Algorithm as Experiment: Machine Learning, Market Design, and Policy Eligibility Rules

Yusuke Narita  
*Yale University*

Kohei Yata  
*Yale University*

Follow this and additional works at: https://elischolar.library.yale.edu/cowles-discussion-paper-series

Part of the Economics Commons

Recommended Citation
https://elischolar.library.yale.edu/cowles-discussion-paper-series/2795

This Discussion Paper is brought to you for free and open access by the Cowles Foundation at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Cowles Foundation Discussion Papers by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.
ALGORITHM AS EXPERIMENT:
MACHINE LEARNING, MARKET DESIGN,
AND POLICY ELIGIBILITY RULES

By

Yusuke Narita and Kohei Yata

May 2024

COWLES FOUNDATION DISCUSSION PAPER NO. 2391

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/
Algorithm as Experiment: Machine Learning, Market Design, and Policy Eligibility Rules*

Yusuke Narita Kohei Yata†‡

May 28, 2024

Abstract

Algorithms make a growing portion of policy and business decisions. We develop a treatment-effect estimator using algorithmic decisions as instruments for a class of stochastic and deterministic algorithms. Our estimator is consistent and asymptotically normal for well-defined causal effects. A special case of our setup is multidimensional regression discontinuity designs with complex boundaries. We apply our estimator to evaluate the Coronavirus Aid, Relief, and Economic Security Act, which allocated many billions of dollars worth of relief funding to hospitals via an algorithmic rule. The funding is shown to have little effect on COVID-19-related hospital activities. Naive estimates exhibit selection bias.

*Keywords: Algorithmic decision making, instrumental variables, propensity score, regression discontinuity design, COVID-19 hospital relief funding

†Narita: Department of Economics, Yale University, email: yusuke.narita@yale.edu. Yata: Department of Economics, University of Wisconsin–Madison, email: yata@wisc.edu.

‡For their suggestions, we are grateful to Joseph Altonji, Josh Angrist, Tim Armstrong, Yingying Dong, Ivan Fernandez-Val, Pat Kline, Michal Kolesár, Chris Walters, and seminar participants at the ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO ’21), American Economic Association, Berkeley, BU, Caltech, Columbia, CEMFI, Counterfactual Machine Learning Workshop, Econometric Society, European Economic Association, Hitotsubashi, JSAI, Michigan, Stanford, UC Irvine, the University of Tokyo, Virtual Market Design Seminar, and Yale. We are especially indebted to Aneesha Parvathaneni, Richard Liu, Richard Gong, and several others for expert research assistance.
1 Introduction

Today’s society increasingly resorts to algorithms for decision making and resource allocation. For example, judges in the US make legal decisions aided by predictions from supervised machine learning algorithms. Supervised learning is also used by governments to detect potential criminals and terrorists, and by banks and insurance companies to screen potential customers. Tech companies like Facebook, Microsoft, and Netflix allocate digital content by reinforcement learning and bandit algorithms. Retailers and e-commerce platforms engage in algorithmic pricing. Similar algorithms are encroaching on high-stakes settings, such as in education, healthcare, and the military.

Other types of algorithms also loom large. School districts, college admissions systems, and labor markets use matching algorithms for position and seat allocations. Objects worth astronomical sums of money change hands every day in algorithmically run auctions. Many public policy domains like Medicaid often use algorithmic rules to decide who is eligible.

All of the above examples share a common trait: a decision-making algorithm makes decisions based only on its observable input variables. Thus conditional on the observable variables, algorithmic treatment decisions are assigned independently of any potential outcome. This property turns algorithm-based treatment decisions into instrumental variables (IVs) that can be used for measuring the causal effect of the final treatment assignment. The algorithm-based IV may produce stratified randomization, regression-discontinuity-style local variation, or some combination of the two.

This paper shows how to use data obtained from algorithmic decision making to identify and estimate causal effects. The analyst observes a random iid sample \( \{(Y_i, X_i, D_i, Z_i)\}_{i=1}^n \), where \( Y_i \) is the outcome of interest, \( X_i \in \mathbb{R}^p \) is a vector of pre-treatment covariates used as the algorithm’s input variables, \( D_i \) is the binary treatment assignment, possibly made by humans, and \( Z_i \) is the binary treatment recommendation made by a known algorithm. The algorithm takes \( X_i \) as input and computes the probability of the treatment recommendation \( A(X_i) = \Pr(Z_i = 1|X_i) \). \( Z_i \) is then randomly determined based on the known probability \( A(X_i) \) independently of everything else conditional on \( X_i \). The algorithm’s recommendation \( Z_i \) may influ-
ence the final treatment assignment $D_i$, determined as $D_i = Z_i D_i(1) + (1 - Z_i) D_i(0)$, where $D_i(z)$ is the potential treatment assignment that would be realized if $Z_i = z$. Finally, the observed outcome $Y_i$ is determined as $Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$, where $Y_i(1)$ and $Y_i(0)$ are potential outcomes that would be realized if the individual were treated and not treated, respectively. This setup is an IV model where the IV satisfies the conditional independence condition but may not satisfy the overlap (full-support) condition. This setup nests the classic propensity-score and regression-discontinuity-design (RDD) setups.

Within this framework, we first characterize the sources of causal-effect identification for a class of data-generating algorithms. This class includes all of the aforementioned examples, nesting both stochastic and deterministic algorithms. The sources of causal-effect identification turn out to be summarized by a suitable modification of the Propensity Score. We call it the Approximate Propensity Score (APS). For each covariate value $x$, the Approximate Propensity Score is the average probability of a treatment recommendation in a shrinking neighborhood around $x$, defined as

$$p^A(x) \equiv \lim_{\delta \to 0} \frac{\int_{B(x,\delta)} A(x^*) dx^*}{\int_{B(x,\delta)} dx^*},$$

where $B(x, \delta)$ is a $p$-dimensional ball with radius $\delta$ centered at $x$. The Approximate Propensity Score provides an easy-to-check condition for what causal effects the data from an algorithm allow us to identify. In particular, we show that the conditional local average treatment effect (LATE; Angrist, Imbens and Rubin, 1996) at covariate value $x$ is identified if and only if the Approximate Propensity Score is nondegenerate, i.e., $p^A(x) \in (0, 1)$.

The identification analysis suggests an estimator. The treatment effects can be estimated by two-stage least squares (2SLS) where we regress the outcome on the treatment with the algorithm’s recommendation as an IV. To make the algorithmic recommendation a conditionally independent IV, we propose to control for the Approximate Propensity Score.
1. For small bandwidth $\delta > 0$ and a large number of simulation draws $S$, compute
\[
p^s(X_i; \delta) = \frac{1}{S} \sum_{s=1}^{S} A(X^*_{i,s}),
\]
where $X^*_{i,1}, \ldots, X^*_{i,S}$ are $S$ independent simulation draws from the uniform distribution on $B(X_i, \delta)$. This $p^s(X_i; \delta)$ is a simulation-based approximation to the Approximate Propensity Score $p^A(X_i)$.

2. Run this 2SLS regression for observations with $p^s(X_i; \delta) \in (0, 1)$:
\[
D_i = \gamma_0 + \gamma_1 Z_i + \gamma_2 p^s(X_i; \delta) + \nu_i \text{ (First Stage)}
\]
\[
Y_i = \beta_0 + \beta_1 D_i + \beta_2 p^s(X_i; \delta) + \epsilon_i \text{ (Second Stage)}.
\]

Let $\hat{\beta}_1^s$ be the estimated coefficient on $D_i$.

As the main theoretical result, we prove the 2SLS estimator $\hat{\beta}_1^s$ is a consistent and asymptotically normal estimator of a well-defined causal effect (weighted average of conditional local average treatment effects). Our result clarifies how to estimate other parameters by reweighting observations. We also show that inference based on the conventional 2SLS heteroskedasticity-robust standard errors is asymptotically valid as long as the bandwidth $\delta$ goes to zero and the number of simulation draws $S$ goes to infinity at appropriate rates. We prove the asymptotic properties by exploiting results from differential geometry and geometric measure theory.

Our estimator is applicable even if the algorithm is deterministic and produces multidimensional regression-discontinuity variation only. In contrast to standard multidimensional RDD methods that define the distance to the nearest boundary as a single running variable, our estimator is straightforward to use even when the multidimensional boundary is complex and the distance is hard to compute. Moreover, our method applies to more general settings with stochastic algorithms, deterministic algorithms, and combinations of the two. In settings that mix stochastic and deterministic algorithms, our estimator exploits both the random-assignment variation.

\[\text{We write these equations just to describe our 2SLS specification. These equations do not represent a structural model.}\]
and RDD variation, producing precision and representability gains compared to RDD and propensity-score estimators only using either variation.

The practical performance of our estimator is demonstrated through simulation and an original application. We first conduct a Monte Carlo simulation mimicking real-world decision making based on machine learning algorithms. We consider a data-generating process combining stochastic and deterministic algorithms. Treatment recommendations are randomly assigned for a small experimental segment of the population and are determined by a high-dimensional, deterministic machine learning algorithm for the rest of the population. Such combinations of small experiments and deterministic treatment allocations arise in the real world when large-scale experiments are prohibited due to ethical, budget, or legislative constraints. Our estimator is shown to be feasible in this high-dimensional setting and has smaller median absolute errors relative to alternative estimators. In particular, our method produces more efficient and representative estimates than a conventional propensity-score approach. The simulation also illustrates a setting where the RDD boundary is complex and no prior multidimensional RDD method is applicable.

Our empirical application is an analysis of COVID-19 hospital relief funding. The Coronavirus Aid, Relief, and Economic Security (CARES) Act designated $175 billion for COVID-19 response efforts and reimbursement to health care entities for expenses or lost revenues (Kakani, Chandra, Mullainathan and Obermeyer, 2020). This policy intended to help hospitals hit hard by the pandemic, as “financially insecure hospitals may be less capable of investing in COVID-19 response efforts” (Khullar, Bond and Schpero, 2020). We ask whether this problem is alleviated by the relief funding for hospitals.

We identify the causal effects of the relief funding by exploiting the funding eligibility rule. The government runs an algorithmic rule on hospital characteristics to decide which hospitals are eligible for funding. This fact allows us to apply our method to estimate the effect of relief funding. Specifically, our 2SLS estimators use funding eligibility status as an IV for funding amounts, while controlling for the Approximate Propensity Score induced by the eligibility-determining algorithm. The funding eligibility IV boosts the funding amount by about $15 million on average.

The resulting 2SLS estimates with Approximate Propensity Score controls suggest
that COVID-19 relief funding has little to no effect on outcomes, such as the number of COVID-19 patients hospitalized at each hospital. The estimated causal effects of relief funding are much smaller and less significant than the naive ordinary least squares (OLS) (with and without controlling for hospital characteristics) or 2SLS estimates with no controls. The OLS estimates, for example, imply that a $1 million increase in funding allows hospitals to accommodate 4.53 more COVID-19 patients. The uncontrolled 2SLS estimates produce similar, slightly smaller effects (2.44 more patients per $1 million of funding). In contrast, the 2SLS estimates with Approximate Propensity Score controls show no or even negative effects (up to 2.21 fewer patients for every $1 million of funding).

The null effect of funding persists several months after the distribution of funding. We also find no clear heterogeneity in the null funding effect across different subgroups of hospitals. Our finding provides causal evidence for the concern that funding in the CARES Act might not have been well targeted to the clinics and hospitals with the greatest needs (Kakani et al., 2020).

Related Literature. Our framework integrates the classic propensity-score (selection-on-observables) scenario with a multidimensional extension of the fuzzy RDD. We analyze this integrated setup in the IV world with noncompliance. This general setting appears to have no prior established estimator. Armstrong and Kolesár (2021) provide an estimator for a related setting with perfect compliance.

When we specialize our estimator to the multidimensional RDD case, our estimator has three features. First, it is a consistent and asymptotically normal estimator of a well-interpreted causal effect (average of conditional treatment effects along the RDD boundary) even if treatment effects are heterogeneous. Second, it uses observations near all the boundary points as opposed to using only observations near one specific boundary point, thus avoiding variance explosion even when \( X_i \) has many elements. Third, it can be easily implemented even in cases with many covariates and complex algorithms (RDD boundaries). No existing estimator appears to have all of these properties (Papay, Willett and Murnane, 2011; Zajonc, 2012; Keele and Titiunik, 2015; Cattaneo, Titiunik, Vazquez-Bare and Keele, 2016; Imbens and Wager, 2019).

A popular approach to the two-dimensional RDD is to use the shortest (Euclidean)
distance from each individual to the boundary as a univariate running variable and apply a univariate RDD method. In general multidimensional RDDs where the boundary is complex or its analytical form is unknown, the distance-based approach requires approximation or estimation of the distance to the boundary. In such cases, no distance-based estimator has been proven to have properties such as consistency and asymptotic normality. We provide the asymptotic properties of our simulation-based estimator for a class of multidimensional fuzzy RDDs, taking into account the simulation errors.

Our estimator is applicable to a class of data-generating algorithms that includes stochastic and deterministic algorithms used in practice. Our results thus nest existing insights on quasi-experimental variation in particular algorithms, such as supervised learning (Cowgill, 2018; Bundorf, Polyakova and Tai-Seale, 2024), bandit, reinforcement learning, and market-design algorithms (Abdulkadiroğlu, Angrist, Narita and Pathak, 2017, 2022; Abdulkadiroğlu, 2013; Narita, 2021a,b; Chen, 2023). Our framework also reveals new sources of identification for algorithms that, at first sight, do not appear to produce a natural experiment.

The Approximate Propensity Score in this paper is related to the local random assignment interpretation of the RDD, discussed by Cattaneo, Frandsen and Titunik (2015), Sekhon and Titunik (2017), Frölich and Huber (2019), and Abdulkadiroğlu et al. (2022). These papers consider special cases of this paper’s framework. For example, Abdulkadiroğlu et al. (2022) focus on Gale and Shapley’s deferred acceptance matching algorithm and its variants. In contrast, this paper’s method is applicable to generic algorithms, including complex machine-learning algorithms and multidimensional RDDs.

---

2 Another common approach to the two-dimensional RDD is to first estimate the conditional average treatment effect for a large number of boundary points (either by the univariate local polynomial regression using the distance to the point as a univariate covariate or by the bivariate local polynomial regression). It then computes a weighted average of the estimated conditional average treatment effects over the boundary (Zajonc, 2012; Keele and Titunik, 2015). However, identifying boundary points from a general decision algorithm is hard unless it has a known analytical form. Even if we can trace out the boundary, it is not straightforward to select a grid of points along the boundary.
2 Framework

Our framework is a mix of the conditional independence, multidimensional RDD, and instrumental variable scenarios. In the setup in the introduction, we are interested in the effect of some binary treatment $D_i \in \{0, 1\}$ on some outcome of interest $Y_i \in \mathbb{R}$. We impose the exclusion restriction that the treatment recommendation $Z_i \in \{0, 1\}$ does not affect the observed outcome other than through the treatment assignment $D_i$. This allows us to define the potential outcomes indexed against the treatment assignment $D_i$ alone. $Y_i(1)$ and $Y_i(0)$ denote potential outcomes when the individual is treated and not treated, respectively.

We consider algorithms that make treatment recommendations based on individual $i$’s predetermined, observable covariates $X_i = (X_{i1}, ..., X_{ip})' \in \mathbb{R}^p$. Let the function $A : \mathbb{R}^p \rightarrow [0, 1]$ represent the decision algorithm, where $A(X_i) = \Pr(Z_i = 1|X_i)$ is the probability that the treatment is recommended for individual $i$ with covariates $X_i$. The central assumption is that the analyst knows function $A$ and is able to simulate it. That is, the analyst is able to compute the recommendation probability $A(x)$ given any input value $x \in \mathbb{R}^p$. The treatment recommendation $Z_i$ for individual $i$ is then randomly determined with probability $A(X_i)$ independently of everything else. Consequently, the following conditional independence holds.

**Property 1 (Conditional Independence).** $Z_i \perp \perp (Y_i(1), Y_i(0), D_i(1), D_i(0))|X_i$.

The codomain of $A$ contains 0 and 1, allowing for deterministic treatment assignments conditional on $X_i$. Our framework therefore nests the RDD as a special case. Another special case is the classic conditional independence scenario with the common support condition ($A(X_i) \in (0, 1)$ almost surely). In addition to these simple settings, this framework nests many other situations, such as multidimensional RDDs and complex machine learning and market-design algorithms, as illustrated in Sections 5–6 (see also Appendix F of a preprint of this paper (Narita and Yata, 2023)).

In typical machine-learning scenarios, an algorithm first applies machine learning on $X_i$ to make some prediction and then uses the prediction to output the recommendation probability $A(X_i)$, as in the following example.
Example. Automated disease detection algorithms use machine learning to detect various diseases and to identify patients at risk (Gulshan et al., 2016). A detection algorithm predicts whether an individual $i$ has a disease ($Z_i = 1$) or not ($Z_i = 0$) based on a digital image $X_i \in \mathbb{R}^p$ of the individual’s body, where each $X_{ij} \in \mathbb{R}$ denotes the intensity value of a pixel in the image. The algorithm uses training data to construct a classifier $A : \mathbb{R}^p \rightarrow \{0, 1\}$. The classifier takes an image of individual $i$ as input and makes a binary prediction of whether the individual has the disease: $Z_i \equiv A(X_i)$. The algorithm’s diagnosis $Z_i$ may influence the doctor’s treatment decision for the individual, denoted by $D_i \in \{0, 1\}$. We are interested in how the treatment decision $D_i$ affects the individual’s health outcome $Y_i$.

Let $Y_{zi}$ be defined as $Y_{zi} \equiv D_i(z)Y_i(1) + (1 - D_i(z))Y_i(0)$ for $z \in \{0, 1\}$. $Y_{zi}$ is the potential outcome when the treatment recommendation is $Z_i = z$. It follows from Property 1 that $Z_i \perp (Y_{1i}, Y_{0i}) | X_i$. We put assumptions on the covariates $X_i$ and the algorithm $A$. For simplicity, we assume that the distribution of $X_i$ is absolutely continuous with respect to the Lebesgue measure. The analysis extends to the case where some covariates in $X_i$ are discrete if we condition on the discrete covariates (see Appendix E.2 of Narita and Yata (2023)). Let $\mathcal{X}$ be the support of $X_i$, $\mathcal{X}_0 = \{x \in \mathcal{X} : A(x) = 0\}$, $\mathcal{X}_1 = \{x \in \mathcal{X} : A(x) = 1\}$, $\mathcal{L}^p$ be the Lebesgue measure on $\mathbb{R}^p$, and $\text{int}(S)$ be the interior of a set $S \subset \mathbb{R}^p$.

Assumption 1.

(a) $A$ is continuous almost everywhere with respect to the Lebesgue measure.

(b) (Measure Zero Boundaries of $\mathcal{X}_0$ and $\mathcal{X}_1$) $\mathcal{L}^p(\mathcal{X}_k) = \mathcal{L}^p(\text{int}(\mathcal{X}_k))$ for $k = 0, 1$.

Assumption 1 (a) allows the function $A$ to be discontinuous on a set of points with the Lebesgue measure zero. For example, $A$ is allowed to be a discontinuous step function as long as it is continuous almost everywhere. Assumption 1 (b) holds if the Lebesgue measures of the boundaries of $\mathcal{X}_0$ and $\mathcal{X}_1$ are zero. Assumption 1 (b) is only for ruling out perverse cases such as the case where $A(x) = 1$ if $x \in \mathbb{R}$ is an irrational number and $A(x) \neq 1$ otherwise.
3 Identification

What causal effects can be learned from data \((Y_i, X_i, D_i, Z_i)\) generated by the algorithm \(A\)? A key step toward answering this question is the \textit{Approximate Propensity Score} (APS). To define it, we first define the fixed-bandwidth \textit{Approximate Propensity Score} as follows:

\[
p^A(x; \delta) \equiv \frac{\int_{B(x, \delta)} A(x^*) \, dx^*}{\int_{B(x, \delta)} \, dx^*},
\]

where \(B(x, \delta) = \{x^* \in \mathbb{R}^p : \|x - x^*\| < \delta\}\) is the (open) \(\delta\)-ball around \(x \in \mathcal{X}\). Here, \(\| \cdot \|\) denotes the Euclidean norm on \(\mathbb{R}^p\). To make a common bandwidth \(\delta\) for all dimensions reasonable, we standardize \(X_{ij}\) to have variance one for each \(j = 1, \ldots, p\). We assume that \(A\) is a \(L^p\)-measurable function so that the integrals exist. We then define APS as follows:

\[
p^A(x) \equiv \lim_{\delta \to 0} p^A(x; \delta).
\]

APS at \(x\) is the average probability of a treatment recommendation in a shrinking ball around \(x\). We call this the \textit{Approximate} Propensity Score, since this score modifies the standard propensity score \(A(X_i)\) to incorporate local variation in the score. APS exists for most covariate points and algorithms (see Appendix C.3). Figure 1 illustrates APS. In the example, \(X_i\) is two dimensional, and the support of \(X_i\) is divided into three sets depending on the value of \(A\). For the interior points of each set, APS is equal to \(A\). On the border of any two sets, APS is the average of the \(A\) values in the two sets. Thus, \(p^A(x) = (0 + 0.5)/2 = 0.25\) for any \(x\) in the open line segment \(AB\), \(p^A(x) = (0.5 + 1)/2 = 0.75\) for any \(x\) in the open line segment \(BC\), and \(p^A(x) = (0 + 1)/2 = 0.5\) for any \(x\) in the open line segment \(BD\).

Our identification analysis uses the following continuity condition.

\textbf{Assumption 2 (Local Mean Continuity).} For \(z \in \{0, 1\}\), the conditional expectation functions \(E[Y_{zi}|X_i]\) and \(E[D_i(z)|X_i]\) are continuous at any point \(x \in \mathcal{X}\) such that

\[\text{We use a ball for simplicity. When we instead use a rectangle, ellipsoid, or any standard kernel function to define } p^A(x; \delta), \text{ the limit } \lim_{\delta \to 0} p^A(x; \delta) \text{ may be different at some points (e.g., at discontinuity points of } A), \text{ but the same identification results hold under suitable conditions.}\]
Assumption 2 is a multivariate extension of the local mean continuity condition frequently assumed in the RDD. In the RDD with a single running variable, the point \( x \) for which \( p^A(x) \in (0, 1) \) and \( A(x) \in \{0, 1\} \) is the cutoff at which the treatment probability discontinuously changes. \( A(x) \in \{0, 1\} \) means that the treatment recommendation \( Z_i \) is deterministic conditional on \( X_i = x \). If APS at the point \( x \) is nondegenerate (\( p^A(x) \in (0, 1) \)), however, there exists a point close to \( x \) that has a different value of \( A \) from \( x \)'s, which creates variation in the treatment recommendation near \( x \). For any such point \( x \), Assumption 2 requires that the points close to \( x \) have similar conditional means of the outcome \( Y_{zi} \) and treatment assignment \( D_i(z) \).

Note that Assumption 2 does not require continuity of the conditional means at \( x \) for which \( A(x) \in (0, 1) \), since the identification of the conditional means at such points follows from Property 1 without continuity.

Under the above assumptions, APS provides an easy-to-check condition for whether an algorithm allows us to identify causal effects. We say that a causal effect is identified if it is uniquely determined by the joint distribution of \((Y_i, X_i, D_i, Z_i)\).

**Proposition 1 (Identification).** Under Assumptions 1 and 2:
(a) $E[Y_{1i} - Y_{0i}|X_i = x]$ and $E[D_i(1) - D_i(0)|X_i = x]$ are identified for every $x \in \text{int}(\mathcal{X})$ such that $p^A(x) \in (0, 1)$.

(b) Let $S$ be any open subset of $\mathcal{X}$ such that $p^A(x)$ exists for all $x \in S$. Then either $E[Y_{1i} - Y_{0i}|X_i \in S]$ or $E[D_i(1) - D_i(0)|X_i \in S]$ or both are identified only if $p^A(x) \in (0, 1)$ for almost every $x \in S$ (with respect to the Lebesgue measure).

Proof. See Appendix D.1. \qed

Proposition 1 characterizes a necessary and sufficient condition for identification. Part (a) says that the average effects of the treatment recommendation $Z_i$ on the outcome $Y_i$ and on the treatment assignment $D_i$ conditional on $X_i = x$ are both identified if APS at $x$ is neither 0 nor 1. Nondegeneracy of APS at $x$ implies that there are both types of individuals who receive $Z_i = 1$ and $Z_i = 0$ among those whose $X_i$ is close to $x$. Assumption 2 ensures that these individuals are similar in terms of average potential outcomes and treatment assignments. We can therefore identify the average effects conditional on $X_i = x$. In Figure 1, $p^A(x) \in (0, 1)$ holds for any $x$ in the shaded region.

Part (b) provides a necessary condition for identification. It says that if the average effect of the treatment recommendation conditional on $X_i$ being in some open set $S$ is identified, then we must have $p^A(x) \in (0, 1)$ for almost every $x \in S$. If, to the contrary, there is a subset of $S$ of nonzero measure for which $p^A(x) = 1$ (or $p^A(x) = 0$), then $Z_i$ has no variation in the subset, which makes it impossible to identify the average effect for the subset.

Proposition 1 concerns causal effects of treatment recommendation, not of treatment assignment. The proposition implies that the conditional average treatment effects and the conditional local average treatment effects are identified under additional assumptions.

**Corollary 1.** Under Assumptions 1 and 2:

(a) The conditional average treatment effect, $E[Y_i(1) - Y_i(0)|X_i = x]$, is identified for every $x \in \text{int}(\mathcal{X})$ with $p^A(x) \in (0, 1)$ and $\Pr(D_i(1) > D_i(0)|X_i = x) = 1$ (perfect compliance).
(b) The conditional local average treatment effect, \(E[Y_i(1) - Y_i(0)|D_i(1) \neq D_i(0), X_i = x]\), is identified for every \(x \in \text{int}(\mathcal{X})\) such that \(p^A(x) \in (0, 1)\), \(\Pr(D_i(1) \geq D_i(0)|X_i = x) = 1\) (monotonicity), and \(\Pr(D_i(1) \neq D_i(0)|X_i = x) > 0\) (existence of compliers).

Nondegeneracy of APS \(p^A(x)\) therefore summarizes what causal effects the data from \(A\) identify. The key condition \((p^A(x) \in (0, 1))\) holds for some points \(x\) for every standard algorithm except trivial algorithms that always recommend a treatment with probability 0 or 1. Therefore, the data from every nondegenerate algorithm identify some causal effect.

## 4 Estimation

The sources of quasi-random assignment characterized in Proposition 1 suggest a way of estimating causal effects of the treatment. In view of Proposition 1, it is possible to nonparametrically estimate conditional average causal effects \(E[Y_i(1) - Y_i(0)|X_i = x]\) and \(E[D_i(1) - D_i(0)|X_i = x]\) for points \(x\) such that \(p^A(x) \in (0, 1)\). This approach is hard to use in practice, however, when \(X_i\) has many elements. We instead seek an estimator that aggregates conditional effects at different points into a single average causal effect. Proposition 1 suggests that conditioning on APS makes algorithm-based treatment recommendation quasi-randomly assigned. This motivates the use of an algorithm’s recommendation as an instrument conditional on APS, which we operationalize as follows.

### 4.1 Two-Stage Least Squares Meets APS

Suppose that we observe a random iid sample \(\{(Y_i, X_i, D_i, Z_i)\}_{i=1}^n\) of size \(n\) from the population whose data-generating process is as described in the introduction and Section 2. Consider the following 2SLS regression using the observations with \(p^A(X_i; \delta_n) \in (0, 1)\):

\[
\begin{align*}
D_i &= \gamma_0 + \gamma_1 Z_i + \gamma_2 p^A(X_i; \delta_n) + \nu_i \tag{1} \\
Y_i &= \beta_0 + \beta_1 D_i + \beta_2 p^A(X_i; \delta_n) + \epsilon_i \tag{2}
\end{align*}
\]
where bandwidth $\delta_n$ shrinks toward zero as the sample size $n$ increases. We drop the constant term if $A(X_i)$ takes only one nondegenerate value in the sample. Let $I_{i,n} = 1\{p^A(X_i; \delta_n) \in (0, 1)\}$, $D_{i,n} = (1, D_i, p^A(X_i; \delta_n))'$, and $Z_{i,n} = (1, Z_i, p^A(X_i; \delta_n))'$. The 2SLS estimator $\hat{\beta}$ is then given by

$$
\hat{\beta} = \left( \sum_{i=1}^{n} Z_{i,n} D_{i,n}' I_{i,n} \right)^{-1} \sum_{i=1}^{n} Z_{i,n} Y_i I_{i,n}.
$$

Let $\hat{\beta}_1$ denote the 2SLS estimator of $\beta_1$ in the above regression.

