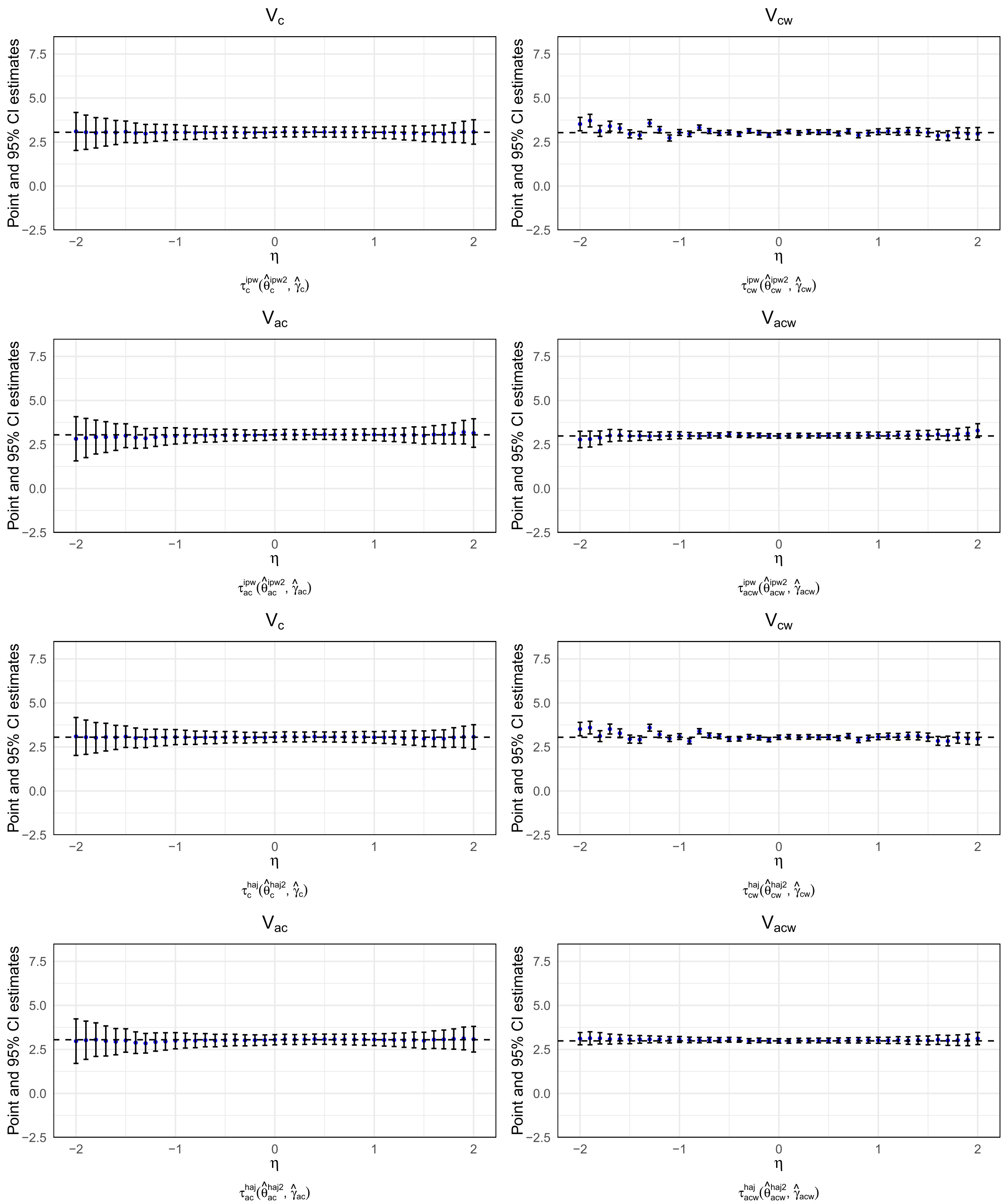

Fig. 3. The point and 95% confidence interval estimates of $\tau_k^{\text{ipw}}(\hat{\theta}_k^{\text{ipw2}}, \hat{\gamma}_k)$ and $\tau_k^{\text{haj}}(\hat{\theta}_k^{\text{haj2}}, \hat{\gamma}_k)$ with different η in propensity score (14) for Case 1.

the smallest variance. For example, the estimated variance of $\tau_{cw}^{\text{ipw}}(\hat{\theta}_{cw}^{\text{ipw2}}, \hat{\gamma}_{cw})$ is 0.57, while that of $\tau_{cw}^{\text{haj}}(\hat{\theta}_{cw}^{\text{haj2}}, \hat{\gamma}_{cw})$ is 0.56. Additionally, employing instrumental variable A and true confounders C in the propensity score model tends to increase the estimated variance in most cases. These findings are consistent with our theoretical results.