The above regression uses true fixed-bandwidth APS $p^A(X_i; \delta_n)$, but it may be difficult to analytically compute if $A$ is complex. In such a case, we propose to approximate $p^A(X_i; \delta_n)$ using brute force simulation. We draw a value of $x$ from the uniform distribution on $B(X_i, \delta_n)$ a number of times, compute $A(x)$ for each draw, and take the average of $A(x)$ over the draws. Formally, let $X_{i,1}^*, ..., X_{i,S_n}^*$ be $S_n$ independent draws from the uniform distribution on $B(X_i, \delta_n)$, and calculate

$$
p^s(X_i; \delta_n) = \frac{1}{S_n} \sum_{s=1}^{S_n} A(X_{i,s}^*).
$$

We compute $p^s(X_i; \delta_n)$ for each $i = 1, ..., n$ independently across $i$ so that $p^s(X_1; \delta_n), ..., p^s(X_n; \delta_n)$ are independent of each other. For fixed $n$ and $X_i$, the approximation error relative to true $p^A(X_i; \delta_n)$ has a $1/\sqrt{S_n}$ rate of convergence. This rate does not depend on the dimension of $X_i$, so the simulation error can be made negligible even when $X_i$ has many elements. We consider the simulation version of the 2SLS regression (1) and (2), where we use the simulated fixed-bandwidth APS $p^s(X_i; \delta_n)$ in place of $p^A(X_i; \delta_n)$. Let $\hat{\beta}_1^s$ denote the 2SLS estimator of $\beta_1$ in the simulation-based regression.

An alternative 2SLS regression is the one controlling for the variable that equals $A(X_i)$ for observations with $A(X_i) \in (0, 1)$ and $p^s(X_i; \delta_n)$ for those with $A(X_i) \in \{0, 1\}$. This allows us to avoid any simulation for those with $A(X_i) \in (0, 1)$. Moreover, this prevents the finite-sample bias due to using fixed-bandwidth APS instead of the standard propensity score for those with $A(X_i) \in (0, 1)$. This modification does not affect the asymptotic results below.

Our method is a simple way to exploit the whole quasi-experimental variation.
embedded in the data. In settings that mix stochastic and deterministic algorithms, our estimator uses both the random-assignment part (i.e., observations with \(A(X_i) \in (0,1)\)) and regression-discontinuity part (i.e., observations with \(A(X_i) \in \{0,1\}\) and \(p^A(X_i; \delta_n) \in (0,1)\)). If the researcher is interested in estimating causal effects for the two parts separately, it is possible to split the data into the two parts and run the 2SLS regression separately.

4.2 Consistency and Asymptotic Normality

We establish the consistency and asymptotic normality of the 2SLS estimators \(\hat{\beta}_1\) and \(\hat{\beta}_s\). We use the following assumptions.

**Assumption 3.**

(a) (Finite Moment) \(E[Y_i^4] < \infty\).

Let \(f_X\) denote the probability density function of \(X_i\) and let \(H^k\) denote the \(k\)-dimensional Hausdorff measure on \(\mathbb{R}^p\).

(b) (Nonzero First Stage) \(\int X p^A(x)(1-p^A(x))E[D_i(1)-D_i(0)|X_i = x]f_X(x)d\mu(x) \neq 0\), where \(\mu\) is the Lebesgue measure \(L^p\) when \(\Pr(A(X_i) \in (0,1)) > 0\) and is the \((p-1)\)-dimensional Hausdorff measure \(H^{p-1}\) when \(\Pr(A(X_i) \in (0,1)) = 0\).

If \(\Pr(A(X_i) \in (0,1)) = 0\), then the following conditions (c)–(f) hold.

(c) (Nonzero Variance) \(\text{Var}(A(X_i)) > 0\).

For a set \(S \subset \mathbb{R}^p\), let \(\text{cl}(S)\) denote the closure of \(S\) and let \(\partial S\) denote the boundary of \(S\), i.e., \(\partial S = \text{cl}(S) \setminus \text{int}(S)\).

(d) (\(C^2\) Boundary of \(\Omega^*\)) There exists a partition \(\{\Omega^*_1, ..., \Omega^*_M\}\) of \(\Omega^* = \{x \in \mathbb{R}^p : A(x) = 1\}\) such that

\[\]
(i) \(\text{dist}(\Omega_m^*, \Omega_{m'}^*) > 0\) for any \(m, m' \in \{1, \ldots, M\}\) such that \(m \neq m'\). Here \(\text{dist}(S, T) = \inf_{x \in S, y \in T} \|x - y\|\) is the distance between two sets \(S\) and \(T \subset \mathbb{R}^p\);

(ii) \(\Omega_m^*\) is nonempty, bounded, open, connected, and twice continuously differentiable for each \(m \in \{1, \ldots, M\}\). Here we say that a bounded open set \(S \subset \mathbb{R}^p\) is twice continuously differentiable if for every \(x \in S\), there exists a ball \(B(x, \epsilon)\) and a one-to-one mapping \(\psi\) from \(B(x, \epsilon)\) onto an open set \(D \subset \mathbb{R}^p\) such that \(\psi\) and \(\psi^{-1}\) are twice continuously differentiable, \(\psi(B(x, \epsilon) \cap S) \subset \{(x_1, \ldots, x_p) \in \mathbb{R}^p : x_p > 0\}\), and \(\psi(B(x, \epsilon) \cap \partial S) \subset \{(x_1, \ldots, x_p) \in \mathbb{R}^p : x_p = 0\}\).

(e) (Regularity of Deterministic \(A\))

(i) \(\mathcal{H}^{p-1}(\partial \Omega^*) < \infty\), and \(\int_{\partial \Omega^*} f_X(x) d\mathcal{H}^{p-1}(x) > 0\);

(ii) There exists \(\delta > 0\) such that \(A(x) = 0\) for almost every \(x \in N(X, \delta) \setminus \Omega^*\), where \(N(S, \delta) = \{x \in \mathbb{R}^p : \|x - y\| < \delta\ \text{for some} \ y \in S\}\) for a set \(S \subset \mathbb{R}^p\) and \(\delta > 0\).

(f) (Conditional Moments and Density Near \(\partial \Omega^*\)) There exists \(\delta > 0\) such that

(i) \(E[Y_{i1}|X_i], E[Y_{0i}|X_i], E[D_i(1)|X_i], E[D_i(0)|X_i]\) and \(f_X\) are continuously differentiable and have bounded partial derivatives on \(N(\partial \Omega^*, \delta)\);

(ii) \(E[Y_{i1}^2|X_i], E[Y_{0i}^2|X_i], E[Y_{i1}D_i(1)|X_i]\) and \(E[Y_{0i}D_i(0)|X_i]\) are continuous on \(N(\partial \Omega^*, \delta)\);

(iii) \(E[Y_{i1}^4|X_i]\) is bounded on \(N(\partial \Omega^*, \delta)\).

Assumption 3 is a set of conditions for establishing consistency. Assumption 3 (b) assumes that the weighted average effect of the algorithm’s recommendation on the treatment assignment is nonzero. Under this assumption, the estimated first-stage coefficient on \(Z_i\) converges to a nonzero quantity. Assumptions 3 (c)–(f) are a set of conditions we require for proving consistency and asymptotic normality of \(\hat{\beta}_1\) when \(A\) is deterministic and produces only multidimensional regression-discontinuity variation. Assumption 3 (c) says that \(A\) produces variation in the treatment recommendation.
Assumption 3 (d) imposes the differentiability of the boundary of \( \Omega^* = \{ x \in \mathbb{R}^p : A(x) = 1 \} \). Our 2SLS regression uses the observations with \( p^A(X_i; \delta_n) \in (0, 1) \) (or \( p^s(X_i; \delta_n) \in (0, 1) \) when we use the simulated fixed-bandwidth APS) only. By definition, if \( p^A(X_i; \delta) \in (0, 1) \), \( X_i \) must be in the \( \delta \)-neighborhood of the boundary of \( \Omega^* \). We use Assumption 3 (d) to derive the limits and convergence rates of the integrals of functions of \( x \) over the \( \delta \)-neighborhood as \( \delta \) shrinks to zero. For example, we show that \( \delta^{-1}E[g(X_i)1\{p^A(X_i; \delta) \in (0, 1)\}] = 2 \int_{\partial \Omega^*} g(x)f_X(x)dH^{p-1}(x) + O(\delta) \) as \( \delta \to 0 \) if \( g \) and \( f_X \) are continuously differentiable and have bounded partial derivatives near the boundary \( \partial \Omega^* \). See Step B.1.3 in Appendix B.1 for the results. The conditions are satisfied if, for example, \( \Omega^* = \{ x \in \mathbb{R}^p : f(x) \geq 0 \} \) for some twice continuously differentiable function \( f : \mathbb{R}^p \to \mathbb{R} \) such that \( \nabla f(x) = \left( \frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_p} \right)' \neq 0 \) for all \( x \in \mathbb{R}^p \) with \( f(x) = 0 \). \( \Omega^* \) takes this form when supervised learning based on smooth models is used to construct a binary classifier \( A \) such that \( A(x) = 1\{f(x) \geq 0\} \), \( x \in \mathbb{R}^p \).

In general, the differentiability of \( \Omega^* \) may not hold. For example, if tree-based algorithms such as Classification And Regression Tree (CART) and random forests are used to construct a classifier \( A(x) = 1\{f(x) \geq 0\} \), then the function \( f \) is not differentiable at some points. Yet, the assumptions approximately hold in that \( \Omega^* \) is arbitrarily well approximated by a set that satisfies the differentiability condition.

Part (i) of Assumption 3 (e) says that the boundary of \( \Omega^* \) is \( (p-1) \)-dimensional and that the boundary has nonzero density. Part (ii) puts a weak restriction on the values \( A \) takes on outside the support of \( X_i \). It requires that \( A(x) = 0 \) for almost every \( x \notin \Omega^* \) outside \( \mathcal{X} \) but in the neighborhood of \( \mathcal{X} \). \( A(x) \) may take on any value if \( x \) is not close to \( \mathcal{X} \). This assumption ensures that observations near the boundary of \( \mathcal{X} \) but distant from the boundary of \( \Omega^* \) have \( p^A(X_i; \delta_n) \in \{0, 1\} \) and are excluded from the regression. These conditions hold in practice. Assumption 3 (f) imposes continuity, continuous differentiability, and boundedness on the conditional moments of potential outcomes and the probability density near the boundary of \( \Omega^* \). Note that Part (i) of Assumption 3 (f) implies Assumption 2.

\footnote{For example, suppose that \( \mathcal{X} = [-1, 1] \), \( A(x) = 1 \) if \( x > 0 \), \( A(x) = 0 \) if \( x \in [-1, 0] \), and \( A(x) = \frac{1}{2} \) if \( x < -1 \) (note that \( \Pr(A(X_i) \in (0, 1)) = 0 \) in this case). In this case, for any sufficiently small \( \delta > 0 \), \( p^A(X_i; \delta) \in (0, 1) \) if \( X_i \in [-1, -1 + \delta] \), but these observations should not be included in the regression as they are not close to the boundary of \( \Omega^* = (0, \infty) \). Such a case is ruled out by Part (ii) of Assumption 3 (e).}
Under the above conditions and technical regularity Assumptions A.1 and A.2 in Appendix A, the 2SLS estimators $\hat{\beta}_1$ and $\hat{\beta}_s^1$ are consistent and asymptotically normal estimators of a weighted average treatment effect.\textsuperscript{6}

**Theorem 1** (Consistency and Asymptotic Normality). Suppose that Assumptions 1 and 3 hold and $\delta_n \rightarrow 0$, $n\delta_n \rightarrow \infty$, and $S_n \rightarrow \infty$ as $n \rightarrow \infty$. Then the 2SLS estimators $\hat{\beta}_1$ and $\hat{\beta}_s^1$ converge in probability to

$$\beta_1 \equiv \lim_{\delta \rightarrow 0} E[\omega_1(\delta)(Y_i(1) - Y_i(0))],$$

where

$$\omega_1(\delta) = \frac{p^A(X_i; \delta)(1 - p^A(X_i; \delta))(D_i(1) - D_i(0))}{E[p^A(X_i; \delta)(1 - p^A(X_i; \delta))(D_i(1) - D_i(0))]},$$

Suppose, in addition, that Assumptions A.1 and A.2 in Appendix A hold and $n\delta_n^2 \rightarrow 0$ as $n \rightarrow \infty$. Then

$$\hat{\sigma}_n^{-1}(\hat{\beta}_1 - \beta_1) \overset{d}{\rightarrow} \mathcal{N}(0, 1),$$

$$(\hat{\sigma}_n^s)^{-1}(\hat{\beta}_s^1 - \beta_1) \overset{d}{\rightarrow} \mathcal{N}(0, 1),$$

where we define $\hat{\sigma}_n^{-1}$ and $(\hat{\sigma}_n^s)^{-1}$ as follows. Let $\hat{\epsilon}_{i,n} = Y_i - D_i'n\hat{\beta}$ and

$$\hat{\Sigma}_n = (\sum_{i=1}^n Z_{i,n}D_i'nI_{i,n})^{-1}(\sum_{i=1}^n \hat{\epsilon}_{i,n}^2Z_{i,n}Z_{i,n}'I_{i,n})(\sum_{i=1}^n D_{i,n}Z_{i,n}'I_{i,n})^{-1}.$$

$\hat{\Sigma}_n$ is the conventional heteroskedasticity-robust estimator for the variance of the 2SLS estimator. $\hat{\sigma}_n^2$ is the second diagonal element of $\hat{\Sigma}_n$. $(\hat{\sigma}_n^s)^2$ is the analogously-defined estimator for the variance of $\hat{\beta}_s^1$ from the simulation-based regression.

**Proof.** See Appendices B and D.2

---

\textsuperscript{6}Assumption A.1 is a set of additional regularity conditions (such as the smoothness of $A$) for proving asymptotic normality of $\hat{\beta}_1$ when $A$ is stochastic ($\Pr(A(X_i) \in (0, 1)) > 0$). Assumption A.2 imposes the condition on the growth rate of the number of simulation draws $S_n$, which we require for proving asymptotic normality of the simulation-based estimator $\hat{\beta}_s^1$. 

18
treatment recommendation \((D_i(1) \neq D_i(0))\). The limit \(\lim_{\delta \to 0} E[\omega_i(\delta)(Y_i(1) - Y_i(0))]\) always exists under the assumptions of Theorem 1. It is possible to estimate other weighted averages and the unweighted average by reweighting different observations appropriately. For example, we can estimate the unweighted average treatment effect by weighting observations by the inverse of fixed-bandwidth APS. Under monotonicity \((\Pr(D_i(1) \geq D_i(0)|X_i) = 1)\), we could also apply Abadie (2003)’s Kappa weighting method using fixed-bandwidth APS instead of the standard propensity score to estimate other weighted averages of treatment effects for compliers (see also Sloczyński (2022); Sloczyński, Uysal and Wooldridge (2024)).

Theorem 1 also shows that inference based on the conventional 2SLS heteroskedasticity-robust standard errors is asymptotically valid if \(\delta_n\) goes to zero at an appropriate rate. The convergence rate of \(\hat{\beta}_1\) is \(O_p(1/\sqrt{n})\) if \(\Pr(A(X_i) \in (0,1)) > 0\) and is \(O_p(1/\sqrt{n\delta_n})\) if \(\Pr(A(X_i) \in (0,1)) = 0\).

Our consistency result requires that \(\delta_n\) go to zero slower than \(n^{-1}\). The rate condition ensures that, when \(\Pr(A(X_i) \in (0,1)) = 0\), we have sufficiently many observations in the \(\delta_n\)-neighborhood of the boundary of \(\Omega^*\). Importantly, the rate condition does not depend on the dimension of \(X_i\), unlike other bandwidth-based estimation methods such as kernel methods. This is because we use all the observations in the \(\delta_n\)-neighborhood of the boundary, and the number of those observations is of order \(n\delta_n\) regardless of the dimension of \(X_i\) if the dimension of the boundary is \(p - 1\), i.e., the dimension of \(X_i\) minus one. When \(\Pr(A(X_i) \in (0,1)) > 0\), this rate condition is not necessary since the effective sample size always goes to infinity at the rate \(n\) regardless of the value of \(\delta_n\).

The asymptotic normality requires that \(\delta_n\) go to zero sufficiently quickly so that \(n\delta_n^2 \to 0\). When \(\Pr(A(X_i) \in (0,1)) > 0\), we need to use a small enough \(\delta_n\) so that \(p^A(X_i; \delta_n)\) converges to \(p^A(X_i)\) fast enough. When \(\Pr(A(X_i) \in (0,1)) = 0\), the asymptotic normality is based on undersmoothing, which eliminates the asymptotic bias by using the observations sufficiently close to the boundary of \(\Omega^*\). In both cases, the bias of our estimator is \(O(\delta_n)\). The standard deviation is \(O(1/\sqrt{n})\) when \(\Pr(A(X_i) \in (0,1)) > 0\) and is \(O(1/\sqrt{n\delta_n})\) when \(\Pr(A(X_i) \in (0,1)) = 0\). The condition that \(n\delta_n^2 \to 0\) ensures that the bias converges to zero faster than the standard deviation in either case. When \(\Pr(A(X_i) \in (0,1)) = 0\), a weaker condition that \(n\delta_n^3 \to 0\)
is sufficient for eliminating the asymptotic bias. When \( \Pr(A(X_i) \in (0, 1)) > 0 \), it is possible to relax the rate condition that \( n\delta_n^2 \to 0 \) at the cost of strengthening the smoothness of \( A \) in Assumption A.1.

Whether or not \( \Pr(A(X_i) \in (0, 1)) = 0 \), when we use simulated fixed-bandwidth APS, the consistency result requires that the number of simulation draws \( S_n \) go to infinity as \( n \) increases. The asymptotic normality result requires a sufficiently fast growth rate of \( S_n \) (satisfying Assumption A.2) to make the bias caused by using \( p^A(X_i; \delta_n) \) negligible.

Finally, the weight \( \omega_i(\delta) \) given in Theorem 1 is negative if \( D_i(1) < D_i(0) \), so \( E[\omega_i(\delta)(Y_i(1) - Y_i(0))] \) may not be a convex combination of treatment effects \( Y_i(1) - Y_i(0) \). This can happen because the treatment effect of those whose treatment assignment switches from 1 to 0 in response to the treatment recommendation (i.e., defiers) negatively contributes to \( E[\omega_i(\delta)(Y_i(1) - Y_i(0))] \). Additional assumptions prevent this problem. If the treatment effect is constant, for example, the 2SLS estimators are consistent for the treatment effect.

**Corollary 2.** Suppose that Assumptions 1 and 3 hold, that the treatment effect is constant, i.e., \( Y_i(1) - Y_i(0) = b \) for some constant \( b \), and that \( \delta_n \to 0, n\delta_n \to \infty \), and \( S_n \to \infty \) as \( n \to \infty \). Then the 2SLS estimators \( \hat{\beta}_1 \) and \( \hat{\beta}^s_1 \) converge in probability to \( b \).

Another approach is to impose monotonicity (Angrist et al., 1996). Let \( \text{LATE}(x) = E[Y_i(1) - Y_i(0)|D_i(1) \neq D_i(0), X_i = x] \) be the local average treatment effect (LATE) conditional on \( X_i = x \).

**Corollary 3.** Suppose that Assumptions 1 and 3 hold, that \( \Pr(D_i(1) \geq D_i(0)|X_i = x) = 1 \) for any \( x \in \mathcal{X} \) with \( p^A(x) \in (0, 1) \) (monotonicity), and that \( \delta_n \to 0, n\delta_n \to \infty \), and \( S_n \to \infty \) as \( n \to \infty \). Then the 2SLS estimators \( \hat{\beta}_1 \) and \( \hat{\beta}^s_1 \) converge in probability to

\[
\lim_{\delta \to 0} E[\omega(X_i; \delta)\text{LATE}(X_i)],
\]

where

\[
\omega(x; \delta) = \frac{p^A(x; \delta)(1 - p^A(x; \delta))E[D_i(1) - D_i(0)|X_i = x]}{E[p^A(X_i; \delta)(1 - p^A(X_i; \delta))(D_i(1) - D_i(0))]}.
\]

\footnote{When \( \Pr(A(X_i) \in (0, 1)) = 0 \) and \( n\delta_n^2 \) goes to some nonzero constant, our estimator can have a nonzero asymptotic bias (see Appendix C.2.3.1 of Narita and Yata (2023) for the asymptotic distribution).}
4.3 Special Cases

Theorem 1 holds whether \( A \) is stochastic (\( \Pr(A(X_i) \in (0,1)) > 0 \)) or deterministic (\( \Pr(A(X_i) \in (0,1)) = 0 \)). If we consider these two underlying cases separately, the probability limit of the 2SLS estimators has a more specific expression, as shown in the proof of Theorem 1 in Appendices B and D.2. If \( \Pr(A(X_i) \in (0,1)) > 0 \),

\[
\plim \hat{\beta}_1 = \plim \hat{\beta}_1^s = \frac{E[A(X_i)(1 - A(X_i))(D_i(1) - D_i(0))(Y_i(1) - Y_i(0))]}{E[A(X_i)(1 - A(X_i))(D_i(1) - D_i(0))]}.
\] (3)

The 2SLS estimators converge to a weighted average of treatment effects for the subpopulation with nondegenerate \( A(X_i) \).

To relate this result to existing work, consider the following 2SLS regression with the (standard) propensity score \( A(X_i) \) control:

\[
D_i = \gamma_0 + \gamma_1 Z_i + \gamma_2 A(X_i) + \nu_i \tag{4}
\]

\[
Y_i = \beta_0 + \beta_1 D_i + \beta_2 A(X_i) + \epsilon_i \tag{5}
\]

Under conditional independence, the 2SLS estimator from this regression converges in probability to the treatment-variance weighted average of treatment effects in (3) (Hull, 2018). Not surprisingly, for this selection-on-observables case, our result shows that the 2SLS estimator is consistent for the same treatment effect whether we control for the propensity score, fixed-bandwidth APS, or simulated fixed-bandwidth APS.

Importantly, using fixed-bandwidth APS as a control allows us to consistently estimate a causal effect even if \( A \) is deterministic and produces multidimensional regression-discontinuity variation. If \( \Pr(A(X_i) \in (0,1)) = 0 \),

\[
\plim \hat{\beta}_1 = \plim \hat{\beta}_1^s = \frac{\int_{\partial \Omega^*} E[(D_i(1) - D_i(0))(Y_i(1) - Y_i(0))|X_i = x]f_X(x)d\mathcal{H}^{p-1}(x)}{\int_{\partial \Omega^*} E[D_i(1) - D_i(0)|X_i = x]f_X(x)d\mathcal{H}^{p-1}(x)}.
\] (6)

The 2SLS estimators converge to a weighted average of treatment effects for the subpopulation who are on the boundary of the treated region.

The estimand in (6) nests parameters considered in the Regression Discontinuity (RD) literature. Under the monotonicity condition in Corollary 3, the estimand in
This estimand can be interpreted as the average treatment effect for the subpopulation of compliers on the boundary (Zajonc, 2012). In the univariate RDD with a single cutoff, this estimand further reduces to the average treatment effect for the compliers at the cutoff, the standard parameter in the fuzzy RDD (Hahn, Todd and van der Klaauw, 2001). If we instead assume $\Pr(D_i(1) > D_i(0) | X_i = x) = 1$ (perfect compliance) for all $x \in \partial \Omega^*$, the estimand in (6) equals

$$
\frac{\int_{\partial \Omega^*} LATE(x) E[D_i(1) - D_i(0) | X_i = x] f_X(x) d\mathcal{H}^{p-1}(x)}{\int_{\partial \Omega^*} E[D_i(1) - D_i(0) | X_i = x] f_X(x) d\mathcal{H}^{p-1}(x)}.
$$

This estimand represents the average treatment effect for the subpopulation who are on the boundary (Zajonc, 2012; Keele and Titiumik, 2015).

### 4.4 Comparison with Existing Approaches

To compare our method with existing ones, first consider the univariate RDD with a single cutoff $c$. In this special case, $p^A(X_i; \delta_n) = (X_i - c)/(2\delta_n) + 1/2$ if $X_i \in [c - \delta_n, c + \delta_n]$ and $p^A(X_i; \delta_n) \in \{0, 1\}$ otherwise. Therefore, the estimator $\hat{\beta}_1$ from the 2SLS regression (1) and (2) is numerically equivalent to a version of the RD local linear estimator (Hahn et al., 2001) that uses a box kernel and places the same slope coefficient of $X_i$ on both sides of the cutoff. When we use the bandwidth $\delta_n$ that converges to zero at the $n^{-1/3}$ rate instead of undersmoothing, our estimator achieves a convergence rate of $n^{-1/3}$. This rate is optimal for the estimation of the conditional LATE at the cutoff under our smoothness condition (continuous differentiability of $E[Y_1|X_i]$ and $E[Y_0|X_i]$ near the cutoff) in Assumption 3 (f).

Our approach is particularly useful in more general scenarios. One such scenario is the multidimensional sharp or fuzzy RDD when the boundary is complex or its analytical form is unknown. Our estimator is feasible for any decision boundary as long as we can simulate the underlying algorithm. More importantly, our estimator is shown to be consistent and asymptotically normal. As far as we know, there appear
to be no existing estimators that are computationally feasible and have theoretical validity for a general class of multidimensional RDDs. Moreover, our method is applicable to a more general setting that mixes stochastic and deterministic algorithms. We illustrate such a case in the next section.

5 Monte Carlo Simulation

This section assesses the feasibility and performance of our method. We do so through a Monte Carlo experiment motivated by decision making by machine learning with high-dimensional data. Consider a government or tech company that applies a machine-learning-based deterministic decision algorithm to a large segment of the population. At the same time, they conduct a randomized controlled trial (RCT) using the rest of the population. They are interested in estimating treatment effects using data from both segments. Our approach offers a way of exploiting not only the RCT segment but also the deterministic algorithm segment. Even if we focus on the deterministic algorithm segment, it is not straightforward to apply existing RDD methods to our simulation setup since the decision boundary is high dimensional and complex. We demonstrate the applicability of our approach in such a setup.

We simulate 1,000 hypothetical samples from the following data-generating process. Each sample \( \{(Y_i, X_i, D_i, Z_i)\}_{i=1}^{n} \) is of size \( n = 10,000 \). There are 100 covariates \( (p = 100) \), and \( X_i \sim \mathcal{N}(0, \Sigma) \). \( Y_i(0) \) is generated as \( Y_i(0) = 0.75X_i'\alpha_0 + 0.25\epsilon_{0i} \), where \( \alpha_0 \in \mathbb{R}^{100} \), and \( \epsilon_{0i} \sim \mathcal{N}(0, 1) \). We consider two models for \( Y_i(1) \), one in which the treatment effect \( Y_i(1) - Y_i(0) \) does not depend on \( X_i \) and one in which the treatment effect depends on \( X_i \).

Model A. \( Y_i(1) = Y_i(0) + \epsilon_{1i} \), where \( \epsilon_{1i} \sim \mathcal{N}(0, 1) \).

Model B. \( Y_i(1) = Y_i(0) + X_i'\alpha_1 \), where \( \alpha_1 \in \mathbb{R}^{100} \).

The choice of parameters \( \Sigma, \alpha_0 \) and \( \alpha_1 \) is explained in Appendix G of Narita and Yata (2023). \( D_i(0) \) and \( D_i(1) \) are generated as \( D_i(0) = 0 \) and \( D_i(1) = 1\{Y_i(1) - Y_i(0) > u_i\} \), where \( u_i \sim \mathcal{N}(0, 1) \).

To generate \( Z_i \), let \( q_{0.495} \) and \( q_{0.505} \) be the 49.5th and 50.5th (empirical) quantiles of the first covariate \( X_{i1} \). Let \( \tau_{pred}(X_i) \) be a real-valued function of \( X_i \), which is
constructed by random forests using an independent sample (see Appendix G of Narita and Yata (2023) for the details). $Z_i$ is then generated as

$$Z_i = \begin{cases} 
Z_i^* \sim \text{Bernoulli}(0.5) & \text{if } X_{i1} \in [q_{0.495}, q_{0.505}] \\
1 & \text{if } X_{i1} \notin [q_{0.495}, q_{0.505}] \text{ and } \tau_{\text{pred}}(X_i) \geq 0 \\
0 & \text{if } X_{i1} \notin [q_{0.495}, q_{0.505}] \text{ and } \tau_{\text{pred}}(X_i) < 0.
\end{cases}$$

The first case corresponds to the RCT segment while the latter two cases to the deterministic algorithm segment. The algorithm function $A$ is given by

$$A(x) = \begin{cases} 
0.5 & \text{if } x_1 \in [q_{0.495}, q_{0.505}] \\
1 & \text{if } x_1 \notin [q_{0.495}, q_{0.505}] \text{ and } \tau_{\text{pred}}(x) \geq 0 \\
0 & \text{if } x_1 \notin [q_{0.495}, q_{0.505}] \text{ and } \tau_{\text{pred}}(x) < 0.
\end{cases}$$

Finally, $D_i$ and $Y_i$ are generated as $D_i = Z_iD_i(1) + (1 - Z_i)D_i(0)$ and $Y_i = D_iY_i(1) + (1 - D_i)Y_i(0)$, respectively.

**Estimands and Estimators.** We consider four parameters as target estimands: $\text{ATE} \equiv E[Y_i(1) - Y_i(0)]$; $\text{ATE(RCT)} \equiv E[Y_i(1) - Y_i(0)|X_{i1} \in [q_{0.495}, q_{0.505}]]$; $\text{LATE} \equiv E[Y_i(1) - Y_i(0)|D_i(1) \neq D_i(0)]$; and $\text{LATE(RCT)} \equiv E[Y_i(1) - Y_i(0)|D_i(1) \neq D_i(0), X_{i1} \in [q_{0.495}, q_{0.505}]]$. In the case where the treatment effect does not depend on $X_i$ (Model A), $\text{ATE}$ and $\text{LATE}$ are the same as $\text{ATE(RCT)}$ and $\text{LATE(RCT)}$, respectively. In the case where the treatment effect depends on $X_i$ (Model B), the conditional effects are heterogeneous. However, since the RCT segment consists of those in the middle of the distribution of $X_{i1}$, the average effect for the RCT segment is close to the unconditional average effect. As a result, $\text{ATE}$ is similar to $\text{ATE(RCT)}$, and $\text{LATE}$ is similar to $\text{LATE(RCT)}$.

We use the data $\{(Y_i, X_i, D_i, Z_i)\}_{i=1}^n$ to estimate the treatment effect parameters. Our main approach is 2SLS with fixed-bandwidth APS controls in Theorem 1. To compute fixed-bandwidth APS, we use $S = 400$ simulation draws for each observation.

We compare our approach with two naive alternatives. The first alternative is OLS of $Y_i$ on a constant and $D_i$ (i.e., the difference in the sample mean of $Y_i$ between the treated group and untreated group) using all observations. The second alternative
is 2SLS with $A(X_i)$ controls. This method uses the observations with $A(X_i) \in (0, 1)$ to run the 2SLS regression of $Y_i$ on $D_i$ and $A(X_i)$ using $Z_i$ as an instrument for $D_i$ (see (4) and (5) in Section 4.3) and reports the coefficient on $D_i$.

For both models, the 2SLS estimator converges in probability to LATE(RCT) (equivalently, the right-hand side of equation (3) in Section 4.3) whether we control for fixed-bandwidth APS or $A(X_i)$. However, 2SLS with $A(X_i)$ controls uses only the RCT segment while 2SLS with fixed-bandwidth APS controls additionally uses the individuals near the decision boundary of the deterministic algorithm (i.e., the boundary of the region for which $\tau_{\text{pred}}(x) \geq 0$). Therefore, 2SLS with fixed-bandwidth APS controls is expected to produce a more precise estimate than 2SLS with $A(X_i)$ controls if the conditional effects for those near the boundary are not far from the target estimand.

We do not apply any multidimensional RD estimators as alternatives, since there appear to be no existing RD estimators applicable to this setup. It is hard to apply distance-based RD methods since it is difficult to compute the distance from each $X_i$ to the high-dimensional random-forests decision boundary. An alternative is to use the individual’s predicted effect $\tau_{\text{pred}}(X_i)$ as a univariate running variable. However, $\tau_{\text{pred}}(X_i)$ may not be a continuous variable since $\tau_{\text{pred}}$ is constructed by tree-based methods.

Performance Measures. 2SLS with a single instrument has no moments, so we cannot consider the bias, standard deviation, or mean squared error. As an alternative, we calculate the median bias, $\text{med}(\hat{\theta}) - \theta$, median absolute deviation from the median, $\text{med}(|\hat{\theta} - \text{med}(\hat{\theta})|)$, and median absolute error, $\text{med}(|\hat{\theta} - \theta|)$, where $\theta$ and $\hat{\theta}$ denote the estimand and estimator, respectively.

Results. Table 1 reports the median bias, median absolute deviation from the median (a measure of dispersion), and median absolute error (an overall performance measure). Panels A and B present the results for the cases where the conditional effects are homogeneous and heterogeneous, respectively. OLS with no controls is significantly biased, showing the importance of correcting for omitted variable bias. 2SLS with fixed-bandwidth APS controls achieves this goal, as demonstrated by its smaller biases across models, target parameters, and small values of the bandwidth $\delta$. 

25
Table 1: Comparison of Estimators and Coverage of Confidence Intervals

<table>
<thead>
<tr>
<th></th>
<th>OLS with No Controls</th>
<th>2SLS with A(Xi) Controls</th>
<th>Our Method: 2SLS with Approximate Propensity Score Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>δ = 0.01</td>
<td>δ = 0.05</td>
<td>δ = 0.1</td>
</tr>
<tr>
<td>Estimand: ATE</td>
<td>0.663</td>
<td>0.562</td>
<td>0.558</td>
</tr>
<tr>
<td>Median Bias</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median Absolute Error</td>
<td>0.663</td>
<td>0.562</td>
<td>0.558</td>
</tr>
<tr>
<td>Coverage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg N</td>
<td>10000</td>
<td>100</td>
<td>397</td>
</tr>
</tbody>
</table>

Panel A: Homogeneous Conditional Effects (Model A)

Estimand: ATE = 0
- Median Bias: 0.663, 0.562, 0.558, 0.618, 0.652, 0.716, 0.811, 0.965
- Median Absolute Error: 0.663, 0.562, 0.558, 0.618, 0.652, 0.716, 0.811, 0.965
- Coverage: 0.4%, 95.2%, 94.4%, 92.7%, 84.0%, 46.0%, 3.1%, 0.0%
- Avg N: 10000, 100, 397, 1175, 1722, 2613, 3349, 3994

Estimand: ATE(RCT) = −0.004
- Median Bias: 0.098, −0.002, −0.006, 0.053, 0.088, 0.152, 0.246, 0.401
- Median Absolute Error: 0.098, 0.222, 0.129, 0.084, 0.091, 0.152, 0.246, 0.401
- Coverage: 0.4%, 95.2%, 94.4%, 92.7%, 84.0%, 46.0%, 3.1%, 0.0%
- Avg N: 10000, 100, 397, 1175, 1722, 2613, 3349, 3994

Estimand: LATE = 0.564
- Median Bias: 0.096, −0.005, −0.008, 0.051, 0.086, 0.150, 0.244, 0.399
- Median Absolute Error: 0.096, 0.223, 0.129, 0.083, 0.089, 0.150, 0.244, 0.399
- Med. Abs. Deviation: 0.014, 0.224, 0.128, 0.075, 0.073, 0.150, 0.244, 0.399
- Coverage: 0.4%, 95.2%, 94.4%, 92.7%, 84.0%, 46.0%, 3.1%, 0.0%
- Avg N: 10000, 100, 397, 1175, 1722, 2613, 3349, 3994

Estimand: LATE(RCT) = 0.566
- Median Bias: 0.096, −0.005, −0.008, 0.051, 0.086, 0.150, 0.244, 0.399
- Median Absolute Error: 0.096, 0.223, 0.129, 0.083, 0.089, 0.150, 0.244, 0.399
- Med. Abs. Deviation: 0.014, 0.224, 0.128, 0.075, 0.073, 0.150, 0.244, 0.399
- Coverage: 0.4%, 95.2%, 94.4%, 92.7%, 84.0%, 46.0%, 3.1%, 0.0%
- Avg N: 10000, 100, 397, 1175, 1722, 2613, 3349, 3994

Panel B: Heterogeneous Conditional Effects (Model B)

Estimand: ATE = 0
- Median Bias: 1.010, 0.561, 0.471, 0.496, 0.525, 0.590, 0.698, 0.882
- Median Absolute Error: 1.010, 0.566, 0.471, 0.496, 0.525, 0.590, 0.698, 0.882
- Coverage: 0.0%, 94.6%, 92.4%, 91.7%, 94.1%, 93.5%, 51.8%, 0.3%
- Avg N: 10000, 100, 397, 1175, 1722, 2613, 3349, 3994

Estimand: ATE(RCT) = −0.004
- Median Bias: 1.015, 0.566, 0.475, 0.500, 0.529, 0.595, 0.702, 0.887
- Median Absolute Error: 1.015, 0.571, 0.475, 0.500, 0.529, 0.595, 0.702, 0.887
- Coverage: 0.0%, 94.6%, 92.4%, 91.7%, 94.1%, 93.5%, 51.8%, 0.3%
- Avg N: 10000, 100, 397, 1175, 1722, 2613, 3349, 3994

Estimand: LATE = 0.564
- Median Bias: 0.446, −0.002, −0.093, −0.068, −0.039, 0.026, 0.134, 0.318
- Median Absolute Error: 0.446, 0.262, 0.163, 0.099, 0.073, 0.057, 0.134, 0.318
- Coverage: 0.0%, 94.6%, 92.4%, 91.7%, 94.1%, 93.5%, 51.8%, 0.3%
- Avg N: 10000, 100, 397, 1175, 1722, 2613, 3349, 3994

Estimand: LATE(RCT) = 0.559
- Median Bias: 0.451, 0.002, −0.089, −0.063, −0.035, 0.031, 0.138, 0.323
- Median Absolute Error: 0.451, 0.265, 0.162, 0.097, 0.074, 0.057, 0.138, 0.323
- Coverage: 0.0%, 94.6%, 92.4%, 91.7%, 94.1%, 93.5%, 51.8%, 0.3%
- Avg N: 10000, 100, 397, 1175, 1722, 2613, 3349, 3994

Note: This table shows the median bias, median of absolute errors, and median of absolute deviations from the median of OLS with no controls, 2SLS with A(Xi) controls, and 2SLS with Approximate Propensity Score controls. These statistics are computed with the estimand set to ATE, ATE(RCT), LATE, or LATE(RCT). The “Coverage” row in each panel shows the probabilities that the 95% confidence intervals of the form [\(\hat{\beta}_1 - 1.96\hat{\sigma}_n, \hat{\beta}_1 + 1.96\hat{\sigma}_n\)] contains LATE(RCT), where \(\hat{\beta}_1\) is the estimate and \(\hat{\sigma}_n\) is its heteroskedasticity-robust standard error. We use 1,000 replications of a size 10,000 simulated sample to compute these statistics. We use several possible values of \(\delta\) to compute the Approximate Propensity Score. All Approximate Propensity Scores are computed by averaging 400 simulation draws of A(Xi). Panel A reports the results under the model in which the treatment effect does not depend on X_i (Model A). Panel B reports the results under the model in which the treatment effect depends on X_i (Model B). The bottom row in each panel shows the average number of observations used for estimation (i.e., the average number of observations for which the Approximate Propensity Score or A(Xi) is strictly between 0 and 1).
2SLS with fixed-bandwidth APS controls shows a consistent pattern; as the bandwidth $\delta$ grows, the bias increases while the absolute deviation from the median declines. For several values of $\delta$, 2SLS with fixed-bandwidth APS controls outperforms 2SLS with $A(X_i)$ controls in terms of the median absolute error. This finding implies that exploiting individuals near the multidimensional decision boundary of the deterministic algorithm can lead to better performance than using only the RCT segment.

We also evaluate our inference procedure based on Theorem 1. Table 1 reports the coverage probabilities of the 95% confidence intervals for LATE(RCT) constructed from the estimates and their heteroskedasticity-robust standard errors. The confidence intervals for 2SLS offer nearly correct coverage when $\delta$ is small, which supports the implication of Theorem 1 that the inference procedure is valid when we use a sufficiently small $\delta$. Overall, Table 1 shows that our estimator works well in this high-dimensional setting and performs better than alternative estimators.

6 Empirical Policy Application

6.1 Hospital Relief Funding during the Pandemic

The COVID-19 pandemic afflicted millions of people across the country and imposed historic challenges for the health system. The pandemic led to revenue losses coupled with skyrocketing expenses, pushing many already overburdened hospitals further to their financial brink.

To deal with this crisis, as part of the 3-phase Coronavirus Aid, Relief, and Economic Security (CARES) Act, the US government distributed tens of billions of dollars of relief funding to hospitals since April 2020. This funding intended to help health care providers hit hardest by the COVID-19 outbreak. The bill specified that providers may (but are not required to) use the funds for COVID-19-related expenses such as construction of temporary structures, purchasing medical supplies and equipment (including personal protective equipment and testing supplies), increased workforce utilization and training, establishing emergency operation centers, retrofitting facilities, and managing the surge in capacity.

We ask whether this funding had a causal impact on hospital operation and ac-
tivities in dealing with COVID-19 patients. Answering this question would help the government design better funding policies to respond to future healthcare crises. We focus on an initial portion of this funding ($10 billion). This portion was allocated to hospitals that qualified as “safety net hospitals” according to a specific eligibility criterion. This eligibility criterion intends to direct funding towards hospitals that “disproportionately provide care to the most vulnerable, and operate on thin margins.” Specifically, an acute care hospital is deemed eligible for funding if the following conditions hold:

- Medicare Disproportionate Patient Percentage (DPP) of 20.2% or greater. DPP is equal to the sum of (1) the percentage of Medicare inpatient days attributable to patients eligible for both Medicare Part A and Supplemental Security Income (SSI), and (2) the percentage of total inpatient days attributable to patients eligible for Medicaid but not Medicare Part A.

- Annual Uncompensated Care (UCC) of at least $25,000 per bed. UCC is a measure of hospital care provided for which no payment was received from the patient or insurer. It is the sum of a hospital’s bad debt and the financial assistance it provides.

- Profit Margin (net income/(net patient revenue + total other income)) of 3.0% or less.

Hospitals that do not qualify on any of the three dimensions are funding ineligible. Figure 2 visualizes how the three dimensions determine funding eligibility. From the original space of the three eligibility determinants, we extract two-dimensional planes to better visualize the structure of quasi-experimental variation. As the bottom two-dimensional planes show, eligibility discontinuously changes as hospitals cross the eligibility boundary in the characteristic space. This setting is a three-dimensional RDD, falling under our framework.

Our treatment is the funding amount, which is calculated as follows. Each eligible hospital is assigned a facility score, which is calculated as the product of DPP and the number of beds in that hospital. This facility score determines the share of funding allocated to the hospital, out of the total $10 billion. The share received by each hospital is determined by the ratio of the hospital’s facility score to the sum of
Figure 2: Regression Discontinuity in Hospital Funding Eligibility

Note: The top figure visualizes the three hospital characteristics that determine funding eligibility. The bottom figures show the data points plotted along 2 out of 3 dimensions. The bottom left panel plots disproportionate patient percentage against profit margin, while the bottom right panel plots uncompensated care per bed against profit margin. We remove hospitals above the 99th percentile of disproportionate patient percentage and uncompensated care per bed, for visibility purposes.
facility scores across all eligible hospitals. The amount of funding that can be received by a hospital is bounded below at $5 million and capped above at $50 million. We compute funding eligibility as well as the funding amount, by using data from the Healthcare Cost Report Information System (HCRIS) for the 2018 financial year.\(^8\)

A majority of eligible hospitals receive the minimum amount of $5 million. A small mass of hospitals receive amounts close to the maximum of $50 million. The distribution of funding amounts received by eligible hospitals is found in Figure A.1 in Narita and Yata (2023).

Our outcomes are a few different versions of the number of COVID patients hospitalized at each hospital. To obtain these outcomes, we use the publicly available COVID-19 Reported Patient Impact and Hospital Capacity by Facility dataset (U.S. Department of Health and Human Services, 2020–2021). This provides facility-level data on hospital utilization aggregated on a weekly basis, from July 31st 2020 onwards. Summary statistics about hospital outcomes and characteristics are documented in Table 2. Eligible hospitals have larger numbers of inpatient and ICU beds occupied by COVID-19 patients. Eligible hospitals also have a higher disproportionate patient percentage, higher uncompensated care per bed, lower profit margins, more employees and beds, and shorter lengths of inpatient stay. These patterns are consistent with the funding’s goal of helping struggling hospitals.

### 6.2 Covariate Balance Estimates

We validate our method by evaluating the balancing property of fixed-bandwidth APS conditioning. We calculate fixed-bandwidth-APS-controlled differences in covariate means for eligible vs ineligible hospitals. We run the following OLS regression of hospital-level characteristics on the eligibility status using observations with 

\[ p^*(X_i; \delta) \in (0, 1): \]

\[ W_i = \gamma_0 + \gamma_1 Z_i + \gamma_2 p^*(X_i; \delta) + \eta_i, \]

where \( W_i \) is one of the predetermined characteristics of the hospital, \( Z_i \) is a funding eligibility dummy, \( X_i \) is a vector of the three input variables (DPP, UCC, and profit

---

\(^8\)We use the methodology detailed in the CARES Act website to project funding based on 2018 financial year cost reports. We use the RAND cleaned version of the dataset (RAND Corporation, 2018). See Appendix H of Narita and Yata (2023) for details on the construction of our dataset.
Table 2: Hospital Characteristics and Outcomes

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Ineligible Hospitals</th>
<th>Eligible Hospitals</th>
<th>Hospitals w/ APS ∈ (0,1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># Confirmed/Suspected COVID Patients</td>
<td>105.59</td>
<td>98.41</td>
<td>136.61</td>
</tr>
<tr>
<td></td>
<td># Confirmed COVID Patients</td>
<td>80.10</td>
<td>73.86</td>
<td>107.83</td>
</tr>
<tr>
<td></td>
<td># Confirmed/Suspected COVID Patients in ICU</td>
<td>26.62</td>
<td>24.41</td>
<td>36.56</td>
</tr>
<tr>
<td></td>
<td># Confirmed COVID Patients in ICU</td>
<td>26.62</td>
<td>24.41</td>
<td>36.56</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>4,008</td>
<td>3,293</td>
<td>715</td>
</tr>
</tbody>
</table>

Panel B: Hospital Characteristics Means

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Ineligible Hospitals</th>
<th>Eligible Hospitals</th>
<th>Hospitals w/ APS ∈ (0,1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beds</td>
<td>143.66</td>
<td>134.60</td>
<td>188.35</td>
<td>205.30</td>
</tr>
<tr>
<td>Interns and residents (full-time equivalents) per bed</td>
<td>.06</td>
<td>.05</td>
<td>.11</td>
<td>.09</td>
</tr>
<tr>
<td>Adult and pediatric hospital beds</td>
<td>120.26</td>
<td>113.29</td>
<td>154.66</td>
<td>169.64</td>
</tr>
<tr>
<td>Ownership: Proprietary (for-profit)</td>
<td>.19</td>
<td>.20</td>
<td>.18</td>
<td>.16</td>
</tr>
<tr>
<td>Ownership: Governmental</td>
<td>.22</td>
<td>.22</td>
<td>.23</td>
<td>.16</td>
</tr>
<tr>
<td>Ownership: Voluntary (non-profit)</td>
<td>.58</td>
<td>.58</td>
<td>.59</td>
<td>.68</td>
</tr>
<tr>
<td>Inpatient length of stay</td>
<td>9.21</td>
<td>10.14</td>
<td>4.66</td>
<td>4.37</td>
</tr>
<tr>
<td>Employees on payroll (full-time equivalents)</td>
<td>973.90</td>
<td>897.31</td>
<td>1351.57</td>
<td>1511.87</td>
</tr>
<tr>
<td>Disproportionate patient percentage</td>
<td>.21</td>
<td>.18</td>
<td>.38</td>
<td>.36</td>
</tr>
<tr>
<td>Uncompensated care per bed ($)</td>
<td>59,850.00</td>
<td>56,556.03</td>
<td>76,096.31</td>
<td>45,575.28</td>
</tr>
<tr>
<td>Profit margin</td>
<td>.02</td>
<td>.04</td>
<td>−.07</td>
<td>−.03</td>
</tr>
<tr>
<td>N</td>
<td>4,633</td>
<td>3,852</td>
<td>781</td>
<td>494</td>
</tr>
</tbody>
</table>

Note: This table reports averages of outcome variables and hospital characteristics by funding eligibility. Panel A reports the outcome variable means. Outcome variable estimates are 7 day sums for the week spanning July 31st 2020 to August 6th 2020. Confirmed or Suspected COVID patients refer to the sum of patients in inpatient beds with lab-confirmed/suspected COVID. Confirmed COVID patients refer to the sum of patients in inpatient beds with lab-confirmed COVID, including those with both lab-confirmed COVID and influenza. Inpatient bed totals also include observation beds. Similarly, Confirmed/Suspected COVID patients in ICU refer to the sum of patients in ICU beds with lab-confirmed or suspected COVID. Confirmed COVID patients in ICU refers to the sum of patients in ICU beds with lab-confirmed COVID, including those with both lab-confirmed COVID and influenza. Panel B reports the means for hospital characteristics for the financial year 2018. Column 1 shows the means for all hospitals. Columns 2 and 3 show the means for hospitals that are ineligible and eligible to receive funding, respectively. Column 4 shows the means for the hospitals with nondegenerate Approximate Propensity Score with bandwidth δ = 0.05. Approximate Propensity Score is computed by averaging 10,000 simulation draws.

margin) that determine the funding eligibility, and \( p^*(X_i; \delta) \) is the simulated fixed-bandwidth APS. We compute fixed-bandwidth APS using \( S = 10,000 \) simulation draws for different bandwidth values. The estimated coefficient on \( Z_i \) is the fixed-bandwidth-APS-controlled difference in the mean of the covariate between eligible and ineligible hospitals. For comparison, we also run the OLS regression with no controls using the whole sample.

Table 3 reports the covariate balance estimates. Column 2 shows that, without controlling for fixed-bandwidth APS, eligible hospitals are significantly different from ineligible hospitals. All the relevant hospital eligibility characteristics are strongly associated with eligibility. Once we control for fixed-bandwidth APS with small enough
Table 3: Covariate Balance Regressions

<table>
<thead>
<tr>
<th></th>
<th>Mean (Ineligible Hospitals)</th>
<th>OLS with No Controls</th>
<th>Our Method: OLS with Approximate Propensity Score Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>δ = 0.01 δ = 0.025 δ = 0.05 δ = 0.075 δ = 0.1 δ = 0.25 δ = 0.5</td>
</tr>
<tr>
<td>Profit margin</td>
<td>0.04</td>
<td>-0.11</td>
<td>-0.01 -0.01 0.02 0.01 0.03 0.05 0.04</td>
</tr>
<tr>
<td>Uncompensated care per bed ($)</td>
<td>56,556</td>
<td>19,540</td>
<td>2,941 -8,408 -10,882 -9,432 -7,232 -8,071</td>
</tr>
<tr>
<td>Disproportionate patient percentage</td>
<td>0.18</td>
<td>0.21</td>
<td>-0.06 -0.09 -0.09 -0.07 -0.07 -0.07 -0.07</td>
</tr>
</tbody>
</table>

Panel A: Determinants of Funding Eligibility

<table>
<thead>
<tr>
<th></th>
<th>Profit margin</th>
<th>Uncompensated care per bed ($)</th>
<th>Disproportionate patient percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=4633</td>
<td>N=4633</td>
<td>N=4633</td>
</tr>
<tr>
<td></td>
<td>454.26</td>
<td>19,540</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>(69.23)</td>
<td>(10,419)</td>
<td>(0.01)</td>
</tr>
<tr>
<td></td>
<td>1973.70</td>
<td>2,941</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>(1,382.13)</td>
<td>6,235</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>155.03</td>
<td>-8,408</td>
<td>-0.09</td>
</tr>
<tr>
<td></td>
<td>(561.57)</td>
<td>-10,882</td>
<td>-0.09</td>
</tr>
<tr>
<td></td>
<td>-58.52</td>
<td>-9,432</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td>(432.30)</td>
<td>-7,232</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td>65.12</td>
<td>-8,071</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td>(354.30)</td>
<td>39.85</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td>39.85</td>
<td>192.77</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td>(178.37)</td>
<td>-3.40</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td>124.68</td>
<td></td>
<td>-0.07</td>
</tr>
</tbody>
</table>

Panel B: Other Hospital Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Full time employees</th>
<th>Medicare net revenue (in millions $)</th>
<th>Occupancy</th>
<th>Operating margin</th>
<th>Beds</th>
<th>Costs per discharge (in thousands $)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>897.32</td>
<td>20.04</td>
<td>0.44</td>
<td>0.02</td>
<td>134.60</td>
<td>66.28</td>
</tr>
<tr>
<td></td>
<td>(69.23)</td>
<td>(2.39)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(7.05)</td>
<td>(17.93)</td>
</tr>
<tr>
<td></td>
<td>1,963.70</td>
<td>18.36</td>
<td>0.17</td>
<td>-0.11</td>
<td>53.75</td>
<td>-49.95</td>
</tr>
<tr>
<td></td>
<td>(1,382.13)</td>
<td>(25.85)</td>
<td>0.04</td>
<td>-0.11</td>
<td>(91.55)</td>
<td>(1.91)</td>
</tr>
<tr>
<td></td>
<td>155.03</td>
<td>-8.58</td>
<td>-0.01</td>
<td>0.01</td>
<td>17.33</td>
<td>3.78</td>
</tr>
<tr>
<td></td>
<td>(561.57)</td>
<td>-6.49</td>
<td>0.01</td>
<td>0.01</td>
<td>(41.41)</td>
<td>3.00</td>
</tr>
<tr>
<td></td>
<td>-58.52</td>
<td>-1.53</td>
<td>0.01</td>
<td>0.01</td>
<td>(33.85)</td>
<td>1.21</td>
</tr>
<tr>
<td></td>
<td>(432.30)</td>
<td>2.67</td>
<td>0.02</td>
<td>0.04</td>
<td>8.62</td>
<td>-5.41</td>
</tr>
<tr>
<td></td>
<td>65.12</td>
<td>2.55</td>
<td>0.04</td>
<td>0.06</td>
<td>12.90</td>
<td>1.46</td>
</tr>
<tr>
<td></td>
<td>(354.30)</td>
<td>-4.40</td>
<td>0.03</td>
<td>0.06</td>
<td>3.92</td>
<td>5.77</td>
</tr>
<tr>
<td></td>
<td>39.85</td>
<td>-4.00</td>
<td>0.03</td>
<td>0.06</td>
<td>12.92</td>
<td>-5.80</td>
</tr>
<tr>
<td></td>
<td>(178.37)</td>
<td>-7.232</td>
<td>0.03</td>
<td>0.06</td>
<td>5.80</td>
<td>-5.80</td>
</tr>
<tr>
<td></td>
<td>124.68</td>
<td>-8.071</td>
<td>0.03</td>
<td>0.06</td>
<td>5.80</td>
<td>-5.80</td>
</tr>
</tbody>
</table>

Panel B: Other Hospital Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Operating margin</th>
<th>Beds</th>
<th>Costs per discharge</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=4624</td>
<td>N=4633</td>
<td>N=3539</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>0.02</td>
<td>-49.95</td>
</tr>
<tr>
<td></td>
<td>(-0.01)</td>
<td>(-0.01)</td>
<td>(17.93)</td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>0.01</td>
<td>3.78</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(1.91)</td>
</tr>
<tr>
<td></td>
<td>0.03</td>
<td>0.03</td>
<td>3.00</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(1.38)</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>0.02</td>
<td>1.21</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(1.08)</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>0.04</td>
<td>-5.41</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(7.00)</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>0.06</td>
<td>1.46</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(5.90)</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>0.06</td>
<td>5.77</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(10.28)</td>
</tr>
</tbody>
</table>

Note: This table shows the results of the covariate balance regressions at the hospital level. The dependent variables for these regressions are drawn from the Healthcare Cost Report Information System for the financial year 2018. Disproportionate patient percentage, profit margin and uncompensated care per bed are used to determine the hospital’s funding eligibility. Other dependent variables shown indicate the financial health and utilization of the hospitals. In column 2, we regress the dependent variables on the eligibility of the hospital with no controls. In columns 3–9, we regress the dependent variables on funding eligibility controlling for the Approximate Propensity Score with different values of bandwidth $\delta$. All Approximate Propensity Scores are computed by averaging 10,000 simulation draws. Column 1 shows the mean of dependent variables for hospitals that are ineligible to receive safety net funding. Robust standard errors are reported in the parenthesis and the number of observations is reported separately for each regression. The last row reports the p-value of the joint significance test.

bandwidth $\delta$, eligible and ineligible hospitals have similar financial and utilization characteristics, as reported in columns 3–7 of Table 3. These estimates are consistent with our theoretical results, establishing the empirical ability of fixed-bandwidth APS
controls to eliminate selection bias.