7. Discussion

In this paper, we present two propensity score based estimators that leverage regression models to improve efficiency. We analyze the optimal solutions of these proposed estimators using both known and estimated propensity scores. Moreover, we compare the

Table 4
Estimation results for SIPP data.

		IPW estimators				Hájek estimators			
Covariate	γ	θ	Point Estimate	Avar	95% CI	θ	Point Estimate	Avar	95% CI
V_c	$\hat{\gamma}_c$	0	20.39	0.62	(19.18, 21.59)	0	20.40	0.61	(19.19, 21.60)
V_{cw}	$\hat{\gamma}_{cw}$	0	20.71	0.59	(19.55, 21.86)	0	20.74	0.59	(19.59, 21.89)
V_{ac}	$\hat{\gamma}_{ac}$	0	20.31	0.62	(19.11, 21.52)	0	20.41	0.61	(19.20, 21.61)
V_{acw}	$\hat{\gamma}_{acw}$	0	20.59	0.59	(19.44, 21.74)	0	20.72	0.59	(19.57, 21.86)
V_c	$\hat{\gamma}_c$	$\hat{\theta}_c^{\mathrm{ipw2}}$	20.09	0.60	(18.91, 21.27)	$\hat{\theta}_c^{\mathrm{haj2}}$	20.06	0.60	(18.88, 21.23)
V_{cw}	$\hat{\gamma}_{cw}$	$\hat{\theta}_{cw}^{\mathrm{ipw2}}$	20.42	0.57	(19.30, 21.53)	$\hat{\theta}_{cw}^{\mathrm{haj2}}$	20.56	0.56	(19.45, 21.67)
V_{ac}	$\hat{\gamma}_{ac}$	$\hat{\theta}_{ac}^{\mathrm{ipw2}}$	19.49	0.59	(18.33, 20.65)	$\hat{\theta}_{ac}^{\mathrm{haj2}}$	19.57	0.59	(18.41, 20.73)
V_{acw}	$\hat{\gamma}_{acw}$	$\hat{\theta}_{acw}^{\mathrm{ipw2}}$	20.16	0.57	(19.05, 21.28)	$\hat{\theta}_{acw}^{\mathrm{haj2}}$	19.94	0.56	(18.85, 21.03)

asymptotic variances of four optimal estimators based on different adjustment sets. Additionally, we investigate the potential of the proposed estimators when utilizing estimated propensity scores, which provides additional opportunities for enhancing estimation efficiency compared to using known propensity scores.

In observational studies, adding more covariates to the propensity score model may seem beneficial for improving estimation efficiency. However, it can be counterintuitive that including additional variables, particularly instrumental variables, can actually harm efficiency (Craycroft et al., 2020; Patrick et al., 2011). A plausible explanation is that the influences on the outcome variable can be entirely explained by the treatment variable Z and the true confounding factors C. As a result, the instrumental variables A become spurious variables for the outcome and do not possess any predictive power for improving estimation efficiency through propensity scores. In fact, including such instrumental variables in the propensity score model may even increase estimation instability (Hahn, 2004).

In practice, we suggest utilizing prior information or nonparametric methods to categorize different types of covariates efficiently. In our simulation studies, we employed a probit model and normal distribution for the data-generating mechanism. Under this special setting, the propensity score model for different adjustment sets can be parameterized. However, in practice, modeling all propensity score models accurately can be challenging. The simulation studies suggest that parameterizing the propensity score models for the true confounders and predictor variables can be most beneficial for improving estimation efficiency.

This paper contributes to the efficiency analysis of estimators based on propensity scores, particularly for different adjustment sets. Our method currently requires centered covariates to ensure consistent estimation, and we recommend centering non-centered covariates before analysis. How to use more flexible models to handle non-centered covariates is another important issue (Luo et al., 2024). Additionally, while our study focuses on binary treatments, extending it to general treatment allocation is an interesting direction. Moreover, investigating regression challenges under different adjustment sets in the presence of spillover effects is another area worth exploring (Liu et al., 2016). These issues are beyond the scope of this paper and are left for future research.

Acknowledgement

Shanshan Luo's research is supported by the National Natural Science Foundation of China (12401378). Xueli Wang's research is supported by the Joint Key Research Project funded by the Beijing Municipal Education Commission and the Beijing Municipal Natural Science Foundation (No. 23JA0006). Wei Li's research is supported by the National Natural Science Foundation of China (12101607, 12471269) and Beijing Natural Science Foundation (1232008).

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.csda.2024.108079.