6.3 Effects of Funding: OLS and 2SLS Estimates

The balancing performance of fixed-bandwidth APS motivates us to estimate the causal effects of funding by using algorithmically-determined funding eligibility as an IV. We study the effect of funding on relevant hospital outcomes, such as the number of inpatient beds occupied by adult COVID patients between July 31st 2020 and August 6th 2020.

We first run the following OLS (reduced-form) regression of each outcome on the binary funding eligibility, while controlling for fixed-bandwidth APS:

\[ Y_i = \alpha_0 + \alpha_1 Z_i + \alpha_2 p^*(X_i; \delta) + \nu_i, \]

where \( Y_i \) is a hospital-level outcome and \( Z_i \) is the binary indicator for funding eligibility. This OLS specification is a special case of the 2SLS specification in the theoretical analysis.

We then estimate the following 2SLS regression using the funding amount as the treatment and funding eligibility as an instrument. We run the regression on two different hospital-level outcome variables, using hospitals with \( p^*(X_i; \delta) \in (0, 1) \):

\[ D_i = \gamma_0 + \gamma_1 Z_i + \gamma_2 p^*(X_i; \delta) + v_i \]
\[ Y_i = \beta_0 + \beta_1 D_i + \beta_2 p^*(X_i; \delta) + \epsilon_i, \]

where \( D_i \) is the funding amount.\(^9\) As benchmarks, we also run OLS and 2SLS with no controls, as well as OLS controlling for the three eligibility determinants.

OLS estimates of funding effects, reported as the benchmark in column 1 of Table 4, indicate that funding is associated with a higher number of adult inpatient beds.

\(^9\)We also implemented 2SLS specifications where the treatment is a binary indicator for receiving any funding more than 10 million USD. These 2SLS specifications are special cases of our theoretical analysis. They produce qualitatively the same results. Also, it would be interesting to extend our theoretical results to the case with a continuous treatment. Our conjecture is that under suitable conditions, the 2SLS estimator converges in probability to a weighted average of the marginal treatment effect over the values of the continuous treatment and over the units on the eligibility boundary. See Angrist, Graddy and Imbens (2000) for a related result.
### Table 4: Effects of Funding on Hospital Behavior

<table>
<thead>
<tr>
<th></th>
<th>OLS with No Controls</th>
<th>OLS with Covariate Controls</th>
<th>2SLS with No Controls</th>
<th>2SLS with Covariate Controls</th>
<th>Our Method: 2SLS with Approximate Propensity Score Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>δ = 0.01</td>
<td>δ = 0.025</td>
<td>δ = 0.05</td>
<td>δ = 0.075</td>
<td>δ = 0.1</td>
</tr>
<tr>
<td>First stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(in millions $)</td>
<td>(0.50)</td>
<td>(5.49)</td>
<td>(3.17)</td>
<td>(2.02)</td>
<td></td>
</tr>
<tr>
<td>Reduced form</td>
<td>33.97</td>
<td>−7.51</td>
<td>−31.50</td>
<td>15.98</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(7.44)</td>
<td>(71.05)</td>
<td>(48.48)</td>
<td>(29.66)</td>
<td></td>
</tr>
<tr>
<td>$1mm of funding</td>
<td>4.53</td>
<td>2.50</td>
<td>−4.71</td>
<td>−2.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.63)</td>
<td>(0.79)</td>
<td>(0.50)</td>
<td>(4.42)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.53)</td>
<td>(1.97)</td>
<td>(1.75)</td>
<td>(1.47)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.01)</td>
<td>(1.01)</td>
<td>(1.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>3558</td>
<td>3558</td>
<td>3558</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td></td>
<td>191</td>
<td>393</td>
<td>550</td>
<td>725</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1388</td>
<td>1949</td>
<td>1388</td>
<td>1949</td>
<td></td>
</tr>
<tr>
<td># Confirmed COVID Patients in ICU:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First stage</td>
<td>13.89</td>
<td>15.03</td>
<td>13.99</td>
<td>15.61</td>
<td></td>
</tr>
<tr>
<td>(in millions $)</td>
<td>(0.50)</td>
<td>(5.62)</td>
<td>(3.19)</td>
<td>(2.08)</td>
<td></td>
</tr>
<tr>
<td>Reduced form</td>
<td>12.16</td>
<td>4.42</td>
<td>−1.71</td>
<td>2.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.58)</td>
<td>(24.19)</td>
<td>(16.89)</td>
<td>(9.89)</td>
<td></td>
</tr>
<tr>
<td>$1mm of funding</td>
<td>1.51</td>
<td>0.82</td>
<td>0.88</td>
<td>0.29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.21)</td>
<td>(0.27)</td>
<td>(0.17)</td>
<td>(1.56)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.29)</td>
<td>(1.20)</td>
<td>(0.63)</td>
<td>(0.56)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.48)</td>
<td>(0.35)</td>
<td>(0.35)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>3503</td>
<td>3503</td>
<td>3503</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td></td>
<td>181</td>
<td>379</td>
<td>526</td>
<td>689</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1342</td>
<td>1893</td>
<td>1342</td>
<td>1893</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** In this table, we estimate the effects of the funding amount on hospital-level outcomes. Column 1 presents OLS results of outcome variables on funding without controls. Column 2 controls for disproportionate patient percentage, uncompensated care per bed, and profit margin. Column 3 shows the results of 2SLS regression with no controls. In columns 4–10, we control for Approximate Propensity Score with different values of bandwidth $\delta$ on the sample with nondegenerate Approximate Propensity Scores. The first stage shows the effect of funding eligibility on the amount of relief funding received by hospitals, in millions of dollars. The reduced form shows the effect of funding eligibility on the outcome. 2SLS estimates show the effect of the amount of relief funding on outcomes. Approximate Propensity Scores are computed by averaging 10,000 simulation draws. The outcome variables are 7 day totals for the week spanning July 31st, 2020 to August 6th, 2020. Confirmed COVID patients refer to the sum of patients in inpatient beds with lab-confirmed COVID-19, including those with both lab-confirmed COVID-19 and influenza. Inpatient bed totals also include observation beds. Confirmed COVID patients in ICU refers to the sum of patients in ICU beds with lab-confirmed COVID-19, including those with both lab-confirmed COVID-19 and influenza. Robust standard errors are reported in parentheses.

and higher number of staffed ICU beds utilized by COVID patients. For example, the estimates indicate that receiving an additional $1 million in funding is associated with 4.53 more beds occupied by patients. These uncontrolled OLS estimates show a similar picture as the descriptive statistics in Table 2. Naive 2SLS estimates with no controls and OLS with covariate controls produce similar significantly positive associations of funding with outcomes.

However, the OLS or uncontrolled 2SLS estimates turn out to be an artifact of selection bias. In contrast with them, our preferred reduced-form and 2SLS estimates with APS controls show a different picture (columns 4–10). The gains in the number
of inpatient beds and staffed ICU beds occupied by COVID patients become much smaller and lose significance across all bandwidth specifications. In fact, even the sign of the estimated funding effect is reversed for several combinations of the outcome and bandwidth. Once we control for fixed-bandwidth APS to eliminate the bias, therefore, funding has little to no effect on the hospital utilization level by COVID-19 patients. These results suggest that fixed-bandwidth APS reveals important selection bias in the naively estimated effects of funding.\footnote{The 2SLS estimates in Table 4 are unlikely to be compromised by differential attrition. Estimates reported in Table A.1 in Narita and Yata (2023) show little difference in outcome availability rates between eligible and ineligible hospitals once we control for fixed-bandwidth APS.}

We also estimate the evolving effects of funding for each week from July 31st, 2020 to April 2nd, 2021 (see Appendix H.3 of Narita and Yata (2023)). The estimated dynamic effects are similar to the initial null effects, suggesting that funding has no substantial effect even in the long run. Furthermore, we estimate dynamic effects for different groups of hospitals defined by hospital size and ownership type. We do not find any strong evidence of heterogeneity at any point in time.

The insignificance of the estimates suggests that funding by the CARES Act had little effect on hospital utilization during the pandemic. The null effect is widely observed for subgroups of hospitals at different points in time. This finding is consistent with policy and media arguments that CARES Act funding was not well targeted toward needy providers. Unlike the previous arguments, our analysis provides causal evidence supporting the concern.

7 Conclusion

As algorithmic decisions become the norm, the world becomes a mountain of natural experiments. We develop a general method to use these algorithm-produced instruments to identify and estimate treatment effects. Our analysis of the CARES Act hospital relief funding uses the proposed method to find that relief funding has little effect on COVID-19-related hospital activities. OLS or uncontrolled 2SLS estimates, by contrast, show larger and more significant effects. The large estimates appear to be an artifact of selection bias; relief funding just went to hospitals with more COVID-19 patients, without helping hospitals accommodate additional patients.
Our analysis provides a few implications for policy and management practices of decision-making algorithms. It is important to record the implementation of algorithms in a replicable way, including what input variables $X_i$ are used to make algorithmic recommendation $Z_i$. Another key lesson is the importance of recording an algorithm’s recommendation $Z_i$ even if they are superseded by a human decision $D_i$. These data retention efforts go a long way to exploit the full potential of algorithms as natural experiments.

An important future direction is estimation and inference details, such as data-driven bandwidth selection. This work needs to extend Imbens and Kalyanaraman (2012) and Calonico, Cattaneo and Titiunik (2014)’s bandwidth selection methods in the univariate RDD to our setting.\textsuperscript{11} Inference on treatment effects in our framework relies on large sample reasoning. It seems natural to additionally consider permutation or randomization inference. It will also be challenging but interesting to develop finite-sample optimal estimation and inference strategies. Finally, we look forward to empirical applications of our method in a variety of business and policy domains.

### A Additional Assumptions for Theorem 1

Here we provide additional assumptions required for proving asymptotic normality of $\hat{\beta}_1$ and $\hat{\beta}^\pi_1$ in Theorem 1. First, when $A$ is stochastic, we assume the following. Let

$$C^* = \{ x \in \mathbb{R}^p : A \text{ is continuously differentiable at } x \},$$

and let $D^* = \mathbb{R}^p \setminus C^*$ be the set of points at which $A$ is not continuously differentiable.

**Assumption A.1.** If $\Pr(A(X_i) \in (0, 1)) > 0$, then the following conditions (a)–(c) hold.

(a) (Probability of Neighborhood of $D^*$) $\Pr(X_i \in N(D^*, \delta)) = O(\delta)$.

\textsuperscript{11}For univariate RDDs, Imbens and Kalyanaraman (2012) and Calonico et al. (2014) estimate the bandwidth that minimizes the asymptotic mean squared error (AMSE). It is not straightforward to estimate the AMSE-optimal bandwidth in our setting with many running variables and complex IV assignment, since it requires nonparametric estimation of functions on the multidimensional covariate space such as conditional mean functions, their derivatives, the curvature of the RDD boundary, etc.
(b) (Bounded Partial Derivatives of $A$) The partial derivatives of $A$ are bounded on $C^*$. 

(c) (Bounded Conditional Mean) $E[Y_i|X_i]$ is bounded on $\mathcal{X}$. 

To explain the role of Assumption A.1 (a), consider a path of points $x_\delta \in N(D^*, \delta) \cap C^*$ indexed by $\delta > 0$. Since $A$ is continuous at $x_\delta$, $p^A(x_\delta) = A(x_\delta)$ (as formally implied by Lemma C.5 in Appendix C.3). However, $p^A(x_\delta; \delta)$ does not necessarily get sufficiently close to $A(x_\delta)$ even as $\delta \to 0$, since $x_\delta$ is in the $\delta$-neighborhood of $D^*$ and hence $A$ may discontinuously change within the $\delta$-ball $B(x_\delta, \delta)$. Assumption A.1 (a) requires that the probability of $X_i$ being in the $\delta$-neighborhood of $D^*$ shrink to zero at the rate of $\delta$, which makes the points in the neighborhood negligible.

Assumption A.1 (a) often holds in practice. If $A$ is continuously differentiable on $\mathcal{X}$, then $D^* \cap \mathcal{X} = \emptyset$, so this condition holds. If, for example, the treatment recommendation is randomly assigned based on a stratified randomized experiment, $D^*$ is the boundary at which the recommendation probability changes discontinuously. For any boundary of standard shape, $Pr(X_i \in N(D^*, \delta))$ vanishes at the rate of $\delta$, and the required condition is satisfied. Assumption A.1 (b) and (c) are regularity conditions, imposing the boundedness of the partial derivatives of $A$ and of the conditional mean of the outcome.

The following is the key to asymptotic normality of the simulation-based estimator $\hat{\beta}_s^*$.

**Assumption A.2** (The Number of Simulation Draws). $(n\delta_n)^{-1/2}S_n \to \infty$, and $Pr(p^A(X_i; \delta_n) \in (0, \gamma \frac{\log n}{S_n}) \cup (1 - \gamma \frac{\log n}{S_n}, 1)) = o(n^{-1/2}\delta_n^{1/2})$ for some $\gamma > \frac{1}{2}$.

Assumption A.2 imposes the condition on the growth rate of the number of simulation draws $S_n$. This assumption ensures that the bias caused by using $p^s(X_i; \delta_n)$ instead of $p^A(X_i; \delta_n)$ is asymptotically negligible. To understand this condition, note that $p^s(X_i; \delta_n)$ enters the 2SLS first-order condition, $\sum_{i=1}^n (1, Z_i, p^s(X_i; \delta_n))'(Y_i - \beta_0 - \beta_1D_i - \beta_2p^s(X_i; \delta_n))1\{p^s(X_i; \delta_n) \in (0, 1)\} = 0$, in two ways. First, $p^s(X_i; \delta_n)$ enters the condition in a nonlinear but smooth way through the $p^s(X_i; \delta_n)^2$ term. The asymptotic bias due to simulation errors is $O(\sqrt{n\delta_n^2}/S_n)$ under Assumption A.1 if $Pr(A(X_i) \in (0, 1)) > 0$ and $n\delta_n^2 \to 0$, and is $O(\sqrt{n\delta_n}/S_n)$ if $Pr(A(X_i) \in (0, 1)) = 0$. The bias diminishes under the first part of Assumption A.2. Second, $p^s(X_i; \delta_n)$ also
enters the first-order condition in a nonsmooth way, since we only use observations for which \( p^s(X_i; \delta_n) \in (0, 1) \). If \( p^A(X_i; \delta_n) \) is nondegenerate but close to zero or one, \( p^s(X_i; \delta_n) \) may be degenerate (i.e., \( A(X^*_{i,s}) = 0 \) for all \( s \) or \( A(X^*_{i,s}) = 1 \) for all \( s \)) with a large probability. The second part of Assumption A.2 ensures that the fraction of such observations goes to zero sufficiently fast, which eliminates the asymptotic bias caused by not using observations with \( p^A(X_i; \delta_n) \in (0, 1) \).

To illustrate how this assumption restricts the rate at which \( S_n \) grows sufficiently fast so that \( n^{1/2\delta_n^2/\log n} = o(1) \). One choice of \((\delta_n, S_n)\) that satisfies both parts of Assumption A.2 is \( \delta_n = \alpha_1 n^{-\kappa_1} \) and \( S_n = \alpha_2 n^{\kappa_2} \) for some \( \alpha_1, \alpha_2 > 0, \kappa_1 \in (1/2, 1) \) and \( \kappa_2 > \frac{1}{2}(1 - \kappa_1) \).

**B Proof of Theorem 1**

As we mention when we define our 2SLS estimator \( \hat{\beta} \) in Section 4.1, we drop the constant term if \( A(X_i) \) takes on only one nondegenerate value in the sample. If \( \text{Pr}(A(X_i) \in (0, 1)) > 0 \), \( \hat{\beta} = \left( \sum_{i=1}^n Z_{i,n} D_{i,n}^c I_{i,n} \right)^{-1} \sum_{i=1}^n Z_{i,n} Y_i I_{i,n} \) with probability approaching one if \( \text{Var}(A(X_i) | A(X_i) \in (0, 1)) > 0 \) and \( \hat{\beta} = \left( \sum_{i=1}^n Z_{i,n}^c (D_{i,n}^c )' I_{i,n} \right)^{-1} \sum_{i=1}^n Z_{i,n}^c Y_i I_{i,n} \) with probability approaching one if \( \text{Var}(A(X_i) | A(X_i) \in (0, 1)) = 0 \), where \( D_{i,n}^c = (D_i, p^A(X_i; \delta_n))' \) and \( Z_{i,n}^c = (Z_i, p^A(X_i; \delta_n))' \) (see Appendix C.2.1 of Narita and Yata (2023) for details). If \( \text{Pr}(A(X_i) \in (0, 1)) = 0 \), \( \hat{\beta} = \left( \sum_{i=1}^n Z_{i,n} D_{i,n}^c I_{i,n} \right)^{-1} \sum_{i=1}^n Z_{i,n} Y_i I_{i,n} \) with probability one.

We provide proofs separately for the two cases, the case in which \( \text{Pr}(A(X_i) \in (0, 1)) > 0 \) and the case in which \( \text{Pr}(A(X_i) \in (0, 1)) = 0 \). For each case, we first prove consistency and asymptotic normality of \( \hat{\beta}_1 \), and then prove those of \( \hat{\beta}_1^* \). For the first case, we only prove the results when \( \text{Var}(A(X_i) | A(X_i) \in (0, 1)) > 0 \), treating \( \hat{\beta} = \left( \sum_{i=1}^n Z_{i,n} D_{i,n} I_{i,n} \right)^{-1} \sum_{i=1}^n Z_{i,n} Y_i I_{i,n} \). The results can be obtained analogously when \( \text{Var}(A(X_i) | A(X_i) \in (0, 1)) = 0 \). Throughout the proof, we omit the subscript \( n \) from \( I_{i,n}, D_{i,n}, Z_{i,n}, \hat{\epsilon}_{i,n}, \hat{\Sigma}_n, \hat{\sigma}_n \), etc. for notational brevity.
Below, we provide the proof of consistency and asymptotic normality of $\hat{\beta}_1$ when $\Pr(A(X_i) \in (0, 1)) = 0$. The rest of the proof is in Appendix D.2.

**Notation.** For a scalar-valued differentiable function $f : S \subset \mathbb{R}^n \to \mathbb{R}$, let $\nabla f : S \to \mathbb{R}^n$ be a gradient of $f$: for every $x \in S$, $\nabla f(x) = \left( \frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_n} \right)'$. When the second-order partial derivatives of $f$ exist, let $D^2 f(x)$ be the $n \times n$ Hessian matrix for each $x \in S$: $(D^2 f(x))_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$. Let $f : S \subset \mathbb{R}^m \to \mathbb{R}^n$ be a function such that its first-order partial derivatives exist. For each $x \in S$, let $J f(x)$ be the $n \times m$ Jacobian matrix of $f$ at $x$: $(J f(x))_{ij} = \frac{\partial f_i(x)}{\partial x_j}$. For a positive integer $n$, let $I_n$ denote the $n \times n$ identity matrix.

Let $S \subset \mathbb{R}^p$ be a twice continuously differentiable set. For each $x \in \partial S$, we denote by $\nu_S(x) \in \mathbb{R}^p$ the inward unit normal vector of $\partial S$ at $x$, that is, the unit vector orthogonal to all vectors in the tangent space of $\partial S$ at $x$ that points toward the inside of $S$. For a set $S \subset \mathbb{R}^p$, let $d_S^* : \mathbb{R}^p \to \mathbb{R}$ be the signed distance function of $S$, defined by $d_S^*(x) = \begin{cases} d(x, \partial S) & \text{if } x \in \text{cl}(S) \\ -d(x, \partial S) & \text{if } x \in \mathbb{R}^p \setminus \text{cl}(S), \end{cases}$ where $d(x, B) = \inf_{y \in B} \|y - x\|$ for any $x \in \mathbb{R}^p$ for a set $B \subset \mathbb{R}^p$. Note that we can write $N(\partial S, \delta) = \{x \in \mathbb{R}^p : -\delta < d_S^*(x) < \delta\}$ for $\delta > 0$.

**B.1 Proof of Asymptotic Properties of $\hat{\beta}_1$ When $\Pr(A(X_i) \in (0, 1)) = 0$**

We use the results provided in Appendix C. By Lemma C.3, under Assumption 3 (d), there exists $\mu > 0$ such that $d_{\Omega^*}^*$ is twice continuously differentiable on $N(\partial \Omega^*, \mu)$ and

$$\int_{N(\partial \Omega^*, \delta)} g(x)dx = \int_{-\delta}^{\delta} \int_{\partial \Omega^*} g(u + \lambda \nu_{\Omega^*}(u)) J_{p-1}^{\Omega^*} \psi_{\Omega^*} (u, \lambda) d\mathcal{H}^{p-1}(u) d\lambda$$

for every $\delta \in (0, \mu)$ and every function $g : \mathbb{R}^p \to \mathbb{R}$ that is integrable on $N(\partial \Omega^*, \delta)$. Here, $J_{p-1}^{\Omega^*} \psi_{\Omega^*}$ is a function defined in Lemma C.3 and satisfies the following: $J_{p-1}^{\Omega^*} \psi_{\Omega^*}(u, \cdot)$ is continuously differentiable in $\lambda$ and $J_{p-1}^{\Omega^*} \psi_{\Omega^*}(u, 0) = 1$ for every
\[ u \in \partial \Omega^*, \text{ and } J_{\mu-1}^{\partial \Omega^*} \psi_{\Omega^*}(\cdot, \cdot) \text{ and } \frac{J_{\mu-1}^{\partial \Omega^*} \psi_{\Omega^*}((\cdot, \cdot))}{\partial \lambda} \text{ are bounded on } \partial \Omega^* \times (-\mu, \mu). \]

Below we show \( \hat{\beta}_1 - \beta_1 \xrightarrow{p} 0 \) if Assumption 3 holds, \( \delta_n \to 0 \), and \( n\delta_n \to \infty \), and \( \hat{\sigma}^{-1}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} N(0,1) \) if \( n\delta_3 \to 0 \) in addition. The proof proceeds in eight steps.

**Step B.1.1.** There exist \( \bar{\delta} > 0 \) and a bounded function \( r : \partial \Omega^* \cap N(\mathcal{X}, \bar{\delta}) \times (-1,1) \times (0,\bar{\delta}) \to \mathbb{R} \) such that

\[
p^A(u + \delta v \nu_{\Omega^*}(u); \delta) = k(v) + \delta r(u, v, \delta)
\]

for every \((u, v, \delta) \in \partial \Omega^* \cap N(\mathcal{X}, \bar{\delta}) \times (-1,1) \times (0,\bar{\delta})\), where

\[
k(v) = \begin{cases} 
1 - \frac{1}{2} I_x(1-v^2)(\frac{p+1}{2}, \frac{1}{2}) & \text{for } v \in [0,1) \\
\frac{1}{2} I_x(1-v^2)(\frac{p+1}{2}, \frac{1}{2}) & \text{for } v \in (-1,0).
\end{cases}
\]

Here \( I_x(\alpha, \beta) \) is the regularized incomplete beta function (the cumulative distribution function of the beta distribution with shape parameters \( \alpha \) and \( \beta \)).

**Proof.** By Assumption 3 (e) (ii), there exists \( \bar{\delta} \in (0, \frac{\mu}{2}) \) such that \( A(x) = 0 \) for almost every \( x \in N(\mathcal{X}, 3\bar{\delta}) \setminus \Omega^* \). By Taylor’s theorem, for every \( u \in \partial \Omega^* \cap N(\mathcal{X}, \bar{\delta}) \) and \( a \in B(0,2\bar{\delta}) \),

\[
d^*_{\Omega^*}(u + a) = d^*_{\Omega^*}(u) + \nabla d^*_{\Omega^*}(u)'a + a'R(u, a)a = \nu_{\Omega^*}(u)'a + a'R(u, a)a,
\]

where \( R(u, a) = \int_0^1 (1 - t)D^2 d^*_{\Omega^*}(u + ta)dt \), and the second equality follows since \( d^*_{\Omega^*}(u) = 0 \) and \( \nabla d^*_{\Omega^*}(u) = \nu_{\Omega^*}(u) \) for every \( u \in \partial \Omega^* \cap N(\mathcal{X}, \bar{\delta}) \) by Lemma C.1. Since \( D^2 d^*_{\Omega^*} \) is continuous and \( \text{cl}(N(\partial \Omega^*, 2\bar{\delta})) \) is bounded and closed, \( D^2 d^*_{\Omega^*} \) is bounded on \( \text{cl}(N(\partial \Omega^*, 2\bar{\delta})) \). Therefore, \( R(\cdot, \cdot) \) is bounded on \( \partial \Omega^* \cap N(\mathcal{X}, \bar{\delta}) \times B(0, 2\bar{\delta}) \).
For \((u, v, \delta) \in \partial \Omega^* \cap \mathcal{N}(\mathcal{X}, \delta) \times (-1, 1) \times (0, \delta)\),

\[
p^A(u + \delta v v_{\Omega^*} (u); \delta) = \frac{\delta^p \int_{B(0,1)} A(u + \delta v v_{\Omega^*} (u) + \delta w) dw}{\delta^p \int_{B(0,1)} dw} = \frac{\int_{B(0,1)} 1\{u + \delta v v_{\Omega^*} (u) + \delta w \in \Omega^*\} dw}{\text{Vol}_p} = \frac{\int_{B(0,1)} 1\{d_{\Omega^*}(u + \delta(v v_{\Omega^*} (u) + w)) \geq 0\} dw}{\text{Vol}_p} = \left[ \int_{B(0,1)} 1\{\delta v_{\Omega^*} (u) \cdot (v v_{\Omega^*} (u) + w) + \delta^2 (v v_{\Omega^*} (u) + w)' R(u, \delta(v v_{\Omega^*} (u) + w)) (v v_{\Omega^*} (u) + w) \geq 0\} dw \right]/\text{Vol}_p,
\]

where \(\text{Vol}_p\) denotes the volume of the \(p\)-dimensional unit ball, and the second equality follows since \(u + \delta v v_{\Omega^*} (u) + \delta w \in \mathcal{N}(\mathcal{X}, 3\delta)\) and hence \(A(u + \delta v v_{\Omega^*} (u) + \delta w) = 0\) for almost every \(w \in B(0,1)\) such that \(u + \delta v v_{\Omega^*} (u) + \delta w \notin \Omega^*\). Observe that

\[
1\{\delta v_{\Omega^*} (u) \cdot (v v_{\Omega^*} (u) + w) + \delta^2 (v v_{\Omega^*} (u) + w)' R(u, \delta(v v_{\Omega^*} (u) + w)) (v v_{\Omega^*} (u) + w) \geq 0\} = 1\{v + \nu_{\Omega^*} (u) \cdot w + \delta(v v_{\Omega^*} (u) + w)' R(u, \delta(v v_{\Omega^*} (u) + w)) (v v_{\Omega^*} (u) + w) \geq 0\} = 1\{v + \nu_{\Omega^*} (u) \cdot w \geq 0\} - a(u, v, w, \delta) + b(u, v, w, \delta),
\]

where \(a(u, v, w, \delta) = 1\{v + \nu_{\Omega^*} (u) \cdot w \geq 0, c(u, v, w, \delta) < 0\}\), \(b(u, v, w, \delta) = 1\{v + \nu_{\Omega^*} (u) \cdot w < 0, c(u, v, w, \delta) \geq 0\}\), and

\[
c(u, v, w, \delta) = v + \nu_{\Omega^*} (u) \cdot w + \delta(v v_{\Omega^*} (u) + w)' R(u, \delta(v v_{\Omega^*} (u) + w)) (v v_{\Omega^*} (u) + w).
\]

\(\{w \in B(0,1) : v + \nu_{\Omega^*} (u) \cdot w \geq 0\}\) is a region of the \(p\)-dimensional unit ball cut off by the plane \(\{w \in \mathbb{R}^p : v + \nu_{\Omega^*} (u) \cdot w = 0\}\). The distance from the unit ball’s center to the plane is \(|v|\). By the formula for the volume of a hyperspherical cap (Li, 2011),

\[
\int_{B(0,1)} 1\{v + \nu_{\Omega^*} (u) \cdot w \geq 0\} dw = \begin{cases} 
\text{Vol}_p (1 - \frac{1}{2} I_{(2(1-v)-(1-v)^2)} (\frac{p+1}{2}, \frac{1}{2})) & \text{if } v \in [0, 1) \\
\frac{1}{2} \text{Vol}_p I_{(2(1+v)-(1+v)^2)} (\frac{p+1}{2}, \frac{1}{2}) & \text{if } v \in (-1, 0].
\end{cases}
\]
Therefore, for every \((u, v, \delta) \in \partial \Omega^* \cap N(\mathcal{X}, \delta) \times (-1, 1) \times (0, \delta)\),

\[
p^A(u + \delta \nu_{\Omega^*}(u); \delta) = k(v) + \frac{\int_{B(0,1)} (-a(u, v, w, \delta) + b(u, v, w, \delta))dw}{\text{Vol}_p}.
\]

Now let \(r(u, v, \delta) = \delta^{-1}(p^A(u + \delta \nu_{\Omega^*}(u); \delta) - k(v))\). Since \(R(\cdot, \cdot)\) is bounded on \(\partial \Omega^* \cap N(\mathcal{X}, \delta) \times B(0, 2\delta)\) and \(\|\nu_{\Omega^*}(u)\| = 1\), there exists \(\bar{r} > 0\) such that

\[
|((\nu_{\Omega^*}(u) + w)'R(u, \delta(\nu_{\Omega^*}(u) + w))(\nu_{\Omega^*}(u) + w)| \leq \bar{r}
\]

for every \((u, v, w, \delta) \in \partial \Omega^* \cap N(\mathcal{X}, \delta) \times (-1, 1) \times B(0, 1) \times (0, \delta)\). Therefore,

\[
0 \leq a(u, v, w, \delta) \leq 1\{0 \leq v + \nu_{\Omega^*}(u) \cdot w < \delta \bar{r}\},
\]

\[
0 \leq b(u, v, w, \delta) \leq 1\{-\delta \bar{r} \leq v + \nu_{\Omega^*}(u) \cdot w < 0\}.
\]

It then follows that

\[
-\int_{B(0,1)} 1\{0 \leq v + \nu_{\Omega^*}(u) \cdot w < \delta \bar{r}\}dw \leq \int_{B(0,1)} (-a(u, v, w, \delta) + b(u, v, w, \delta))dw
\]

\[
\leq \int_{B(0,1)} 1\{-\delta \bar{r} \leq v + \nu_{\Omega^*}(u) \cdot w < 0\}dw.
\]

The set \(\{w \in B(0,1) : 0 \leq v + \nu_{\Omega^*}(u) \cdot w < \delta \bar{r}\}\) is a region of the \(p\)-dimensional unit ball cut off by the two planes \(\{w \in \mathbb{R}^p : v + \nu_{\Omega^*}(u) \cdot w = 0\}\) and \(\{w \in \mathbb{R}^p : v + \nu_{\Omega^*}(u) \cdot w = \delta \bar{r}\}\). Its Lebesgue measure is at most the volume of the \((p - 1)\)-dimensional unit ball times the distance between the two planes, so

\[
-\delta \text{Vol}_{p-1}\bar{r} \leq -\int_{B(0,1)} 1\{0 \leq v + \nu_{\Omega^*}(u) \cdot w < \delta \bar{r}\}dw.
\]

Likewise,

\[
\int_{B(0,1)} 1\{-\delta \bar{r} \leq v + \nu_{\Omega^*}(u) \cdot w < 0\}dw \leq \delta \text{Vol}_{p-1}\bar{r}.
\]

Therefore,

\[
\frac{-\delta \text{Vol}_{p-1}\bar{r}}{\text{Vol}_p} \leq \frac{\int_{B(0,1)} (-a(u, v, w, \delta) + b(u, v, w, \delta))dw}{\text{Vol}_p} \leq \frac{\delta \text{Vol}_{p-1}\bar{r}}{\text{Vol}_p}.
\]
It follows that
\[
    r(u, v, \delta) = \delta^{-1} \int_{B(0,1)} (-a(u, v, w, \delta) + b(u, v, w, \delta))dw \in \left[ -\frac{\text{Vol}_{p-1}r}{\text{Vol}_p}, \frac{\text{Vol}_{p-1}r}{\text{Vol}_p} \right],
\]
and hence \( r \) is bounded on \( \partial\Omega^* \cap N(\mathcal{X}, \tilde{\delta}) \times (-1, 1) \times (0, \tilde{\delta}). \)

\[\boxed{\text{Step B.1.2.} \text{ For every } (u, v, \delta) \in \partial\Omega^* \cap N(\mathcal{X}, \tilde{\delta}) \times (-1, 1) \times (0, \tilde{\delta}), p^A(u + \delta\nu_{\Omega^*}(u); \delta) \in (0, 1).}\]

**Proof.** Fix \((u, v, \delta) \in \partial\Omega^* \cap N(\mathcal{X}, \tilde{\delta}) \times (-1, 1) \times (0, \tilde{\delta}).\) Let \(\epsilon \in (0, \delta(1 - |v|)).\) Note that \(B(u, \epsilon) \subset B(u + \delta\nu_{\Omega^*}(u), \delta),\) since for any \(x \in B(u, \epsilon), \|u + \delta\nu_{\Omega^*}(u) - x\| \leq \|\delta\nu_{\Omega^*}(u)\| + \|u - x\| \leq |v| + \epsilon < \delta.\) By Step B.1.1, \(p^A(u) = \lim_{\delta \to 0} p^A(u; \delta') = k(0) = \frac{1}{2}.\) There exists \(\epsilon' \in (0, \epsilon)\) such that \(p^A(u; \epsilon') \in (0, 1).\) It then follows that \(0 < \mathcal{L}^p(B(u, \epsilon') \cap \Omega^*) \leq \mathcal{L}^p(B(u, \epsilon) \cap \Omega^*) \leq \mathcal{L}^p(B(u + \delta\nu_{\Omega^*}(u), \delta) \cap \Omega^*)\) and that \(0 < \mathcal{L}^p(B(u, \epsilon') \setminus \Omega^*) \leq \mathcal{L}^p(B(u, \epsilon) \setminus \Omega^*) \leq \mathcal{L}^p(B(u + \delta\nu_{\Omega^*}(u), \delta) \setminus \Omega^*).\) Therefore, \(p^A(u + \delta\nu_{\Omega^*}(u); \delta) = \frac{\mathcal{L}^p(B(u + \delta\nu_{\Omega^*}(u), \delta) \setminus \Omega^*)}{\mathcal{L}^p(B(u + \delta\nu_{\Omega^*}(u), \delta))} \in (0, 1).\)

\[\boxed{\text{Step B.1.3.} \text{ Let } I_i(\delta) = 1\{p^A(X_i; \delta) \in (0, 1)\}. \text{ Let } g : \mathbb{R}^p \to \mathbb{R} \text{ be a function that is bounded on } N(\partial\Omega^*, \delta') \cap N(\mathcal{X}, \delta') \text{ for some } \delta' > 0. \text{ Then, for } l \geq 0, \text{ there exist } \tilde{\delta} > 0 \text{ and constant } C > 0 \text{ such that} \}
\]
\[
    |\delta^{-1}E[p^A(X_i; \delta)'g(X_i)I_i(\delta)]| \leq C
\]
for every \(\delta \in (0, \tilde{\delta}).\) If \(g\) is continuous on \(N(\partial\Omega^*, \delta') \cap N(\mathcal{X}, \delta')\) for some \(\delta' > 0,\) then
\[
\delta^{-1}E[p^A(X_i; \delta)'g(X_i)I_i(\delta)] = \int_{-1}^{1} k(v)'dv \int_{\partial\Omega^*} g(x)f_X(x)d\mathcal{H}^{p-1}(x) + o(1)
\]
\[
\delta^{-1}E[Z_i p^A(X_i; \delta)'g(X_i)I_i(\delta)] = \int_{0}^{1} k(v)'dv \int_{\partial\Omega^*} g(x)f_X(x)d\mathcal{H}^{p-1}(x) + o(1)
\]
for \(l \geq 0.\) Furthermore, if \(g\) is continuously differentiable and \(\nabla g\) is bounded on \(N(\partial\Omega^*, \delta') \cap N(\mathcal{X}, \delta')\) for some \(\delta' > 0,\) then for \(l \geq 0,\)
\[
\delta^{-1}E[p^A(X_i; \delta)'g(X_i)I_i(\delta)] = \int_{-1}^{1} k(v)'dv \int_{\partial\Omega^*} g(x)f_X(x)d\mathcal{H}^{p-1}(x) + O(\delta)
\]
\[
\delta^{-1}E[Z_i p^A(X_i; \delta)'g(X_i)I_i(\delta)] = \int_{0}^{1} k(v)'dv \int_{\partial\Omega^*} g(x)f_X(x)d\mathcal{H}^{p-1}(x) + O(\delta).
\]
43
Proof. Let \( \tilde{\delta} \) be given in Step B.1.1. Under Assumption 3 (f), there exists \( \tilde{\delta} \in (0, \tilde{\delta}) \) such that \( f_X \) is bounded, is continuously differentiable, and has bounded partial derivatives on \( N(\partial \Omega^*, 2\tilde{\delta}) \cap N(\mathcal{X}, 2\tilde{\delta}) \). Let \( \tilde{\delta} \in (0, \tilde{\delta}) \) such that both \( g \) and \( f_X \) are bounded on \( N(\partial \Omega^*, 2\tilde{\delta}) \cap N(\mathcal{X}, 2\tilde{\delta}) \). We first show \( p^A(x; \delta) \in \{0, 1\} \) for every \( x \in \mathcal{X} \setminus N(\partial \Omega^*, \tilde{\delta}) \) for every \( \delta \in (0, \tilde{\delta}) \). Pick \( x \in \mathcal{X} \setminus N(\partial \Omega^*, \tilde{\delta}) \) and \( \delta \in (0, \tilde{\delta}) \). Since \( B(x, \delta) \cap \partial \Omega^* = \emptyset \), either \( B(x, \delta) \subset \text{int}(\Omega^*) \) or \( B(x, \delta) \subset \text{int}(\mathbb{R}^p \setminus \Omega^*) \). If \( B(x, \delta) \subset \text{int}(\Omega^*) \), \( p^A(x; \delta) = 1 \). If \( B(x, \delta) \subset \text{int}(\mathbb{R}^p \setminus \Omega^*) \), \( p^A(x; \delta) = 0 \), since \( A(x') = 0 \) for almost every \( x' \in B(x, \delta) \subset N(\mathcal{X}, 3\tilde{\delta}) \setminus \Omega^* \) by the choice of \( \tilde{\delta} \). Thus \( \{x \in \mathcal{X} : p^A(x; \delta) \in (0, 1)\} \subset N(\partial \Omega^*, \tilde{\delta}) \) for every \( \delta \in (0, \tilde{\delta}) \). By this and Lemma C.3, for \( \delta \in (0, \tilde{\delta}) \),

\[
\delta^{-1} E[p^A(X_i; \delta)^l g(X_i) I_i(\delta)] = \delta^{-1} \int p^A(x; \delta)^l g(x) 1\{p^A(x; \delta) \in (0, 1)\} f_X(x) dx \\
= \delta^{-1} \int_{N(\partial \Omega^*, \delta)} p^A(x; \delta)^l g(x) 1\{p^A(x; \delta) \in (0, 1)\} f_X(x) dx \\
= \delta^{-1} \int_{\partial \Omega^*} p^A(u + \lambda \nu_{\Omega^*}(u); \delta)^l g(u + \lambda \nu_{\Omega^*}(u)) 1\{p^A(u + \lambda \nu_{\Omega^*}(u); \delta) \in (0, 1)\} \\
\times f_X(u + \lambda \nu_{\Omega^*}(u)) J_{p-1}^{\Omega^*} \psi_{\Omega^*}(u, \lambda) d\mathcal{H}^{p-1}(u) d\lambda.
\]

With change of variables \( v = \frac{\lambda}{\delta} \), the last expression equals

\[
\int_{-1}^{1} \int_{\partial \Omega^*} p^A(u + \delta \nu_{\Omega^*}(u); \delta)^l 1\{p^A(u + \delta \nu_{\Omega^*}(u); \delta) \in (0, 1)\} \\
\times g(u + \delta \nu_{\Omega^*}(u)) f_X(u + \delta \nu_{\Omega^*}(u)) J_{p-1}^{\Omega^*} \psi_{\Omega^*}(u, \delta v) d\mathcal{H}^{p-1}(u) dv.
\]

For every \( (u, v, \delta) \in \partial \Omega^* \setminus N(\mathcal{X}, \tilde{\delta}) \times (-1, 1) \times (0, \tilde{\delta}) \), \( u + \delta \nu_{\Omega^*}(u) \notin \mathcal{X} \), so

\[
\delta^{-1} E[p^A(X_i; \delta)^l g(X_i) I_i(\delta)] = \int_{-1}^{1} \int_{\partial \Omega^* \cap N(\mathcal{X}, \tilde{\delta})} p^A(u + \delta \nu_{\Omega^*}(u); \delta)^l 1\{p^A(u + \delta \nu_{\Omega^*}(u); \delta) \in (0, 1)\} \\
\times g(u + \delta \nu_{\Omega^*}(u)) f_X(u + \delta \nu_{\Omega^*}(u)) J_{p-1}^{\Omega^*} \psi_{\Omega^*}(u, \delta v) d\mathcal{H}^{p-1}(u) dv \\
= \int_{-1}^{1} \int_{\partial \Omega^* \cap N(\mathcal{X}, \tilde{\delta})} (k(v) + \delta r(u, v, \delta))^l \\
\times g(u + \delta \nu_{\Omega^*}(u)) f_X(u + \delta \nu_{\Omega^*}(u)) J_{p-1}^{\Omega^*} \psi_{\Omega^*}(u, \delta v) d\mathcal{H}^{p-1}(u) dv,
\]

44
where the second equality is by Steps B.1.1 and B.1.2. By Lemma C.3, \( J_{\mu-1}^{\partial \Omega^*} \psi_{\Omega^*} (\cdot, \cdot) \) is bounded on \( \partial \Omega^* \times (-\tilde{\delta}, \tilde{\delta}) \). Since \( r, g \) and \( f_X \) are bounded, for some constant \( C > 0 \),

\[
|\delta^{-1} E[p^A(X_i; \delta)\]g(X_i)I(\delta)]| \leq C \int_{-1}^{1} \int_{\partial \Omega^* \cap N(\mathcal{X}, \tilde{\delta})} d\mathcal{H}^{p-1}(u) dv,
\]

which is finite by Assumption 3 (e) (i). Moreover, if \( g \) and \( f_X \) are continuous on \( N(\partial \Omega^*, 2\tilde{\delta}) \cap N(\mathcal{X}, 2\tilde{\delta}) \), by the Dominated Convergence Theorem,

\[
\delta^{-1} E[p^A(X_i; \delta)\]g(X_i)I(\delta)] \to \int_{-1}^{1} k(v)dv \int_{\partial \Omega^*} g(u)f_X(u)d\mathcal{H}^{p-1}(u),
\]

where we use the fact from Lemma C.3 that \( J_{\mu-1}^{\partial \Omega^*} \psi_{\Omega^*} (u, \lambda) \) is continuous in \( \lambda \) and 
\( J_{\mu-1}^{\partial \Omega^*} \psi_{\Omega^*} (u, 0) = 1 \).

Note that \( A(x) = 1 \) for every \( x \in \Omega^* \) and \( A(x) = 0 \) for almost every \( x \in N(\mathcal{X}, 2\tilde{\delta}) \setminus \Omega^* \). Also, for every \( (u, v, \delta) \in \partial \Omega^* \cap N(\mathcal{X}, \tilde{\delta}) \times (-1, 1) \times (0, \tilde{\delta}) \), \( u + \delta v \nu_{\Omega^*}(u) \in \Omega^* \) if \( v \in (0, 1) \) and \( u + \delta v \nu_{\Omega^*}(u) \in N(\mathcal{X}, 2\tilde{\delta}) \setminus \Omega^* \) if \( v \in (-1, 0) \). Therefore,

\[
\delta^{-1} E[Z_i p^A(X_i; \delta)\]g(X_i)I(\delta)] = \delta^{-1} E[A(X_i)p^A(X_i; \delta)\]g(X_i)I(\delta)]
\]

\[
= \int_{-1}^{1} \int_{\partial \Omega^* \cap N(\mathcal{X}, \tilde{\delta})} A(u + \delta v \nu_{\Omega^*}(u))(k(v) + \delta r(u, v, \delta))g(u + \delta v \nu_{\Omega^*}(u))
\]

\[
\times f_X(u + \delta v \nu_{\Omega^*}(u)) J_{\mu-1}^{\partial \Omega^*} \psi_{\Omega^*}(u, \delta v) d\mathcal{H}^{p-1}(u) dv
\]

\[
= \int_{0}^{1} \int_{\partial \Omega^* \cap N(\mathcal{X}, \tilde{\delta})} (k(v) + \delta r(u, v, \delta))g(u + \delta v \nu_{\Omega^*}(u))
\]

\[
\times f_X(u + \delta v \nu_{\Omega^*}(u)) J_{\mu-1}^{\partial \Omega^*} \psi_{\Omega^*}(u, \delta v) d\mathcal{H}^{p-1}(u) dv
\]

\[
\to \int_{0}^{1} k(v)dv \int_{\partial \Omega^*} g(u)f_X(u)d\mathcal{H}^{p-1}(u).
\]

Now suppose that \( g \) and \( f_X \) are continuously differentiable on \( N(\partial \Omega^*, 2\tilde{\delta}) \cap N(\mathcal{X}, 2\tilde{\delta}) \) and that \( \nabla g \) and \( \nabla f \) are bounded on \( N(\partial \Omega^*, 2\tilde{\delta}) \cap N(\mathcal{X}, 2\tilde{\delta}) \). Using the mean-value theorem, we obtain that, for any \( (u, v, \delta) \in \partial \Omega^* \cap N(\mathcal{X}, \tilde{\delta}) \times (-1, 1) \times (0, \tilde{\delta}) \),

\[
g(u + \delta v \nu_{\Omega^*}(u)) = g(u) + \nabla g(y_g(u, \delta v \nu_{\Omega^*}(u)))' \delta v \nu_{\Omega^*}(u),
\]

\[
f_X(u + \delta v \nu_{\Omega^*}(u)) = f_X(u) + \nabla f_X(y_f(u, \delta v \nu_{\Omega^*}(u)))' \delta v \nu_{\Omega^*}(u)
\]

45
for some $y_g(u, \delta v \nu_{\Omega^*}(u))$ and $y_f(u, \delta v \nu_{\Omega^*}(u))$ that are on the line segment connecting $u$ and $u + \delta v \nu_{\Omega^*}(u)$. In addition,

$$J_{p-1}^{\partial \Omega^*} \psi_{\Omega^*}(u, \delta v) = J_{p-1}^{\partial \Omega^*} \psi_{\Omega^*}(u, 0) + \frac{\partial J_{p-1}^{\partial \Omega^*} \psi_{\Omega^*}(u, y_f(u, \delta v))}{\partial \lambda} \delta v$$

$$= 1 + \frac{\partial J_{p-1}^{\partial \Omega^*} \psi_{\Omega^*}(u, y_f(u, \delta v))}{\partial \lambda} \delta v$$

for some $y_f(u, \delta v)$ that is on the line segment connecting $0$ and $\delta v$. By Lemma C.3, $\frac{\partial J_{p-1}^{\partial \Omega^*} \psi_{\Omega^*}(\cdot)}{\partial \lambda}$ is bounded on $\partial \Omega^* \times (-\tilde{\delta}, \tilde{\delta})$. We then have

$$\delta^{-1} E[p^A(X_i; \delta)^l g(X_i) I_i(\delta)]$$

$$= \int_{-1}^{1} \int_{\partial \Omega^* \cap N(X, \delta)} (k(v) + \delta r(u, v, \delta))(|g(u) + \nabla g(y_g(u, \delta v \nu_{\Omega^*}(u))|) \delta v \nu_{\Omega^*}(u))$$

$$\times (f_X(u) + \nabla f_X(y_f(u, \delta v \nu_{\Omega^*}(u))|) \delta v \nu_{\Omega^*}(u))(1 + \frac{\partial J_{p-1}^{\partial \Omega^*} \psi_{\Omega^*}(u, y_f(u, \delta v))}{\partial \lambda} \delta v) d\mathcal{H}^{p-1}(u) dv$$

$$= \int_{-1}^{1} \int_{\partial \Omega^* \cap N(X, \delta)} (k(v) g(u) f_X(u) + \delta h(u, v, \delta)) d\mathcal{H}^{p-1}(u) dv$$

$$= \int_{-1}^{1} k(v)^l dv \int_{\partial \Omega^*} g(u) f_X(u) d\mathcal{H}^{p-1}(u) + \delta \int_{-1}^{1} \int_{\partial \Omega^* \cap N(X, \delta)} h(u, v, \delta) d\mathcal{H}^{p-1}(u) dv$$

for some function $h$ bounded on $\partial \Omega^* \cap N(X, \tilde{\delta}) \times (-1, 1) \times (0, \tilde{\delta})$. It then follows that

$$\delta^{-1} E[p^A(X_i; \delta)^l g(X_i) I_i(\delta)] = \int_{-1}^{1} k(v)^l dv \int_{\partial \Omega^*} g(u) f_X(u) d\mathcal{H}^{p-1}(u) + O(\delta).$$

Similarly,

$$\delta^{-1} E[Z_i p^A(X_i; \delta)^l g(X_i) I_i(\delta)] = \int_{0}^{1} k(v)^l dv \int_{\partial \Omega^*} g(u) f_X(u) d\mathcal{H}^{p-1}(u) + O(\delta). \quad \square$$

**Step B.1.4.** Let $S_D = \lim_{\delta \to 0} \delta^{-1} E[Z_i D_i 1\{p^A(X_i; \delta) \in (0, 1)\}]$ and $S_Y = \lim_{\delta \to 0} \delta^{-1} E[Z_i Y_i 1\{p^A(X_i; \delta) \in (0, 1)\}]$. Then the second element of $S_D^{-1} S_Y$ is $\beta_1$.

**Proof.** Note that $D_i = Z_i D_i(1) + (1 - Z_i) D_i(0)$ and $Y_i = Z_i Y_i(1) + (1 - Z_i) Y_i(0)$. By Step
B.1.3, it then follows that $x^p$ for any sufficiently small $\delta > 0$ so that $B(\bar{S}, 0) \subset A$.

$$S_D = \begin{bmatrix}
2\tilde{f}_X & \int_{\partial \Omega^*} E[D_i(1) - D_i(0)|X_i = x]f_X(x)d\mathcal{H}^{p-1}(x) & \int_{-1}^1 k(v)dv \tilde{f}_X \\
\tilde{f}_X & \int_{\partial \Omega^*} E[D_i(1)|X_i = x]f_X(x)d\mathcal{H}^{p-1}(x) & \int_{0}^1 k(v)dv \tilde{f}_X \\
\int_{-1}^1 k(v)dv \tilde{f}_X & \int_{\partial \Omega^*} E[D_i(1)|X_i = x]f_X(x)d\mathcal{H}^{p-1}(x) & \int_{-1}^1 k(v)^2dv \tilde{f}_X
\end{bmatrix},$$

where $\tilde{f}_X = \int_{\partial \Omega^*} f_X(x)d\mathcal{H}^{p-1}(x)$. After a few lines of algebra, we have

$$\det(S_D) = \tilde{f}_X^2 \int_{\partial \Omega^*} E[D_i(1) - D_i(0)|X_i = x]f_X(x)d\mathcal{H}^{p-1}(x)$$

$$\times \left( \int_{-1}^0 (k(v) - \int_{-1}^0 k(s)ds)^2dv + \int_{0}^1 (k(v) - \int_{0}^1 k(s)ds)^2dv \right).$$

We verify that $\det(S_D)$ is nonzero. Since $\tilde{f}_X > 0$ by Assumption 3 (e) (i), it suffices to show that $\int_{\partial \Omega^*} E[D_i(1) - D_i(0)|X_i = x]f_X(x)d\mathcal{H}^{p-1}(x) \neq 0$. To do so, we first show that $p^A(x) \in \{0, 1\}$ for every $x \in \mathcal{X} \setminus \partial \Omega^*$. Pick $x \in \mathcal{X} \setminus \partial \Omega^*$. By definition, either $x \in \mathcal{X} \cap \text{int}(\Omega^*)$ or $x \in \mathcal{X} \cap (\mathbb{R}^p \setminus \text{cl}(\Omega^*))$. If $x \in \mathcal{X} \cap \text{int}(\Omega^*)$, then $B(x, \delta) \subset \text{int}(\Omega^*)$ for any sufficiently small $\delta > 0$ so that $p^A(x) = 1$. If $x \in \mathcal{X} \cap (\mathbb{R}^p \setminus \text{cl}(\Omega^*))$, then $B(x, \delta) \subset N(\mathcal{X}, \delta') \cap (\mathbb{R}^p \setminus \text{cl}(\Omega^*))$ for any sufficiently small $\delta > 0$, where $\delta' > 0$ satisfies Assumption 3 (e) (ii). Since $A(x') = 0$ for almost every $x' \in N(\mathcal{X}, \delta') \setminus \Omega^*$, $p^A(x) = 0$. Note also that $p^A(x) = \lim_{\delta \to 0} p^A(x; \delta) = k(0) = \frac{1}{2}$ for every $x \in \partial \Omega^* \cap \mathcal{X}$ by Step B.1.1. It then follows that

$$\int_{\partial \Omega^*} E[D_i(1) - D_i(0)|X_i = x]f_X(x)d\mathcal{H}^{p-1}(x)$$

$$= 4 \int_{\partial \Omega^* \cap \mathcal{X}} p^A(x)(1 - p^A(x))E[D_i(1) - D_i(0)|X_i = x]f_X(x)d\mathcal{H}^{p-1}(x)$$

$$= 4 \int_{\mathcal{X}} p^A(x)(1 - p^A(x))E[D_i(1) - D_i(0)|X_i = x]f_X(x)d\mathcal{H}^{p-1}(x),$$

47
which is nonzero under Assumption 3 (b). The second element of $S_D^{-1}S_Y$ is

$$
\int_{\partial \Omega^*} E[(D_i(1) - D_i(0))(Y_i(1) - Y_i(0))|X_i = x]f_X(x)d\mathcal{H}^{p-1}(x) 
\int_{\partial \Omega^*} E[D_i(1) - D_i(0)|X_i = x]f_X(x)d\mathcal{H}^{p-1}(x).
$$

On the other hand, by Step B.1.3,

$$
\beta_1 = \lim_{\delta \to 0} E[\omega_i(\delta)(Y_i(1) - Y_i(0))]
\leq \lim_{\delta \to 0} \delta^{-1}E[p^A(X_i; \delta)(1-p^A(X_i; \delta))(D_i(1) - D_i(0))(Y_i(1) - Y_i(0))I_i(\delta)]
\leq \lim_{\delta \to 0} \delta^{-1}E[p^A(X_i; \delta)(1-p^A(X_i; \delta))(D_i(1) - D_i(0))I_i(\delta)]
\leq \lim_{\delta \to 0} \frac{\int_{-1}^1 k(v)(1-k(v))dv \int_{\partial \Omega^*} E[(D_i(1) - D_i(0))(Y_i(1) - Y_i(0))|X_i = x]f_X(x)d\mathcal{H}^{p-1}(x)}{\int_{\partial \Omega^*} E[D_i(1) - D_i(0)|X_i = x]f_X(x)d\mathcal{H}^{p-1}(x)},
$$

where $I_i(\delta) = 1\{p^A(X_i; \delta) \in (0, 1)\}$. \hfill \Box

**Step B.1.5.** If $\delta_n \to 0$ and $n\delta_n \to \infty$ as $n \to \infty$, then $\hat{\beta}_1 \overset{p}{\to} \beta_1$.