References

Abadie, A., Imbens, G.W., 2016. Matching on the estimated propensity score. Econometrica 84, 781–807.
Athey, S., Imbens, G., 2016. Recursive partitioning for heterogeneous causal effects. Proc. Natl. Acad. Sci. 113, 7353–7360.
Bloniarz, A., Liu, H., Zhang, C.H., Sekhon, J.S., Yu, B., 2016. Lasso adjustments of treatment effect estimates in randomized experiments. Proc. Natl. Acad. Sci. 113, 7383–7390.
Breiman, L., 2017. Classification and regression trees. Routledge.
Brookhart, M.A., Schneeweiss, S., Rothman, K.J., Glynn, R.J., Avorn, J., Stürmer, T., 2006. Variable selection for propensity score models. Am. J. Epidemiol. 163, 1149–1156.
Chipman, H.A., George, E.I., McCulloch, R.E., 2010. Bart: Bayesian additive regression trees. Ann. Appl. Stat. 4, 266–298.
Craycroft, J.A., Huang, J., Kong, M., 2020. Propensity score specification for optimal estimation of average treatment effect with binary response. Stat. Methods Med. Res. 29, 3623–3640.
De Luna, X., Waernbaum, I., Richardson, T.S., 2011. Covariate selection for the nonparametric estimation of an average treatment effect. Biometrika 98, 861–875.
Fisher, R.A., 1966. The Design of Experiments, vol. 21. Springer.
Franklin, J.M., Eddings, W., Glynn, R.J., Schneeweiss, S., 2015. Regularized regression versus the high-dimensional propensity score for confounding adjustment in secondary database analyses. Am. J. Epidemiol. 182, 651–659.
Hahn, J., 2004. Functional restriction and efficiency in causal inference. Rev. Econ. Stat. 86, 73–76.

Hájek, J., 1971. Comment on “An Essay on the Logical Foundations of Survey Sampling, Part One”. In: The Foundations of Survey Sampling, p. 236.
Hirano, K., Imbens, G.W., Ridder, G., 2003. Efficient estimation of average treatment effects using the estimated propensity score. Econometrica 71, 1161–1189.
Horowitz, J.L., Manski, C.F., 1995. Identification and robustness with contaminated and corrupted data. Econometrica, 281–302.
Horvitz, D.G., Thompson, D.J., 1952. A generalization of sampling without replacement from a finite universe. J. Am. Stat. Assoc. 47, 663–685.
Li, F., Morgan, K.L., Zaslavsky, A.M., 2018. Balancing covariates via propensity score weighting. J. Am. Stat. Assoc. 113, 390–400.
Li, X., Ding, P., 2017. General forms of finite population central limit theorems with applications to causal inference. J. Am. Stat. Assoc. 112, 1759–1769.
Lin, W., 2013. Agnostic notes on regression adjustments to experimental data: reexamining Freedman’s critique. Ann. Appl. Stat. 7, 295–318.
Liu, L., Hudgens, M.G., Becker-Dreps, S., 2016. On inverse probability-weighted estimators in the presence of interference. Biometrika 103, 829–842.
Lunceford, J.K., Davidian, M., 2004. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. Stat. Med. 23, 2937–2960.
Luo, S., Shi, M., Li, W., Wang, X., Geng, Z., 2024. Efficiency-improved doubly robust estimation with non-confounding predictive covariates. arXiv:2402.14438.
Neyman, J., 1923. On the application of probability theory to agricultural experiments. Stat. Sci. 5, 4.
Pan, Y., Zhao, Y.Q., 2021. Improved doubly robust estimation in learning optimal individualized treatment rules. J. Am. Stat. Assoc. 116, 283–294.
Patrick, A.R., Schneeweiss, S., Brookhart, M.A., Glynn, R.J., Rothman, K.J., Avorn, J., Stürmer, T., 2011. The implications of propensity score variable selection strategies in pharmacoepidemiology: an empirical illustration. Pharmacoepidemiol. Drug Saf. 20, 551–559.
Rosenbaum, P.R., Rubin, D.B., 1983. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. J. R. Stat. Soc., Ser. B, Methodol. 45, 212–218.
Rubin, D.B., 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. J. Educ. Psychol. 66, 688.
Rubin, D.B., 1990. Comment: Neyman (1923) and causal inference in experiments and observational studies. Stat. Sci. 5, 472–480.
Stefanski, L.A., Boos, D.D., 2002. The calculus of m-estimation. Am. Stat. 56, 29–38.
Ting, Y., Jun, S., Yanyao, Y., Qingyuan, Z., 2023. Toward better practice of covariate adjustment in analyzing randomized clinical trials. J. Am. Stat. Assoc. 118, 2370–2382.
Van der Vaart, A.W., 2000. Asymptotic Statistics, vol. 3. Cambridge University Press.
Vansteelandt, S., Daniel, R., 2014. On regression adjustment for the propensity score. Stat. Med. 33, 4053–4072.