**Proof.** It suffices to verify the variance of each element of $\frac{1}{n\delta_n} \sum_{i=1}^n Z_i D'_i I_i$ and $\frac{1}{n\delta_n} \sum_{i=1}^n Z_i Y_i I_i$ is $o(1)$. We only verify $\text{Var}(\frac{1}{n\delta_n} \sum_{i=1}^n p^A(X_i; \delta_n)Y_i I_i) = o(1)$. Note $E[Y_i^2 | X_i] = E[Z_i Y_i^2 + (1 - Z_i) Y_i^2|X_i] \leq E[Y_i^2 + Y_i^2|X_i]$. Under Assumption 3 (f), there exists $\delta' > 0$ such that $E[Y_i^2 + Y_i^2|X_i]$ is continuous on $N(\partial \Omega^*, \delta')$. Since $\text{cl}(N(\partial \Omega^*, \frac{1}{2}\delta'))$ is closed and bounded, $E[Y_i^2 + Y_i^2|X_i]$ is bounded on $\text{cl}(N(\partial \Omega^*, \frac{1}{2}\delta'))$. For some $C > 0$, for any sufficiently large $n$,

$$
\text{Var}\left(\frac{1}{n\delta_n} \sum_{i=1}^n p^A(X_i; \delta_n)Y_i I_i\right) \leq \frac{1}{n\delta_n} \delta_n^{-1}E[p^A(X_i; \delta_n)^2 E[Y_i^2|X_i] I_i] \leq \frac{1}{n\delta_n} C,
$$

where the last inequality holds by Step B.1.3. The conclusion follows since $n\delta_n \to \infty$. \hfill \Box

Now let $\beta = (\beta_0, \beta_1, \beta_2)' = S_D^{-1}S_Y$ and let $\epsilon_i = Y_i - D'_i \beta$. We can write

$$
\sqrt{n\delta_n} (\hat{\beta} - \beta) = (\frac{1}{n\delta_n} \sum_{i=1}^n Z_i D'_i I_i)^{-1} \frac{1}{\sqrt{n\delta_n}} \sum_{i=1}^n \{(Z_i \epsilon_i I_i - E[Z_i \epsilon_i I_i]) + E[Z_i \epsilon_i I_i]\}.
$$

48
Step B.1.6. \( \frac{1}{\sqrt{n\delta_n}} \sum_{i=1}^{n}(Z_i\epsilon_iI_i - E[Z_i\epsilon_iI_i]) \xrightarrow{d} \mathcal{N}(0, V) \), where \( V = \lim_{n \to \infty} \delta_n^{-1}E[\epsilon_i^2Z_iZ_iI_i] \).

Proof. We use the triangular-array Lyapunov CLT and the Cramér-Wold device. Pick a nonzero \( \lambda \in \mathbb{R}^p \), and let \( V_{i,n} = \frac{1}{\sqrt{n\delta_n}}\lambda'(Z_i\epsilon_iI_i - E[Z_i\epsilon_iI_i]) \). First, by Step B.1.3, \( E[Z_i\epsilon_iI_i] = E[Z_i(Y_i - D_i'(\beta)I_i)] = O(\delta_n) \), so \( \delta_n^{-1}E[Z_i\epsilon_iI_i]E[Z_i\epsilon_iI_i] = o(1) \). We have

\[
E[\epsilon_i^2Z_iZ_iI_i] = E[(Y_i - \beta_0 - \beta_1D_i - \beta_2p^4(X_i; \delta_n))^2Z_iZ_iI_i]
\]

\[
= E[Z_i(Y_i - \beta_0 - \beta_1D_i(1) - \beta_2p^4(X_i; \delta_n))^2Z_iZ_iI_i]
+ E[(1 - Z_i)(Y_{0i} - \beta_0 - \beta_1D_i(0) - \beta_2p^4(X_i; \delta_n))^2Z_iZ_iI_i].
\]

Since \( E[Y_{i1}|X_i] \), \( E[Y_{0i}|X_i] \), \( E[D_i(1)|X_i] \), \( E[D_i(0)|X_i] \), \( E[Y_{i1}^2|X_i] \), \( E[Y_{01}^2|X_i] \), \( E[Y_{1i}D_i(1)|X_i] \) and \( E[Y_{0i}D_i(0)|X_i] \) are continuous on \( N(\partial \Omega^*, \delta') \) for some \( \delta' > 0 \) under Assumption 3 (f), \( \lim_{n \to \infty} \delta_n^{-1}E[\epsilon_i^2Z_iZ_iI_i] \) exists and finite. Therefore, \( \sum_{i=1}^{n} E[V_{i,n}^2] = \delta_n^{-1}\lambda'(E[\epsilon_i^2Z_iZ_iI_i] - E[Z_i\epsilon_iI_i]E[Z_i\epsilon_iI_i])\lambda \to \lambda'V\lambda \).

We next verify the Lyapunov condition: for some \( t > 0 \), \( \sum_{i=1}^{n} E[|V_{i,n}|^{2+t}] \to 0 \).

Note

\[
\sum_{i=1}^{n} E[|V_{i,n}|^4] = \frac{1}{n\delta_n} \delta_n^{-1}E[|\lambda'(Z_i\epsilon_iI_i - E[Z_i\epsilon_iI_i])|^4]
\leq \frac{2^3}{n\delta_n} \delta_n^{-1} \{ E[|\lambda'Z_i\epsilon_iI_i|^4] + |\lambda'|E[Z_i\epsilon_iI_i]|^4 \}
\]

by the \( c_r \)-inequality. Repeating using the \( c_r \)-inequality gives

\[
\delta_n^{-1}E[|\lambda'Z_i\epsilon_iI_i|^4] = \delta_n^{-1}E[|\lambda'Z_i(Y_i - \beta_0 - \beta_1D_i - \beta_2p^4(X_i; \delta_n))|^4I_i]
\leq 2^3\delta_n^{-1}E[|\lambda'|^4(|Y_i|^4 + |\beta_0|^4 + |\beta_1|^4|I_i| + |\beta_2|^4p^4(X_i; \delta_n)^4I_i)]
\leq 2^3(|\lambda_1| + |\lambda_2| + |\lambda_3|)\delta_n^{-1}E[(Y_i'^4 + \beta_0'^4 + \beta_1'^4 + \beta_2'^4)I_i] = 2^3cO(1)
\]

for some finite constant \( c \), where the last equality holds by Step B.1.3 under Assumption 3 (f). Moreover, \( \delta_n^{-1}|\lambda'|E[Z_i\epsilon_iI_i]|^4 = \delta_n^3|\lambda'|\delta_n^{-1}E[Z_i\epsilon_iI_i]|^4 = \delta_n^3O(1) = o(1) \). Therefore, when \( n\delta_n \to \infty \), \( \sum_{i=1}^{n} E[|V_{i,n}|^4] \to 0 \), and the conclusion follows from the Lyapunov CLT and the Cramér-Wold device.

Step B.1.7. \( n\delta_n \tilde{\Sigma} \xrightarrow{p} S_D^{-1}V(S_D')^{-1} \).
Proof. Using the result that \( \hat{\beta} - \beta = o_p(1) \) and Step B.1.3, we have

\[
\frac{1}{n \delta_n} \sum_{i=1}^{n} \epsilon_i^2 Z_i Z_i' I_i = \frac{1}{n \delta_n} \sum_{i=1}^{n} (Y_i - D_i \hat{\beta})^2 Z_i Z_i' I_i = \frac{1}{n \delta_n} \sum_{i=1}^{n} (\epsilon_i - D_i (\hat{\beta} - \beta))^2 Z_i Z_i' I_i
\]

\[
= \frac{1}{n \delta_n} \sum_{i=1}^{n} \epsilon_i^2 Z_i Z_i' I_i + \frac{1}{n \delta_n} \sum_{i=1}^{n} ((\hat{\beta}_0 - \beta_0) + D_i (\hat{\beta}_1 - \beta_1) + p^A(X_i; \delta_n)(\hat{\beta}_2 - \beta_2))^2 Z_i Z_i' I_i
\]

\[
- \frac{2}{n \delta_n} \sum_{i=1}^{n} (Y_i - D_i \beta)((\hat{\beta}_0 - \beta_0) + D_i (\hat{\beta}_1 - \beta_1) + p^A(X_i; \delta_n)(\hat{\beta}_2 - \beta_2)) Z_i Z_i' I_i
\]

\[
= \frac{1}{n \delta_n} \sum_{i=1}^{n} \epsilon_i^2 Z_i Z_i' I_i + o_p(1)O_p(1).
\]

To show \( \frac{1}{n \delta_n} \sum_{i=1}^{n} \epsilon_i^2 Z_i Z_i' I_i \xrightarrow{p} V \), it suffices to verify that the variance of each element of \( \frac{1}{n \delta_n} \sum_{i=1}^{n} \epsilon_i^2 Z_i Z_i' I_i \) is \( o(1) \). We only verify that \( \text{Var}(\frac{1}{n \delta_n} \sum_{i=1}^{n} \epsilon_i^2 p^A(X_i; \delta_n)^2 I_i) = o(1) \). Using the \( c_r \)-inequality, we have that for some constant \( c \),

\[
\text{Var} \left( \frac{1}{n \delta_n} \sum_{i=1}^{n} \epsilon_i^2 p^A(X_i; \delta_n)^2 I_i \right) \leq \frac{1}{n \delta_n} \delta_n^{-1} E[\epsilon_i^4 I_i]
\]

\[
= \frac{1}{n \delta_n} \delta_n^{-1} E[(Y_i - \beta_0 - \beta_1 D_i - \beta_2 p^A(X_i))^4 I_i]
\]

\[
\leq \frac{1}{n \delta_n} 2^{3c} \delta_n^{-1} E[(Y_i^4 + \beta_0^4 + \beta_1^4 D_i + \beta_2^4 p^A(X_i)^4) I_i] = \frac{1}{n \delta_n} 2^{3c} O(1) = o(1),
\]

where the second last equality holds by Step B.1.3 under Assumption 3 (f). Thus \( \frac{1}{n \delta_n} \sum_{i=1}^{n} \epsilon_i^2 Z_i Z_i' I_i \xrightarrow{p} V \). It follows that

\[
n \delta_n \hat{\Sigma} = \left( \frac{1}{n \delta_n} \sum_{i=1}^{n} Z_i D_i' I_i \right)^{-1} \left( \frac{1}{n \delta_n} \sum_{i=1}^{n} \epsilon_i^2 Z_i Z_i' I_i \right) \left( \frac{1}{n \delta_n} \sum_{i=1}^{n} D_i Z_i' I_i \right)^{-1} \xrightarrow{p} S_D^{-1} V(S_D')^{-1}.
\]

\( \square \)

**Step B.1.8.** \( \hat{\sigma}^{-1}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} \mathcal{N}(0, 1) \).
Proof. Let \( \beta_n = S_D^{-1} \delta_n^{-1} E[Z_i Y_i I_i] \). Since \( \epsilon_i = Y_i - D' \beta = Y_i - D_i' \beta_n + D_i'(\beta_n - \beta) \),

\[
\frac{1}{\sqrt{n \delta_n}} \sum_{i=1}^{n} E[Z_i \epsilon_i I_i] = \sqrt{n \delta_n \delta_n^{-1} E[Z_i (Y_i - D' \beta) I_i]}
\]

\[
= \sqrt{n \delta_n \delta_n^{-1}} \{ E[Z_i Y_i I_i] - E[Z_i D_i' I_i] \beta_n + E[Z_i D_i' I_i](\beta_n - \beta) \}
\]

\[
= \sqrt{n \delta_n \{ (S_D - \delta_n^{-1} E[Z_i D_i' I_i]) S_D^{-1} \delta_n^{-1} E[Z_i Y_i I_i] \\
+ \delta_n^{-1} E[Z_i D_i' I_i] S_D^{-1} (\delta_n^{-1} E[Z_i Y_i I_i] - S_Y) \}}
\]

\[
= \sqrt{n \delta_n (O(\delta_n) O(1) + O(1)O(\delta_n)) = O(\sqrt{n \delta_n \delta_n})},
\]

where we use Step B.1.3 for the second last equality. Thus, when \( n \delta_n^3 \to 0 \),

\[
\sqrt{n \delta_n (\hat{\beta} - \beta)} = \left( \frac{1}{n \delta_n} \sum_{i=1}^{n} Z_i D_i' I_i \right)^{-1} \frac{1}{\sqrt{n \delta_n}} \sum_{i=1}^{n} \{(Z_i \epsilon_i I_i - E[Z_i \epsilon_i I_i]) + E[Z_i \epsilon_i I_i] \}
\]

\[
\xrightarrow{d} \mathcal{N}(0, S_D^{-1} \mathbf{V}(S_D')^{-1}).
\]

The conclusion then follows from Step B.1.7. \( \square \)

References


C Lemmas

Lemmas C.1, C.2, and C.3 are variations of existing results in differential geometry and geometric measure theory. Hence, their proofs are omitted here. Self-contained proofs are available in Appendices D.1, D.2, and D.3 of Narita and Yata (2023).

C.1 Differential Geometry

We provide some concepts and facts from differential geometry of twice continuously differentiable sets, following Crasta and Malusa (2007). We use the notation introduced in Appendix B. Let $S \subset \mathbb{R}^p$ be a twice continuously differentiable set, and let

$$\Pi_{\partial S}(x) = \{ y \in \partial S : \| y - x \| = d(x, \partial S) \}$$

be the set of projections of $x$ on $\partial S$.

**Lemma C.1** (Corollary of Theorem 4.16, Crasta and Malusa (2007)). Let $S \subset \mathbb{R}^p$ be nonempty, bounded, open, connected and twice continuously differentiable. Then the function $d^*_S$ is twice continuously differentiable on $N(\partial S, \mu)$ for some $\mu > 0$. In addition, for every $x_0 \in \partial S$, $\Pi_{\partial S}(x_0 + t\nu_S(x_0)) = \{ x_0 \}$ for every $t \in (-\mu, \mu)$. Furthermore, for every $x \in N(\partial S, \mu)$, $\Pi_{\partial S}(x)$ is a singleton, $\nabla d^*_S(x) = \nu_S(y)$ and $x = y + d^*_S(x)\nu_S(y)$ for $y \in \Pi_{\partial S}(x)$, and $\| \nabla d^*_S(x) \| = 1$.

We say that a set $S \subset \mathbb{R}^p$ is an $m$-dimensional $C^1$ submanifold of $\mathbb{R}^p$ if for every point $x \in S$, there exist an open neighborhood $V \subset \mathbb{R}^p$ of $x$ and a one-to-one continuously differentiable function $\phi$ from an open set $U \subset \mathbb{R}^m$ to $\mathbb{R}^p$ such that the Jacobian matrix $J\phi(u)$ is of rank $m$ for all $u \in U$, and $\phi(U) = V \cap S$.

**Lemma C.2.** Let $S \subset \mathbb{R}^p$ be nonempty, bounded, open, connected and twice continuously differentiable. Then $\partial S$ is a $(p - 1)$-dimensional $C^1$ submanifold of $\mathbb{R}^p$. 

C.2 Geometric Measure Theory

We provide concepts and facts from geometric measure theory, following Krantz and Parks (2008). Let $S$ be an $m$-dimensional $C^1$ submanifold of $\mathbb{R}^p$. Let $x \in S$ and let $\phi : U \subset \mathbb{R}^m \rightarrow \mathbb{R}^p$ be as in the definition of $m$-dimensional $C^1$ submanifold. We denote by $T_S(x)$ the tangent space of $S$ at $x$, $\{J\phi(u)v : v \in \mathbb{R}^m\}$, where $u = \phi^{-1}(x)$.

Also, for each $x \in S$ and let $d_S^p$ is differentiable and for each $\lambda \in \mathbb{R}$, let $\psi_S(x, \lambda) = x + \lambda \nabla d_S^p(x)$. Furthermore, for a Lipschitz function $f : \mathbb{R}^p \rightarrow \mathbb{R}^\nu$ with $\nu \geq m$, let $J_m^S f(x) = \frac{\mathcal{H}^m((Jf(x))_{y \in P})}{\mathcal{H}^m(P)}$ for each $x \in \mathbb{R}^p$ at which $f$ is differentiable, where $P$ is an arbitrary $m$-dimensional parallelepiped contained in $T_S(x)$.

**Lemma C.3.** For $\Omega \subset \mathbb{R}^p$, suppose that there is a partition $\{\Omega_1, ..., \Omega_M\}$ of $\Omega$ with

(i) $\text{dist}(\Omega_m, \Omega_{m'}) > 0$ for any $m, m' \in \{1, ..., M\}$ such that $m \neq m'$;

(ii) $\Omega_m$ is nonempty, bounded, open, connected and twice continuously differentiable for each $m \in \{1, ..., M\}$.

Then there is $\mu > 0$ such that $d_\Omega^p$ is twice continuously differentiable on $N(\partial \Omega, \mu)$ and

$$
\int_{N(\partial \Omega, \delta)} g(x)dx = \int_{-\delta}^{\delta} \int_{\partial \Omega} g(u + \lambda \nu_{\Omega}(u))J_{p-1}^{\partial \Omega} \psi_{\Omega}(u, \lambda) d\mathcal{H}^{p-1}(u)d\lambda
$$

for every $\delta \in (0, \mu)$ and every function $g : \mathbb{R}^p \rightarrow \mathbb{R}$ integrable on $N(\partial \Omega, \delta)$, where for each fixed $\lambda \in (-\mu, \mu)$, $J_{p-1}^{\partial \Omega} \psi_{\Omega}(\cdot, \lambda)$ is calculated by applying the operation $J_{p-1}^{\partial \Omega}$ to the function $\psi_{\Omega}(\cdot, \lambda)$. Furthermore, $J_{p-1}^{\partial \Omega} \psi_{\Omega}(x, \cdot)$ is continuously differentiable in $\lambda$ and $J_{p-1}^{\partial \Omega} \psi_{\Omega}(x, 0) = 1$ for every $x \in \partial \Omega$, and $J_{p-1}^{\partial \Omega} \psi_{\Omega}(\cdot, \cdot)$ and $\frac{\partial J_{p-1}^{\partial \Omega} \psi_{\Omega}(\cdot, \cdot)}{\partial \lambda}$ are bounded on $\partial \Omega \times (-\mu, \mu)$.

C.3 Existence of the Approximate Propensity Score

The proofs of the following lemmas and the results on the existence of APS at nonsmooth points are available in Appendix E.1 of Narita and Yata (2023).

**Lemma C.4.** $p^A(x)$ exists and is equal to $A(x)$ for almost every $x \in \mathcal{X}$ (with respect to the Lebesgue measure).

**Lemma C.5.** If $A$ is continuous at $x \in \mathcal{X}$, then $p^A(x)$ exists and $p^A(x) = A(x)$.
C.4 Other Lemmas

Lemma C.6. Let \( \{V_i\}_{i=1}^{\infty} \) be i.i.d. random variables such that \( E[V_i^2] < \infty \). If Assumption 1 holds, then for \( l \geq 0 \) and \( m = 0, 1 \),

\[
E[V_ip^A(X_i; \delta)^l1\{p^A(X_i; \delta) \in (0, 1)\}^m] \rightarrow E[V_iA(X_i)^l1\{A(X_i) \in (0, 1)\}]^m
\]
as \( \delta \to 0 \). Moreover, if, in addition, \( \delta_n \to 0 \) as \( n \to \infty \), then for \( l \geq 0 \), as \( n \to \infty \),

\[
\frac{1}{n} \sum_{i=1}^{n} V_ip^A(X_i; \delta_n)^lI_{i,n} \overset{p}{\rightarrow} E[V_iA(X_i)^l1\{A(X_i) \in (0, 1)\}].
\]

Proof. See Appendix E.1.

Lemma C.7. Let \( \{ (\delta_n, S_n) \}_{n=1}^{\infty} \) be any sequence of positive numbers and positive integers. Fix \( x \in X \), and let \( X_{s_1}^{*}, \ldots, X_{s_n}^{*} \) be \( S_n \) independent draws from the uniform distribution on \( B(x, \delta_n) \) so that \( p^*(x; \delta_n) = \frac{1}{S_n} \sum_{s=1}^{S_n} A(X_s^*) \). Then,

\[
E[p^*(x; \delta_n) - p^A(x; \delta_n)] = 0, \quad E[(p^*(x; \delta_n) - p^A(x; \delta_n))^2] \leq \frac{1}{S_n},
\]

\[
|E[p^*(x; \delta_n)^2 - p^A(x; \delta_n)^2]| \leq \frac{1}{S_n}, \quad E[(p^*(x; \delta_n)^2 - p^A(x; \delta_n)^2)^2] \leq \frac{4}{S_n}, \quad \text{and}
\]

\[
\Pr(p^*(x; \delta_n) \in \{0, 1\}) \leq (1 - p^A(x; \delta_n))^{S_n} + p^A(x; \delta_n)^{S_n}.
\]

Moreover, for any \( \epsilon > 0 \), \( E[|p^*(x; \delta_n) - p^A(x; \delta_n)|] \leq \frac{1}{S_n \epsilon^2} + \epsilon \), and if \( S_n \to \infty \), then \( E[|p^*(x; \delta_n) - p^A(x; \delta_n)|] \to 0 \) as \( n \to \infty \).

Proof. See Appendix E.2.

Lemma C.8. Suppose \( \Pr(A(X_i) \in (0, 1)) > 0 \). Let \( I_{i,n}^* = 1\{p^*(X_i; \delta_n) \in (0, 1)\} \) and \( \{V_i\}_{i=1}^{\infty} \) be i.i.d. random variables such that \( E[V_i^2] < \infty \). If Assumption 1 holds, \( S_n \to \infty \), and \( \delta_n \to 0 \), then

\[
\frac{1}{n} \sum_{i=1}^{n} V_ip^*(X_i; \delta_n)^lI_{i,n}^* - \frac{1}{n} \sum_{i=1}^{n} V_ip^A(X_i; \delta_n)^lI_{i,n} = o_p(1)
\]

for \( l = 0, 1, 2, 3, 4 \). If, in addition, Assumptions A.1 and A.2 hold, \( n \delta_n^2 \to 0 \), and
\[ E[V_i|X_i] \text{ is bounded, then for } l = 0, 1, 2, \]
\[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i p^*(X_i; \delta_n) l_{i,n} - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i p^A(X_i; \delta_n) l_{i,n} = o_p(1). \]

Proof. See Appendix E.3. \( \square \)

D Proofs of Main Results

D.1 Proof of Proposition 1

Suppose that Assumptions 1 and 2 hold. Here, we only show that

(a) \( E[Y_{i1} - Y_{0i}|X_i = x] \) is identified for every \( x \in \text{int}(\mathcal{X}) \) such that \( p^A(x) \in (0, 1) \).

(b) Let \( S \) be any open subset of \( \mathcal{X} \) such that \( p^A(x) \) exists for all \( x \in S \). Then \( E[Y_{i1} - Y_{0i}|X_i \in S] \) is identified only if \( p^A(x) \in (0, 1) \) for almost every \( x \in S \).

The proofs for \( E[D_i(1) - D_i(0)|X_i = x] \) and \( E[D_i(1) - D_i(0)|X_i \in S] \) are similar.

Proof of Part (a). Pick \( x \in \text{int}(\mathcal{X}) \) with \( p^A(x) \in (0, 1) \). If \( A(x) \in (0, 1) \), \( E[Y_{i1} - Y_{0i}|X_i = x] \) is identified by Property 1: \( E[Y_i|X_i = x, Z_i = 1] - E[Y_i|X_i = x, Z_i = 0] = E[Y_{i1} - Y_{0i}|X_i = x] \). Next consider the case where \( A(x) \in \{0, 1\} \). Since \( x \in \text{int}(\mathcal{X}) \), \( B(x, \delta) \subset \mathcal{X} \) for any sufficiently small \( \delta > 0 \). Since \( p^A(x) = \lim_{\delta \to 0} p^A(x; \delta) \in (0, 1) \), \( p^A(x; \delta) \in (0, 1) \) for any sufficiently small \( \delta > 0 \). We can find points \( x_{0, \delta}, x_{1, \delta} \in B(x, \delta)(\subset \mathcal{X}) \) such that \( A(x_{0, \delta}) < 1 \) and \( A(x_{1, \delta}) > 0 \) for any sufficiently small \( \delta > 0 \), for otherwise \( p^A(x; \delta) \in \{0, 1\} \). Since \( x_{0, \delta} \to x \) and \( x_{1, \delta} \to x \) as \( \delta \to 0 \),

\[ \lim_{\delta \to 0} (E[Y_i|X_i = x_{1, \delta}, Z_i = 1] - E[Y_i|X_i = x_{0, \delta}, Z_i = 0]) \]
\[ = \lim_{\delta \to 0} (E[Y_{i1}|X_i = x_{1, \delta}] - E[Y_{0i}|X_i = x_{0, \delta}]) = E[Y_{i1} - Y_{0i}|X_i = x], \]

where the two equalities follow from Property 1 and Assumption 2, respectively. \( \square \)

Proof of Part (b). Suppose to the contrary that \( \mathcal{L}^p(\{x \in S : p^A(x) \in \{0, 1\}\}) > 0 \). Without loss of generality, assume \( \mathcal{L}^p(\{x \in S : p^A(x) = 1\}) > 0 \).
Step D.1.1. \( \mathcal{L}^p(S \cap \mathcal{X}_1) > 0. \)

Proof. By Assumption 1, \( A \) is continuous almost everywhere. Lemma C.5 then implies \( p^A(x) = A(x) \) for almost every \( x \in \{x^* \in S : p^A(x^*) = 1\} \). Since \( \mathcal{L}^p(\{x \in S : p^A(x) = 1\}) > 0, \mathcal{L}^p(\{x \in S : p^A(x) = 1, p^A(x) = A(x)\}) > 0 \), and hence \( \mathcal{L}^p(S \cap \mathcal{X}_1) > 0. \) \qed

Step D.1.2. \( S \cap \text{int}(\mathcal{X}_1) \neq \emptyset. \)

Proof. Suppose \( S \cap \text{int}(\mathcal{X}_1) = \emptyset. \) Then, we must have \( S \cap \mathcal{X}_1 \subset \mathcal{X}_1 \setminus \text{int}(\mathcal{X}_1) \). It then follows that \( \mathcal{L}^p(S \cap \mathcal{X}_1) \leq \mathcal{L}^p(\mathcal{X}_1 \setminus \text{int}(\mathcal{X}_1)) = \mathcal{L}^p(\mathcal{X}_1) - \mathcal{L}^p(\text{int}(\mathcal{X}_1)) = 0 \), where the last equality holds by Assumption 1. But this is a contradiction to Step D.1.1. \qed

Step D.1.3. \( p^A(x) = 1 \) for any \( x \in \text{int}(\mathcal{X}_1). \)

Proof. Pick any \( x \in \text{int}(\mathcal{X}_1). \) By the definition of interior, \( B(x, \delta) \subset \mathcal{X}_1 \) for any sufficiently small \( \delta > 0. \) Therefore, \( p^A(x; \delta) = 1 \) for any sufficiently small \( \delta > 0. \) \qed

Step D.1.4. \( E[Y_{i1} - Y_{0i}|X_i \in S] \) is not identified.

Proof. We first introduce some notation. Let \( Q \) be the set of all distributions of \((Y_{i1}, Y_{0i}, X_i, Z_i)\) satisfying Property 1 and Assumptions 1 and 2. Let \( P \) be the set of all distributions of \((Y_i, X_i, Z_i)\). Let \( T : Q \rightarrow P \) be a function such that, for \( Q \in Q \), \( T(Q) \) is the distribution of \((Z_i Y_{i1} + (1 - Z_i) Y_{0i}, X_i, Z_i)\), where the distribution of \((Y_{i1}, Y_{0i}, X_i, Z_i)\) is \( Q \). Let \( Q_0 \) and \( P_0 \) denote the true distributions of \((Y_{i1}, Y_{0i}, X_i, Z_i)\) and \((Y_i, X_i, Z_i)\), respectively. Given \( P_0 \), the identified set of \( E[Y_{i1} - Y_{0i}|X_i \in S] \) is given by \( \{E_Q[Y_{i1} - Y_{0i}|X_i \in S] : P_0 = T(Q), Q \in Q\} \), where \( E_Q[\cdot] \) is the expectation under distribution \( Q \). We show that this set contains two distinct values. In what follows, \( \Pr(\cdot) \) and \( E[\cdot] \) without a subscript denote the probability and expectation under the true distributions \( Q_0 \) and \( P_0 \) as up until now.

Now pick any \( x^* \in S \cap \text{int}(\mathcal{X}_1). \) Since \( S \) and \( \text{int}(\mathcal{X}_1) \) are open, there is a \( \delta > 0 \) such that \( B(x^*, \delta) \subset S \cap \text{int}(\mathcal{X}_1). \) Let \( \epsilon = \frac{\delta}{2} \), and consider a function \( f : \mathcal{X} \rightarrow \mathbb{R} \) such that \( f(x) = E[Y_{0i}|X = x] \) for all \( x \in \mathcal{X} \setminus B(x^*, \epsilon) \) and \( f(x) = E[Y_{0i}|X = x] - 1 \) for all \( x \in B(x^*, \epsilon) \). Below, we show that \( f \) is continuous at any point \( x \in \mathcal{X} \) such that \( p^A(x) \in (0, 1) \) and \( A(x) \in \{0, 1\} \). Pick any \( x \in \mathcal{X} \) such that \( p^A(x) \in (0, 1) \) and \( A(x) \in \{0, 1\} \). Since \( B(x^*, \delta) \subset \text{int}(\mathcal{X}_1) \) and \( \text{int}(\mathcal{X}_1) \subset \{x' \in \mathcal{X} : p^A(x') = 1\} \) by Step D.1.3, \( x \notin B(x^*, \delta) \). Hence, \( B(x, \epsilon) \subset \mathcal{X} \setminus B(x^*, \epsilon) \). By Assumption 2 and the definition of \( f, f \) is continuous at \( x \).
Now take any vector \((Y_{1i}^*, Y_{0i}^*, X_i^*, Z_i^*)\) distributed according to the true distribution \(Q_0\). Let \(Q\) be the distribution of \((Y_{1i}^Q, Y_{0i}^Q, X_i^Q, Z_i^Q)\), where \((Y_{1i}^Q, X_i^Q, Z_i^Q) = (Y_{1i}^*, X_i^*, Z_i^*)\), and

\[
Y_{0i}^Q = \begin{cases} 
Y_{0i}^* & \text{if } X_i^* \in \mathcal{X} \setminus B(x^*, \epsilon) \\
Y_{0i}^* - 1 & \text{if } X_i^* \in B(x^*, \epsilon).
\end{cases}
\]

Note first that \(Q \in \mathcal{Q}\), since \(E_Q[Y_{11i}^Q|X_i^Q = x] = E[Y_{11i}^*|X_i^* = x]\) and \(E_Q[Y_{0i}^Q|X_i^Q = x] = f(x)\), where \(E[Y_{11i}^*|X_i^*]\) and \(f\) are both continuous at any point \(x \in \mathcal{X}\) such that \(p^A(x) \in (0,1)\) and \(A(x) \in \{0,1\}\). Also, \(Z_i^Q = Z_i^* = 1\) if \(X_i^* \in B(x^*, \epsilon)\). It then follows

\[
Y_{1i}^Q = Z_i^Q Y_{1i}^* + (1 - Z_i^Q) Y_{0i}^* = \begin{cases} 
Z_i^* Y_{11i}^* + (1 - Z_i^*) Y_{0i}^* & \text{if } X_i^* \in \mathcal{X} \setminus B(x^*, \epsilon) \\
Z_i^* Y_{11i}^* & \text{if } X_i^* \in B(x^*, \epsilon),
\end{cases}
\]

\[
Y_i^* = Z_i^* Y_{11i}^* + (1 - Z_i^*) Y_{0i}^* = \begin{cases} 
Z_i^* Y_{11i}^* + (1 - Z_i^*) Y_{0i}^* & \text{if } X_i^* \in \mathcal{X} \setminus B(x^*, \epsilon) \\
Z_i^* Y_{11i}^* & \text{if } X_i^* \in B(x^*, \epsilon).
\end{cases}
\]

Thus, \(Y_{1i}^Q = Y_i^*\), and hence \(T(Q) = T(Q_0) = P_0\).

Using \(E_Q[Y_{11i}^Q|X_i^Q = x] = E[Y_{11i}^*|X_i^* = x]\) and \(E_Q[Y_{0i}^Q|X_i^Q = x] = f(x)\), we have

\[
E_Q[Y_{11i}^Q - Y_{0i}^Q|X_i^Q \in S] = E_Q[Y_{11i}^* - Y_{0i}^*|X_i^* \in S]
\]

\[
= E_Q[E_Q[Y_{1i}^Q|X_i^Q|X_i^Q \in S], X_i^Q \notin B(x^*, \epsilon)]Pr_Q(X_i^Q \notin B(x^*, \epsilon)|X_i^Q \in S)
\]

\[
= E_Q[E_Q[Y_{0i}^Q|X_i^Q|X_i^Q \in B(x^*, \epsilon)]Pr_Q(X_i^Q \in B(x^*, \epsilon)|X_i^Q \in S)
\]

\[
= E[Y_{1i}^Q|X_i^* \in S] - E[f(X_i^*)|X_i^* \in S, X_i^* \notin B(x^*, \epsilon)]Pr(X_i^* \notin B(x^*, \epsilon)|X_i^* \in S)
\]

\[
= E[f(X_i^*)|X_i^* \in B(x^*, \epsilon)]Pr(X_i^* \in B(x^*, \epsilon)|X_i^* \in S)
\]

By the definition of support, \(Pr(X_i^* \in B(x^*, \epsilon)) > 0\). Since \(T(Q) = T(Q_0) = P_0\) but \(E_Q[Y_{11i}^Q - Y_{0i}^Q|X_i^Q \in S] \neq E[Y_{11i}^* - Y_{0i}^*|X_i^* \in S]\), \(E[Y_{11i}^Q - Y_{0i}^Q|X_i \in S]\) is not identified.
D.2  Proof of Theorem 1 (Continued from Appendix B)

D.2.1  Proof of Asymptotic Properties of $\hat{\beta}_i$ When Pr$(A(X_i) \in (0, 1)) = 0$

Let $I_i^s = 1\{p^s(X_i; \delta_n) \in (0, 1)\}$, $D_i^s = (1, D_i, p^s(X_i; \delta_n))'$ and $Z_i^s = (1, Z_i, p^s(X_i; \delta_n))'$. $\hat{\beta}$ and $\Sigma^s$ are given by $\hat{\beta} = (\sum_{i=1}^n Z_i^s(D_i^s)'I_i^s)^{-1}\sum_{i=1}^n Z_i^sY_iI_i^s$ and $\Sigma^s = (\sum_{i=1}^n Z_i^s(D_i^s)'I_i^s)^{-1}(\sum_{i=1}^n (\hat{\epsilon}_i^s)^2 Z_i^s(Z_i^s)'I_i^s)(\sum_{i=1}^n D_i^s(Z_i^s)'I_i^s)^{-1}$, where $\hat{\epsilon}_i^s = Y_i - (D_i^s)'\hat{\beta}$. It is sufficient to show that $\hat{\beta} - \beta = o_p(1)$ if $S_n \to \infty$, and that $\sqrt{n\delta_n}(\hat{\beta} - \beta) = o_p(1)$ and $n\delta_n\Sigma^s \overset{p}{\to} S_{D^1}^{-1}V(S_{D^1}')^{-1}$ if Assumption A.2 holds.

Step D.2.1.1. Let $\{V_i\}_{i=1}^\infty$ be i.i.d. random variables. If $E[V_i|X_i]$ and $E[V_i^2|X_i]$ are bounded on $N(\partial \Omega^*, \delta') \cap N(\chi, \delta')$ for some $\delta' > 0$, and $S_n \to \infty$, then

$$\frac{1}{n\delta_n} \sum_{i=1}^n V_ip^s(X_i; \delta_n)'I_i^s - \frac{1}{n\delta_n} \sum_{i=1}^n V_ip^A(X_i; \delta_n)'I_i = o_p(1)$$

for $l = 0, 1, 2, 3, 4$. If, in addition, Assumption A.2 holds, then for $l = 0, 1, 2,$

$$\frac{1}{\sqrt{n\delta_n}} \sum_{i=1}^n V_ip^s(X_i; \delta_n)'I_i^s - \frac{1}{\sqrt{n\delta_n}} \sum_{i=1}^n V_ip^A(X_i; \delta_n)'I_i = o_p(1).$$

Proof. We have

$$\frac{1}{n\delta_n} \sum_{i=1}^n V_ip^s(X_i; \delta_n)'I_i^s - \frac{1}{n\delta_n} \sum_{i=1}^n V_ip^A(X_i; \delta_n)'I_i = \frac{1}{n\delta_n} \sum_{i=1}^n V_ip^s(X_i; \delta_n)'(I_i^s - I_i) + \frac{1}{n\delta_n} \sum_{i=1}^n V_i(p^s(X_i; \delta_n)' - p^A(X_i; \delta_n)')I_i.$$ 

Consider the second term. By the argument used in Step B.1.3 in Appendix B.1,

$$\left| E \left[ \frac{1}{n\delta_n} \sum_{i=1}^n V_i(p^s(X_i; \delta_n)' - p^A(X_i; \delta_n)')I_i \right] \right| \leq \delta_n^{-1} E[|E[V_i|X_i]|E[p^s(X_i; \delta_n)' - p^A(X_i; \delta_n)']|I_i]$$

$$= \int_{\partial \Omega^* \cap N(\chi, \delta)} |E[V_i|X_i = u + \delta_n v\psi_{\Omega^*}(u)]| E[p^s(u + \delta_n v\psi_{\Omega^*}(u); \delta_n)']$$

$$- p^A(u + \delta_n v\psi_{\Omega^*}(u); \delta_n)']| f_X(u + \delta_n v\psi_{\Omega^*}(u)) J_{p-1}^{\psi_{\Omega^*}}(u, \delta_n) dv dh.$$ 

A-7
where the choice of $\bar{\delta}$ is as in the proof of Step B.1.3. By Lemma C.7,

$$
\left| E \left[ \frac{1}{n\delta_n} \sum_{i=1}^{n} V_i (p^s(X_i; \delta_n)^l - p^A(X_i; \delta_n)^l) I_i \right] \right|
\leq \frac{1}{S_n} \int_{-1}^{1} \int_{\partial X^* \cap N(x, \delta)} |E[V_i X_i = u + \delta_n \nu_{X^*}(u)]|
\times f_X(u + \delta_n \nu_{X^*}(u)) J_{p,1}^{\Omega^*} \psi_{X^*}(u, \delta_n v) d\mathcal{H}^{p-1}(u) dv = O(S_n^{-1})
$$

for $l = 0, 1, 2$. Likewise, by Lemma C.7,

$$
\left| E \left[ \frac{1}{n\delta_n} \sum_{i=1}^{n} V_i (p^s(X_i; \delta_n)^3 - p^A(X_i; \delta_n)^3) I_i \right] \right|
\leq \delta_n^{-1} E[|E[V_i X_i]|] E[(p^s(X_i; \delta_n) - p^A(X_i; \delta_n))
\times (p^s(X_i; \delta_n)^2 + p^s(X_i; \delta_n)p^A(X_i; \delta_n) + p^A(X_i; \delta_n)^2)|X_i]|I_i]
\leq 3\delta_n^{-1} E[|E[V_i X_i]|] E[|p^s(X_i; \delta_n) - p^A(X_i; \delta_n)||X_i]|I_i] \leq \left( \frac{1}{S_n\epsilon^2} + \epsilon \right) O(1)
$$

for every $\epsilon > 0$. We can make the right-hand side arbitrarily close to zero by taking sufficiently small $\epsilon > 0$ and sufficiently large $S_n$, which implies that $|E[\frac{1}{n\delta_n} \sum_{i=1}^{n} V_i (p^s(X_i; \delta_n)^3 - p^A(X_i; \delta_n)^3) I_i]| = o(1)$ if $S_n \to \infty$. Likewise, we can show $|E[\frac{1}{n\delta_n} \sum_{i=1}^{n} V_i (p^s(X_i; \delta_n)^4 - p^A(X_i; \delta_n)^4) I_i]| = o(1)$. As for variance, for $l = 0, 1, 2$,

$$
\text{Var} \left( \frac{1}{n\delta_n} \sum_{i=1}^{n} V_i (p^s(X_i; \delta_n)^l - p^A(X_i; \delta_n)^l) I_i \right)
\leq \frac{1}{n\delta_n} \delta_n^{-1} E[V_i^2 (p^s(X_i; \delta_n)^l - p^A(X_i; \delta_n)^l)^2 I_i]
= \frac{1}{n\delta_n} \delta_n^{-1} E[V_i^2 |X_i| E[(p^s(X_i; \delta_n)^l - p^A(X_i; \delta_n)^l)^2 |X_i]|I_i]
\leq \frac{4}{n\delta_n S_n} \delta_n^{-1} E[V_i^2 |X_i]|I_i] = O((n\delta_n S_n)^{-1}),
$$

and for $l = 3, 4$,

$$
\text{Var} \left( \frac{1}{n\delta_n} \sum_{i=1}^{n} V_i (p^s(X_i; \delta_n)^4 - p^A(X_i; \delta_n)^4) I_i \right)
\leq \frac{1}{n\delta_n} \delta_n^{-1} E[V_i^2 I_i] = o(1).
$$
Therefore, \( \frac{1}{n\delta_n} \sum_{i=1}^{n} V_i(p^*(X_i; \delta_n)^l - p^A(X_i; \delta_n)^l)I_i = o_p(1) \) if \( S_n \to \infty \) for \( 0 \leq l \leq 4 \), and \( \frac{1}{\sqrt{n\delta_n}} \sum_{i=1}^{n} V_i(p^*(X_i; \delta_n)^l - p^A(X_i; \delta_n)^l)I_i = o_p(1) \) if \( (n\delta_n)^{-1/2}S_n \to \infty \) for \( l = 0, 1, 2 \).

We next show \( \frac{1}{n\delta_n} \sum_{i=1}^{n} V_i p^*(X_i; \delta_n)^l(I_i^* - I_i) = o_p(1) \) if \( S_n \to \infty \) for \( l \geq 0 \). Note

\[
\left| E \left[ \frac{1}{n\delta_n} \sum_{i=1}^{n} V_i p^*(X_i; \delta_n)^l(I_i^* - I_i) \right] \right| \leq \delta_n^{-1} E[|E[V_i|X_i]|E[|I_i^* - I_i||X_i]|].
\]

Since \( I_i^* - I_i \leq 0 \) with strict inequality only if \( I_i = 1 \),

\[
E[|I_i^* - I_i||X_i|] = -E[I_i^* - I_i|X_i]I_i = (1 - E[I_i^*|X_i])I_i = \Pr(p^*(X_i; \delta_n) \in \{0, 1\}|X_i)I_i.
\]

We then have

\[
\left| E \left[ \frac{1}{n\delta_n} \sum_{i=1}^{n} V_i p^*(X_i; \delta_n)^l(I_i^* - I_i) \right] \right| \leq \delta_n^{-1} E[|E[V_i|X_i]|((1 - p^A(X_i; \delta_n))^S_n + p^A(X_i; \delta_n)^S_n)I_i]
\]

\[
\leq \int_{-1}^{1} \int_{\partial \Theta^* \cap N(X, \tilde{\delta})} |E[V_i|X_i = u + \delta_n v \psi_{\Theta^*}(u)]| \{|1 - p^A(u + \delta_n v \psi_{\Theta^*}(u); \delta_n)|^S_n
\]

\[
+ p^A(u + \delta_n v \psi_{\Theta^*}(u); \delta_n)^S_n \}\cdot f_X(u + \delta_n v \psi_{\Theta^*}(u))J_{p-1} \psi_{\Theta^*}(u, \delta_n v) d\mathcal{H}^{p-1}(u) dv,
\]

where the second inequality follows from Lemma C.7. For every \( (u, v) \in \partial \Theta^* \cap N(X, \tilde{\delta}) \times (-1, 1), \lim_{\delta \to 0} p^A(u + \delta v \psi_{\Theta^*}(u); \delta) = k(v) \in (0, 1) \) by Step B.1.1. Since \( E[V_i|X_i] \), \( f_X \) and \( J_{p-1} \psi_{\Theta^*} \) are bounded, by the Bounded Convergence Theorem, \( |E[\frac{1}{n\delta_n} \sum_{i=1}^{n} V_i p^*(X_i; \delta_n)^l(I_i^* - I_i)]| = o(1) \) if \( S_n \to \infty \). As for variance,

\[
\text{Var} \left( \frac{1}{n\delta_n} \sum_{i=1}^{n} V_i p^*(X_i; \delta_n)^l(I_i^* - I_i) \right) \leq \frac{1}{n\delta_n} \delta_n^{-1} E[|E[V_i^2|X_i]|E[|I_i^* - I_i||X_i]|] = o(1).
\]

Lastly, we show that, for \( l \geq 0 \), \( \frac{1}{\sqrt{n\delta_n}} \sum_{i=1}^{n} V_i p^*(X_i; \delta_n)^l(I_i^* - I_i) = o_p(1) \) if Assump-
tion A.2 holds. Let \( \eta_n = \gamma \frac{\log n}{S_n} \), where \( \gamma > 1/2 \). We have

\[
\left| E \left[ \frac{1}{\sqrt{n\delta_n}} \sum_{i=1}^{n} V_i p^\delta(X_i; \delta_n) I_i^* - I_i \right] \right|
\]

\[
\leq \sqrt{n\delta_n^{-1}} E[|E[V_i|X_i]|((1 - p^A(X_i; \delta_n))^{S_n} + p^A(X_i; \delta_n)^{S_n}) I_i]
\]

\[
= \sqrt{n\delta_n^{-1}} E[|E[V_i|X_i]|((1 - p^A(X_i; \delta_n))^{S_n} + p^A(X_i; \delta_n)^{S_n})
\]

\[
	imes (1\{p^A(X_i; \delta_n) \in (0, \eta_n) \cup (1 - \eta_n, 1)\} + 1\{p^A(X_i; \delta_n) \in (\eta_n, 1 - \eta_n)\})]
\]

\[
\leq \left( \sup_{x \in N(\partial I^*, 2\delta)} |E[V_i|X_i = x]| \right) \left( \sqrt{n\delta_n^{-1}} \Pr(p^A(X_i; \delta_n) \in (0, \eta_n) \cup (1 - \eta_n, 1))
\]

\[
+ 2\sqrt{n\delta_n(1 - \eta_n)^{S_n}\delta_n^{-1}} E[1\{p^A(X_i; \delta_n) \in (\eta_n, 1 - \eta_n)\}] \right).
\]

By Assumption A.2, \( \sqrt{n\delta_n^{-1}} \Pr(p^A(X_i; \delta_n) \in (0, \eta_n) \cup (1 - \eta_n, 1)) = o(1) \). For the second term,

\[
2\sqrt{n\delta_n(1 - \eta_n)^{S_n}\delta_n^{-1}} E[1\{p^A(X_i; \delta_n) \in (\eta_n, 1 - \eta_n)\}] \leq 2\sqrt{n\delta_n(1 - \eta_n)^{S_n}\delta_n^{-1}} E[I_i]
\]

\[
= 2\sqrt{n\delta_n(1 - \eta_n)^{S_n}} O(1).
\]

Using the fact that \( e^t \geq 1 + t \) for every \( t \in \mathbb{R} \), we have

\[
\sqrt{n\delta_n(1 - \eta_n)^{S_n}} \leq \sqrt{n\delta_n(e^{-\eta_n})^{S_n}} = \sqrt{n\delta_n e^{-\gamma \log n}} = \sqrt{n\delta_n n^{-\gamma}} = n^{1/2 - \gamma} \delta_n^{1/2} \rightarrow 0,
\]

since \( \gamma > 1/2 \). As for variance,

\[
\text{Var} \left( \frac{1}{\sqrt{n\delta_n}} \sum_{i=1}^{n} V_i p^\delta(X_i; \delta_n) I_i^* - I_i \right) \leq \delta_n^{-1} E[|E[V_i^2|X_i]|E[I_i^* - I_i||X_i|]
\]

\[
\leq \delta_n^{-1} E[|E[V_i^2|X_i]| \Pr(p^\delta(X_i; \delta_n) \in \{0, 1\}|X_i) I_i] = o(1). \]

\( \square \)
We have

\[
\hat{\beta}^s - \hat{\beta} = \left(\frac{1}{n\delta_n} \sum_{i=1}^{n} Z_i^s (D_i^s)' I_i^s\right)^{-1} \frac{1}{n\delta_n} \sum_{i=1}^{n} Z_i^s Y_i^s - \left(\frac{1}{n\delta_n} \sum_{i=1}^{n} Z_i D_i I_i\right)^{-1} \frac{1}{n\delta_n} \sum_{i=1}^{n} Z_i Y_i I_i
\]

\[
= \left(\frac{1}{n\delta_n} \sum_{i=1}^{n} Z_i^s (D_i^s)' I_i^s\right)^{-1} \left(\frac{1}{n\delta_n} \sum_{i=1}^{n} Z_i^s Y_i^s - \frac{1}{n\delta_n} \sum_{i=1}^{n} Z_i Y_i I_i\right) - \left(\frac{1}{n\delta_n} \sum_{i=1}^{n} Z_i^s (D_i^s)' I_i^s\right)^{-1} \frac{1}{n\delta_n} \sum_{i=1}^{n} Z_i Y_i I_i.
\]

By Step D.2.1.1, \( \hat{\beta}^s - \hat{\beta} = o_p(1) \) if \( S_n \to \infty \), and \( \sqrt{n\delta_n} (\hat{\beta}^s - \hat{\beta}) = o_p(1) \) if Assumption A.2 holds. By proceeding as in Step B.1.7 in Appendix B.1, we have

\[
\frac{1}{n\delta_n} \sum_{i=1}^{n} (\hat{\epsilon}_i^s)^2 Z_i^s (Z_i^s)' I_i^s = \frac{1}{n\delta_n} \sum_{i=1}^{n} (\epsilon_i^s)^2 Z_i^s (Z_i^s)' I_i^s + o_p(1),
\]

where \( \epsilon_i^s = Y_i - (D_i^s)' \beta \). Then, by Step D.2.1.1,

\[
\frac{1}{n\delta_n} \sum_{i=1}^{n} (\hat{\epsilon}_i^s)^2 Z_i^s (Z_i^s)' I_i^s - \frac{1}{n\delta_n} \sum_{i=1}^{n} \epsilon_i^s Z_i Z_i' I_i
\]

\[
= \frac{1}{n\delta_n} \sum_{i=1}^{n} (Y_i - (D_i^s)' \beta)^2 Z_i^s (Z_i^s)' I_i^s - \frac{1}{n\delta_n} \sum_{i=1}^{n} (Y_i - D_i^s \epsilon_i^s)^2 Z_i Z_i' I_i + o_p(1) = o_p(1)
\]

so that \( \frac{1}{n\delta_n} \sum_{i=1}^{n} (\hat{\epsilon}_i^s)^2 Z_i^s (Z_i^s)' I_i^s \overset{p}{\to} V \). Also, \( \frac{1}{n\delta_n} \sum_{i=1}^{n} Z_i^s (D_i^s)' I_i^s \overset{p}{\to} S_D \) by using Step D.2.1.1, implying the conclusion. \( \square \)

**D.2.2 Proof of Asymptotic Properties of \( \hat{\beta}_1 \) When \( \Pr(A(X_i) \in (0, 1)) > 0 \)**

We first show that \( \beta_1 = \frac{E[A(X_i)(1-A(X_i))(D_i(1)-D_i(0))(Y_i(1)-Y_i(0))]}{E[A(X_i)(1-A(X_i))(D_i(1)-D_i(0))]} \). By Lemma C.6,

\[
\lim_{\delta \to 0} E[p^A(X_i; \delta)(1-p^A(X_i; \delta))(D_i(1)-D_i(0))] = E[A(X_i)(1-A(X_i))(D_i(1)-D_i(0))].
\]

When \( \Pr(A(X_i) \in (0, 1)) > 0 \), \( E[A(X_i)(1-A(X_i))(D_i(1)-D_i(0))] = E[p^A(X_i)(1-p^A(X_i))(D_i(1)-D_i(0))] \), since \( p^A(x) = A(x) \) for almost every \( x \in \mathcal{X} \) by Lemma C.4. Note that \( E[p^A(X_i)(1-p^A(X_i))(D_i(1)-D_i(0))] = \int_\mathcal{X} p^A(x)(1-p^A(x))E[D_i(1)-D_i(0)|X_i = x] f_X(x) dx \). Hence, under Assumption 3 (b), \( E[p^A(X_i)(1-p^A(X_i))(D_i(1)-D_i(0))] = \)
\[ D_i(0)) \neq 0. \] Again by Lemma C.6,

\[
\beta_1 = \lim_{\delta \to 0} E[\omega_i(\delta)(Y_i(1) - Y_i(0))] = \frac{E[A(X_i)(1 - A(X_i))(D_i(1) - D_i(0))(Y_i(1) - Y_i(0))]}{E[A(X_i)(1 - A(X_i))(D_i(1) - D_i(0))]}.
\]

Let \( \tilde{D}_i = (1, D_i, A(X_i))', \tilde{Z}_i = (1, Z_i, A(X_i))', \) and \( I_i^A = 1\{A(X_i) \in (0, 1)\}. \) Below we first show that \( \hat{\beta}_1 \xrightarrow{p} \beta_1 \) if Assumptions 1 and 3 hold and \( \delta_n \to 0. \) We then show that \( \hat{\sigma}^{-1}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} N(0, 1) \) if, in addition, Assumption A.1 holds and \( n\delta_n^2 \to 0. \)

**Proof of Consistency.** Note first that, a few lines of algebra gives

\[
\text{det}(E[\tilde{Z}_i\tilde{D}_i'I_i^A]) = \Pr(I_i^A = 1)^2 \text{Var}(A(X_i)|I_i^A = 1) E[D_i(Z_i - A(X_i))I_i^A]
\]

\[
= \Pr(I_i^A = 1)^2 \text{Var}(A(X_i)|I_i^A = 1) E[(Z_iD_i(1) + (1 - Z_i)D_i(0))(Z_i - A(X_i))I_i^A]
\]

\[
= \Pr(I_i^A = 1)^2 \text{Var}(A(X_i)|I_i^A = 1) E[((Z_i - Z_iA(X_i))D_i(1) - (1 - Z_i)A(X_i)D_i(0))I_i^A]
\]

\[
= \Pr(I_i^A = 1)^2 \text{Var}(A(X_i)|I_i^A = 1) E[A(X_i)(1 - A(X_i))(D_i(1) - D_i(0))I_i^A]
\]

\[
= \Pr(I_i^A = 1)^2 \text{Var}(A(X_i)|I_i^A = 1) E[A(X_i)(1 - A(X_i))(D_i(1) - D_i(0))],
\]

where the fourth equality follows from Property 1. Therefore, \( E[\tilde{Z}_i\tilde{D}_i'I_i^A] \) is invertible when \( \text{Var}(A(X_i)|I_i^A = 1) > 0. \) Another few lines of algebra gives

\[
(E[\tilde{Z}_i\tilde{D}_i'I_i^A])^{-1} = \frac{1}{E[A(X_i)(1 - A(X_i))(D_i(1) - D_i(0))]} \begin{bmatrix}
* & * & * \\
0 & 1 & -1 \\
* & * & *
\end{bmatrix}.
\]
Therefore, when \( \text{Var}(A(X_i)|I_i^A) = 1 > 0 \), by Lemma C.6,

\[
\hat{\beta} = \left( \sum_{i=1}^{n} Z_i \hat{D}_i(I_i) \right)^{-1} \sum_{i=1}^{n} Z_i Y_i I_i \xrightarrow{p} (E[\hat{Z}_i \hat{D}_i^t I_i^A])^{-1} E[\hat{Z}_i Y_i I_i^A]
\]

\[
= \frac{E[Z_i Y_i I_i^A] - E[A(X_i) Y_i I_i^A]}{E[A(X_i)(1 - A(X_i))(D_i(1) - D_i(0))]} \]

\[
= \frac{E[A(X_i) Y_i I_i^A] - E[A(X_i)(A(X_i) Y_i + (1 - A(X_i)) Y_0) I_i^A]}{E[A(X_i)(1 - A(X_i))(D_i(1) - D_i(0))]} \]

\[
= \frac{E[A(X_i)(1 - A(X_i)) (Y_i - Y_0) I_i^A]}{E[A(X_i)(1 - A(X_i))(D_i(1) - D_i(0))]} \]

\[
= \frac{E[A(X_i)(1 - A(X_i))(D_i(1) - D_i(0)) (Y_i(1) - Y_i(0))]}{E[A(X_i)(1 - A(X_i))(D_i(1) - D_i(0))]} = \beta_1.
\]

where the third line follows from Property 1, and the second last equality follows from the definitions of \( Y_{1i} \) and \( Y_{0i} \).

\[ \square \]

**Proof of Asymptotic Normality.** The proof proceeds in six steps.

**Step D.2.2.1.** Let \( \tilde{\beta}_n = (E[\hat{Z}_i \hat{D}_i^t I_i])^{-1} E[\hat{Z}_i Y_i I_i] \), and let \( \tilde{\beta}_{1,n} \) denote the second element of \( \tilde{\beta}_n \). Then \( \tilde{\beta}_{1,n} = \beta_1 \) for any choice of \( \delta_n > 0 \).

**Proof.** Note first that, for every \( \delta > 0 \), \( p^A(x; \delta) \in (0, 1) \) for almost every \( x \in \{ x' \in \mathcal{X} : A(x') \in (0, 1) \} \), since by almost everywhere continuity of \( A \), for almost every \( x \in \{ x' \in \mathcal{X} : A(x') \in (0, 1) \} \), there exists an open ball \( B \subset B(x, \delta) \) such that \( A(x') \in (0, 1) \) for every \( x' \in B \). After a few lines of algebra, we have

\[
\text{det}(E[\hat{Z}_i \hat{D}_i^t I_i]) = \text{Pr}(I_i = 1)^2 \text{Var}(A(X_i)|I_i = 1) E[D_i(Z_i - A(X_i))I_i]
\]

\[
= \text{Pr}(I_i = 1)^2 \text{Var}(A(X_i)|I_i = 1) E[A(X_i)(1 - A(X_i))(D_i(1) - D_i(0))I_i]
\]

\[
= \text{Pr}(I_i = 1)^2 \text{Var}(A(X_i)|I_i = 1) E[A(X_i)(1 - A(X_i))(D_i(1) - D_i(0))],
\]

where the last equality holds since \( p^A(x; \delta) \in (0, 1) \) for almost every \( x \in \{ x' \in \mathcal{X} : A(x') \in (0, 1) \} \).
\( A(x') \in (0, 1) \}. \) By the law of total conditional variance,

\[
\text{Var}(A(X_i)|I_i = 1) \\
= E[\text{Var}(A(X_i)|I_i = 1, I_i^A)|I_i = 1] + \text{Var}(E[A(X_i)|I_i = 1, I_i^A]|I_i = 1)
\]

\[
\geq \sum_{t \in \{0, 1\}} \text{Var}(A(X_i)|I_i = 1, I_i^A = t) \Pr(I_i^A = t|I_i = 1)
\]

\[
\geq \text{Var}(A(X_i)|I_i = 1, I_i^A = 1) \Pr(I_i^A = 1|I_i = 1)
\]

\[
= \text{Var}(A(X_i)|I_i = 1, I_i^A = 1) \Pr(I_i^A = 1|I_i = 1) > 0.
\]

Therefore, \( E[\tilde{Z}_i \tilde{D}_i'I_i] \) is invertible. Another few lines of algebra gives

\[
(E[\tilde{Z}_i \tilde{D}_i'I_i])^{-1} = \frac{1}{E[A(X_i)(1 - A(X_i))(D_i(1) - D_i(0))]} \begin{bmatrix} * & * & * \\ 0 & 1 & -1 \\ * & * & * \end{bmatrix}.
\]

It follows that

\[
\tilde{\beta}_{1,n} = \frac{E[Z_i Y_i I_i] - E[A(X_i)Y_i I_i]}{E[A(X_i)(1 - A(X_i))(D_i(1) - D_i(0))]}
\]

\[
= \frac{E[A(X_i)(1 - A(X_i))(D_i(1) - D_i(0))(Y_i(1) - Y_i(0))I_i]}{E[A(X_i)(1 - A(X_i))(D_i(1) - D_i(0))]} = \beta_1. \quad \square
\]

We can write

\[
\sqrt{n}(\hat{\beta} - \tilde{\beta}_n) = \left( \frac{1}{n} \sum_{i=1}^{n} Z_i \tilde{D}_i'I_i \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i Y_i I_i - \left( \frac{1}{n} \sum_{i=1}^{n} \tilde{Z}_i \tilde{D}_i'I_i \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{Z}_i Y_i I_i
\]

\[
\left( A \right)
\]

\[
+ \left( \frac{1}{n} \sum_{i=1}^{n} \tilde{Z}_i \tilde{D}_i'I_i \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{Z}_i Y_i I_i - (E[\tilde{Z}_i \tilde{D}_i'I_i])^{-1} \sqrt{n} E[\tilde{Z}_i Y_i I_i].
\]

\[
\left( B \right)
\]

We first consider \( B \). Let \( \tilde{\epsilon}_{i,n} = Y_i - \tilde{D}_i'\tilde{\beta}_n \) so that \( E[\tilde{Z}_i \tilde{\epsilon}_{i,n} I_i] = E[\tilde{Z}_i (Y_i - \tilde{D}_i'\tilde{\beta}_n) I_i] = \)
$E[\tilde{Z}_n Y_i I_i] - E[\tilde{Z}_n \tilde{D}_i I_i] \tilde{\beta}_n = 0$. Then (B) equals

$$
\frac{1}{n} \sum_{i=1}^{n} \tilde{Z}_n \tilde{D}_i I_i^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{Z}_i (\tilde{D}_i' \tilde{\beta}_n + \tilde{\epsilon}_i I_i) I_i = (E[\tilde{Z}_n \tilde{D}_i I_i])^{-1} \sqrt{n} E[\tilde{Z}_i (\tilde{D}_i' \tilde{\beta}_n + \tilde{\epsilon}_i I_i)] = \frac{1}{n} \sum_{i=1}^{n} \tilde{Z}_i I_i^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{Z}_i I_i I_i^{-1} (E[\tilde{Z}_i \tilde{D}_i I_i])^{-1} \sqrt{n} E[\tilde{Z}_i \tilde{\epsilon}_i I_i] = (E[\tilde{Z}_n \tilde{D}_i I_i])^{-1} \sqrt{n} E[\tilde{Z}_i \tilde{\epsilon}_i I_i].
$$

Step D.2.2.2. Let $\beta = (E[\tilde{Z}_n \tilde{D}_i I_i])^{-1} E[\tilde{Z}_n Y_i I_i]$ and $\tilde{\epsilon}_i = Y_i - \tilde{D}_i \beta$. Then

$$
\frac{1}{n} \sum_{i=1}^{n} \tilde{Z}_i \tilde{\epsilon}_i I_i \overset{d}{\to} N(0, E[\tilde{Z}_n \tilde{D}_i I_i])
$$

Proof: We use the triangular-array Lyapunov CLT and the Cramér-Wold device. Pick a nonzero $\lambda \in \mathbb{R}^p$, and let $V_i = \frac{1}{\sqrt{n}} \lambda' \tilde{Z}_i \tilde{\epsilon}_i I_i$. First, by Lemma C.6, $\tilde{\beta}_n \to (E[\tilde{Z}_n \tilde{D}_i I_i])^{-1} E[\tilde{Z}_n Y_i I_i] = \beta$ as $n \to \infty$. We have

$$
E[\tilde{Z}_n \tilde{\epsilon}_i I_i] = E[(Y_i - \tilde{D}_i' \tilde{\beta}_n)^2 \tilde{Z}_i \tilde{Z}_i' I_i] = E[(\tilde{\epsilon}_i - \tilde{D}_i' (\tilde{\beta}_n - \beta))^2 \tilde{Z}_i \tilde{Z}_i' I_i]
$$

$$
= E[\tilde{Z}_i \tilde{\epsilon}_i I_i] - 2E[\tilde{\epsilon}_i ((\tilde{\beta}_0, n - \beta_0) + D_i (\tilde{\beta}_1, n - \beta_1) + A(X_i) (\tilde{\beta}_2, n - \beta_2)) \tilde{Z}_i \tilde{Z}_i' I_i]
$$

$$
+ E[((\tilde{\beta}_0, n - \beta_0) + D_i (\tilde{\beta}_1, n - \beta_1) + A(X_i) (\tilde{\beta}_2, n - \beta_2))^2 \tilde{Z}_i \tilde{Z}_i' I_i] \to E[\tilde{Z}_n \tilde{D}_i I_i]
$$

where the convergence follows from Lemma C.6 and $\tilde{\beta}_n \to \beta$. Therefore,

$$
\sum_{i=1}^{n} E[V_i^2] = \lambda E[\tilde{Z}_n \tilde{D}_i I_i] \lambda \to \lambda E[\tilde{Z}_n \tilde{D}_i I_i]
$$

We now verify the Lyapunov condition: for some $t > 0$, $\sum_{i=1}^{n} E[|V_i|^t] \to 0$. Consider

$$
E[|V_i|^t] = \frac{1}{n} E[|\lambda' \tilde{Z}_i \tilde{\epsilon}_i I_i|^t]
$$

We use the $c_r$-inequality: $E[|X + Y|^r] \leq 2^{r-1} E[|X|^r + |Y|^r]$ for $r \geq 1$. Repeating using the $c_r$-inequality gives

$$
E[|\lambda' \tilde{Z}_i \tilde{\epsilon}_i I_i|^t] \leq 2^{2c} [E[|\lambda' \tilde{Z}_i|^t] E[|Y_i|^t + |\tilde{\beta}_0, n|^t + |\tilde{\beta}_1, n|^t |D_i + |\tilde{\beta}_2, n|^t A(X_i)^t I_i]}
$$

$$
\leq 2^{3c} (|\lambda| + |\lambda_2|^4 |E[Y_i|^4] + |\tilde{\beta}_0, n|^4 + |\tilde{\beta}_1, n|^4 |D_i + |\tilde{\beta}_2, n|^4 A(X_i)^4 I_i]
$$

for some constant $c$, where the last equality is by Assumption 3 (a). Thus, $\sum_{i=1}^{n} E[|V_i|^t] \to 0$, and the conclusion follows from the Lyapunov CLT and the Cramér-Wold device.

A-15
We next consider (A). We can write
\[
(A) = \left( \frac{1}{n} \sum_{i=1}^{n} Z_i D_i'I_i \right) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Z_i Y_i I_i - \tilde{Z}_i Y_i I_i)
- \left( \frac{1}{n} \sum_{i=1}^{n} Z_i D_i'I_i \right) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Z_i D_i'I_i - \tilde{Z}_i D_i'I_i)\left( \frac{1}{n} \sum_{i=1}^{n} \tilde{Z}_i D_i'I_i \right) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{Z}_i Y_i I_i.
\]

**Step D.2.2.3.** Let \( I_i(\delta) = 1\{p^A(X_i; \delta) \in (0, 1)\} \). Let \( \{V_i\}_{i=1}^{\infty} \) be i.i.d. random variables such that \( E[|V_i|] < \infty \) and that \( E[V_i|X_i] \) is bounded on \( N(D^*, \delta') \cap \mathcal{X} \) for some \( \delta' > 0 \). Then, for \( l = 0, 1 \),
\[
E[V_i p^A(X_i; \delta)'(p^A(X_i; \delta) - A(X_i))I_i(\delta)] = O(\delta).
\]

**Proof.** For every \( x \notin N(D^*, \delta) \), \( B(x, \delta) \cap D^* = \emptyset \), so \( A \) is continuously differentiable on \( B(x, \delta) \). By the mean value theorem, for every \( x \notin N(D^*, \delta) \) and \( a \in B(0, \delta) \),
\[
A(x + a) = A(x) + \nabla A(y(x, a))'a
\]
for some point \( y(x, a) \) on the line segment connecting \( x \) and \( x + a \). For every \( x \notin N(D^*, \delta) \),
\[
p^A(x; \delta) = \frac{\delta^p \int_{B(0, 1)} A(x + \delta u) du}{\delta^p \int_{B(0, 1)} du} = A(x) + \delta \frac{\int_{B(0, 1)} \nabla A(y(x, \delta u))'u du}{\int_{B(0, 1)} du}.
\]

It follows that
\[
|E[V_i p^A(X_i; \delta)'(p^A(X_i; \delta) - A(X_i))I_i(\delta)1\{X_i \notin N(D^*, \delta)\}]|
\leq \delta E \left[ |V_i| p^A(X_i; \delta)' \sum_{k=1}^{p} \frac{\partial A(y(x, \delta u))}{\partial x_k} |u_k| du \right] I_i(\delta) \times 1\{X_i \notin N(D^*, \delta)\}
\leq \delta E \left[ |V_i| p^A(X_i; \delta)' \sup_{x \in C^*} \left| \frac{\partial A(x)}{\partial x_k} \right| \frac{\int_{B(0, 1)} |u_k| du}{\int_{B(0, 1)} du} \right] = O(\delta),
\]

\[A-16\]
where we use the assumption that the partial derivatives of $A$ is bounded on $C^*$. Furthermore, for sufficiently small $\delta > 0$,

$$
|E[V_i p^A(X_i; \delta)^I(p^A(X_i; \delta) - A(X_i))I_i(\delta)]1\{X_i \in N(D^*, \delta)\}]| \\
\leq E[|E[V_i X_i]|1\{X_i \in N(D^*, \delta)\}] \\
\leq CE[1\{X_i \in N(D^*, \delta)\}] = C \Pr(X_i \in N(D^*, \delta)) = O(\delta),
$$

where $C$ is a constant. The second inequality follows from the assumption that $E[V_i |X_i]$ is bounded on $N(D^*, \delta') \cap X$ for some $\delta' > 0$. The last equality is by Assumption A.1 (a). Putting these together, we have

$$
E[V_ip^A(X_i; \delta)^I(p^A(X_i; \delta) - A(X_i))I_i(\delta)] \\
= E[V_ip^A(X_i; \delta)^I(p^A(X_i; \delta) - A(X_i))I_i(\delta)1\{X_i \notin N(D^*, \delta)\} + 1\{X_i \in N(D^*, \delta)\}] \\
= O(\delta).
$$

**Step D.2.2.4.** $\frac{1}{\sqrt{n}} \sum_{i=1}^n \left(Z_i Y_i I_i - \tilde{Z}_i Y_i I_i\right) = o_p(1)$ and $\frac{1}{\sqrt{n}} \sum_{i=1}^n \left(Z_i D_i' I_i - \tilde{Z}_i \tilde{D}_i' I_i\right) = o_p(1)$.

**Proof.** We only show that $\frac{1}{\sqrt{n}} \sum_{i=1}^n (p^A(X_i; \delta_n)^2 - A(X_i)^2)I_i = o_p(1)$. The proofs for the other elements are similar. As for bias,

$$
E \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^n (p^A(X_i; \delta_n)^2 - A(X_i)^2)I_i \right] = \sqrt{n}E[(p^A(X_i; \delta_n)^2 - A(X_i)^2)I_i] \\
= \sqrt{n}E[(p^A(X_i; \delta_n) + A(X_i))(p^A(X_i; \delta_n) - A(X_i))I_i] = \sqrt{n}O(\delta_n) = 0,
$$

where the third equality follows from Step D.2.2.3 and the last from the assumption that $n\delta_n^2 \rightarrow 0$. As for variance, by Lemma C.6,

$$
\text{Var} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n (p^A(X_i; \delta_n)^2 - A(X_i)^2)I_i \right) \leq E[(p^A(X_i; \delta_n)^2 - A(X_i)^2)^2I_i] \rightarrow 0.
$$

**Step D.2.2.5.** $n \tilde{\Sigma} \xrightarrow{p} (E[\tilde{Z}_i \tilde{D}_i' I_i^A])^{-1} E[\tilde{e}_i^2 \tilde{Z}_i I_i^A](E[\tilde{D}_i \tilde{Z}_i' I_i^A])^{-1}$. 

A-17
Proof. Let $\epsilon_i = Y_i - D_i^T \beta$. We have
\[
\frac{1}{n} \sum_{i=1}^{n} \epsilon_i^2 Z_i Z_i'I_i = \frac{1}{n} \sum_{i=1}^{n} (Y_i - D_i^T \hat{\beta})^2 Z_i Z_i'I_i = \frac{1}{n} \sum_{i=1}^{n} (\epsilon_i - D_i^T (\hat{\beta} - \beta))^2 Z_i Z_i'I_i
\]
\[
= \frac{1}{n} \sum_{i=1}^{n} \epsilon_i^2 Z_i Z_i'I_i + \frac{1}{n} \sum_{i=1}^{n} ((\hat{\beta}_0 - \beta_0) + D_i(\hat{\beta}_1 - \beta_1) + p_i A(X_i; \delta_n)(\hat{\beta}_2 - \beta_2))^2 Z_i Z_i'I_i
\]
\[
- \frac{2}{n} \sum_{i=1}^{n} (Y_i - D_i^T \hat{\beta})((\hat{\beta}_0 - \beta_0) + D_i(\hat{\beta}_1 - \beta_1) + p_i A(X_i; \delta_n)(\hat{\beta}_2 - \beta_2)) Z_i Z_i'I_i
\]
\[
= \frac{1}{n} \sum_{i=1}^{n} \epsilon_i^2 Z_i Z_i'I_i + o_p(1) O_p(1),
\]
where the last equality follows from the result that $\hat{\beta} - \beta = o_p(1)$ and from Lemma C.6. The conclusion is implied by the following consequence of Lemma C.6:
\[
\frac{1}{n} \sum_{i=1}^{n} \epsilon_i^2 Z_i Z_i'I_i = \frac{1}{n} \sum_{i=1}^{n} (Y_i^2 - 2Y_i D_i^T \beta + \beta' D_i D_i^T(\beta) Z_i Z_i'I_i)
\]
\[
\xrightarrow{p} E[(Y_i^2 - 2Y_i D_i^T \beta + \beta' D_i D_i^T(\beta) Z_i Z_i'I_i)] = E[\epsilon_i^2 Z_i Z_i'I_i],
\]
\[
\frac{1}{n} \sum_{i=1}^{n} Z_i D_i I_i \xrightarrow{p} E[Z_i \hat{D}_i I_i^A]
\]
\[
\square
\]

Step D.2.2.6. $(\hat{\sigma})^{-1}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} N(0,1)$.

Proof. By combining the results from Steps D.2.2.2–D.2.4.4 and by Lemma C.6,
\[
(A) \xrightarrow{p} 0, \quad (B) \xrightarrow{d} N(0, (E[Z_i \hat{D}_i I_i^A])^{-1} E[\epsilon_i^2 Z_i Z_i'I_i](E[Z_i \hat{D}_i I_i^A])^{-1}),
\]
and therefore, $\sqrt{n}(\hat{\beta} - \beta_n) \xrightarrow{d} N(0, (E[Z_i \hat{D}_i I_i^A])^{-1} E[\epsilon_i^2 Z_i Z_i'I_i](E[Z_i \hat{D}_i I_i^A])^{-1})$. The conclusion then follows from Steps D.2.2.1 and D.2.2.5.

D.2.3 Proof of Asymptotic Properties of $\hat{\beta}_i$ When $\Pr(A(X_i) \in (0,1)) > 0$

Let $I_i = 1\{p_i(X_i; \delta_n) \in (0,1)\}$, $D_i = (1, D_i, p_i(X_i; \delta_n))'$, and $Z_i = (1, Z_i, p_i(X_i; \delta_n))'$. $\hat{\beta}$ and $\Sigma$ are given by $\hat{\beta} = (\sum_{i=1}^{n} Z_i^T (D_i')I_i^s)^{-1} \sum_{i=1}^{n} Z_i Y_i I_i^s$ and $\Sigma = (\sum_{i=1}^{n} Z_i^T (D_i')I_i^s)^{-1} (\sum_{i=1}^{n} (\epsilon_i^2 T_i Z_i^T (Z_i')I_i^s)^{-1})(\sum_{i=1}^{n} D_i^T (Z_i')I_i^s)^{-1}$, where $\epsilon_i = Y_i - (D_i^T \hat{\beta})$. 

A-18
We show $\hat{\beta}_1^s \overset{p}{\to} \beta_1$ if $S_n \to \infty$ and $(\hat{\alpha}^s)^{-1}(\hat{\beta}_1^s - \beta_1) \overset{d}{\to} \mathcal{N}(0, 1)$ if Assumptions A.1 and A.2 hold and $n\delta_n^2 \to 0$. It suffices to show that $\hat{\beta}_1^s - \hat{\beta} = o_p(1)$ if $S_n \to \infty$ and that $\sqrt{n}(\hat{\beta}_1^s - \hat{\beta}) = o_p(1)$ and $n\Sigma^s \overset{p}{\to} (E[\tilde{Z}_i \tilde{D}_i' I_i^A])^{-1} E[\tilde{c}_i^2 \tilde{Z}_i \tilde{I}_i I_i^A](E[\tilde{D}_i \tilde{Z}_i I_i^A])^{-1}$ if Assumptions A.1 and A.2 hold and $n\delta_n^2 \to 0$. We have
\[
\hat{\beta}_1^s - \hat{\beta} = (\frac{1}{n} \sum_{i=1}^{n} Z_i^s(D_i^s)'I_i^s)^{-1} \frac{1}{n} \sum_{i=1}^{n} Z_i^s Y_i I_i^s - \frac{1}{n} \sum_{i=1}^{n} Z_i^s D_i^s I_i - \frac{1}{n} \sum_{i=1}^{n} Z_i Y_i I_i
\]
\[
= (\frac{1}{n} \sum_{i=1}^{n} Z_i^s(D_i^s)'I_i^s)^{-1}(\frac{1}{n} \sum_{i=1}^{n} Z_i Y_i I_i^s - \frac{1}{n} \sum_{i=1}^{n} Z_i Y_i I_i)
\]
\[
- (\frac{1}{n} \sum_{i=1}^{n} Z_i^s(D_i^s)'I_i^s)^{-1}(\frac{1}{n} \sum_{i=1}^{n} Z_i Y_i I_i^s - \frac{1}{n} \sum_{i=1}^{n} Z_i Y_i I_i + (\frac{1}{n} \sum_{i=1}^{n} Z_i^s(D_i^s)'I_i^s)^{-1} \frac{1}{n} \sum_{i=1}^{n} Z_i Y_i I_i)
\]

By Lemma C.8, $\hat{\beta}_1^s - \hat{\beta} = o_p(1)$ if $S_n \to \infty$, and $\sqrt{n}(\hat{\beta}_1^s - \hat{\beta}) = o_p(1)$ if Assumptions A.1 and A.2 hold and $n\delta_n^2 \to 0$. By proceeding as in Step D.2.2.5 in Appendix D.2.2,
\[
\frac{1}{n} \sum_{i=1}^{n} (\hat{c}_i^s)^2 Z_i^s(Z_i^s)'I_i^s = \frac{1}{n} \sum_{i=1}^{n} (\hat{c}_i^s)^2 Z_i^s(Z_i^s)'I_i^s + o_p(1),
\]
where $\epsilon_i^s = Y_i - (D_i^s)'\beta$. Then, by Lemma C.8,
\[
\frac{1}{n} \sum_{i=1}^{n} (\hat{c}_i^s)^2 Z_i^s(Z_i^s)'I_i^s = \frac{1}{n} \sum_{i=1}^{n} \epsilon_i^s Z_i I_i
\]
\[
= \frac{1}{n} \sum_{i=1}^{n} (Y_i^2 - 2Y_i(D_i^s)'\beta + \beta' D_i^s(D_i^s)'\beta)Z_i^s(Z_i^s)'I_i^s
\]
\[
- \frac{1}{n} \sum_{i=1}^{n} (Y_i^2 - 2Y_i D_i^s \beta + \beta' D_i I_i^A)Z_i I_i^A + o_p(1) = o_p(1)
\]
so that $\frac{1}{n} \sum_{i=1}^{n} (\hat{c}_i^s)^2 Z_i^s(Z_i^s)'I_i^s \overset{p}{\to} E[\hat{c}_i^2 \tilde{Z}_i \tilde{I}_i I_i^A]$. Also, $\frac{1}{n} \sum_{i=1}^{n} Z_i^s(D_i^s)'I_i^s \overset{p}{\to} E[\tilde{Z}_i \tilde{D}_i^s I_i^A]$ by using Lemma C.8, implying the conclusion. \qed
E  Proofs of Lemmas

E.1  Proof of Lemma C.6

We show that $E[V_ip^A(X_i;\delta) \mathbb{1}\{p^A(X_i;\delta) \in (0,1)\}^m] \to E[V_iA(X_i)\mathbb{1}\{A(X_i) \in (0,1)\}^m]$ for $l \geq 0$ and $m = 0, 1$ as $\delta \to 0$, and that $\text{Var}(\frac{1}{n} \sum_{i=1}^n V_ip^A(X_i;\delta)I_{i,n}) \to 0$ for $l \geq 0$ as $n \to \infty$. We first prove the first part. Suppose $A$ is continuous at $x$ and $A(x) \in (0,1)$. Then $\lim_{\delta \to 0} p^A(x;\delta) = A(x)$ by Lemma C.5, and hence $p^A(x;\delta) \in (0,1)$ for sufficiently small $\delta > 0$. It follows that $1\{p^A(x;\delta) \in (0,1)\} \to 1 = 1\{A(x) \in (0,1)\}$ as $\delta \to 0$. Suppose $x \in \text{int}(\mathcal{X}_0) \cup \text{int}(\mathcal{X}_1)$. Then $B(x,\delta) \subset \mathcal{X}_0$ or $B(x,\delta) \subset \mathcal{X}_1$ for sufficiently small $\delta > 0$ by the fact that $\text{int}(\mathcal{X}_0)$ and $\text{int}(\mathcal{X}_1)$ are open, and hence $1\{p^A(x;\delta) \in (0,1)\} \to 0 = 1\{A(x) \in (0,1)\}$ as $\delta \to 0$. Therefore, $\lim_{\delta \to 0} p^A(x;\delta) = A(x)$ and $\lim_{\delta \to 0} 1\{p^A(x;\delta) \in (0,1)\} = 1\{A(x) \in (0,1)\}$ for almost every $x \in \mathcal{X}$, since $A$ is continuous at $x$ for almost every $x \in \mathcal{X}$ by Assumption 1 (a), and either $A(x) \in (0,1)$ or $x \in \text{int}(\mathcal{X}_0) \cup \text{int}(\mathcal{X}_1)$ for almost every $x \in \mathcal{X}$ by Assumption 1 (b). By the Dominated Convergence Theorem,

$$E[V_ip^A(X_i;\delta) \mathbb{1}\{p^A(X_i;\delta) \in (0,1)\}^m] \to E[V_iA(X_i)\mathbb{1}\{A(X_i) \in (0,1)\}^m]$$

as $\delta \to 0$. As for variance, as $n \to \infty$,

$$\text{Var}\left(\frac{1}{n} \sum_{i=1}^n V_ip^A(X_i;\delta)I_{i,n}\right) \leq \frac{1}{n} E[V_i^2p^A(X_i;\delta)I_{i,n}^2] \leq \frac{1}{n} E[V_i^2] \to 0.$$  \hfill $\Box$

E.2  Proof of Lemma C.7

By construction, $E[A(X_s^*)] = p^A(x;\delta_n)$, so

$$E[p^s(x;\delta_n) - p^A(x;\delta_n)] = E\left[\frac{1}{S_n} \sum_{s=1}^{S_n} A(X_s^*)\right] - p^A(x;\delta_n) = 0,$$

$$E[(p^s(x;\delta_n) - p^A(x;\delta_n))^2] = E[p^s(x;\delta_n)^2 - p^A(x;\delta_n)^2]$$

$$= \text{Var}(p^s(x;\delta_n)) = \text{Var}\left(\frac{1}{S_n} \sum_{s=1}^{S_n} A(X_s^*)\right) = \frac{1}{S_n} \text{Var}(A(X_s^*)) \leq \frac{1}{S_n} E[A(X_s^*)^2] \leq \frac{1}{S_n},$$
We have the following bounds on $\Pr(A(X_s^*) = 0)$ and $\Pr(A(X_s^*) = 1)$:

$$0 \leq \Pr(A(X_s^*) = 0) \leq 1 - p^A(x; \delta_n), \quad 0 \leq \Pr(A(X_s^*) = 1) \leq p^A(x; \delta_n).$$

It follows that

$$\Pr(p^s(x; \delta_n) \in \{0, 1\}) = \Pr(A(X_s^*) = 0)^{S_n} + \Pr(A(X_s^*) = 1)^{S_n} \leq (1 - p^A(x; \delta_n))^{S_n} + p^A(x; \delta_n)^{S_n}.$$

Lastly, for any $\epsilon > 0$,

$$E[|p^s(x; \delta_n) - p^A(x; \delta_n)|]$$

$$= E[|p^s(x; \delta_n) - p^A(x; \delta_n)||p^s(x; \delta_n) - p^A(x; \delta_n)| \geq \epsilon \Pr(|p^s(x; \delta_n) - p^A(x; \delta_n)| \geq \epsilon)]$$

$$+ E[|p^s(x; \delta_n) - p^A(x; \delta_n)||p^s(x; \delta_n) - p^A(x; \delta_n)| < \epsilon \Pr(|p^s(x; \delta_n) - p^A(x; \delta_n)| < \epsilon)] < 1 \cdot \frac{\text{Var}(p^s(x; \delta_n))}{\epsilon^2} + \epsilon \leq \frac{1}{S_n \epsilon^2} + \epsilon,$$

where the first inequality is by Chebyshev's inequality. We can make $E[|p^s(x; \delta_n) - p^A(x; \delta_n)|]$ arbitrarily close to zero by taking sufficiently small $\epsilon > 0$ and sufficiently large $S_n$, which implies that $E[|p^s(x; \delta_n) - p^A(x; \delta_n)|] = o(1)$ if $S_n \to \infty$. \hfill \Box

### E.3 Proof of Lemma C.8

We have

$$\frac{1}{n} \sum_{i=1}^{n} V_i p^s(X_i; \delta_n) I_{i,n} - \frac{1}{n} \sum_{i=1}^{n} V_i p^A(X_i; \delta_n) I_{i,n}$$

$$= \frac{1}{n} \sum_{i=1}^{n} V_i p^s(X_i; \delta_n) (I_{i,n} - I_{i,n}) + \frac{1}{n} \sum_{i=1}^{n} V_i (p^s(X_i; \delta_n) - p^A(X_i; \delta_n)) I_{i,n}.$$
We first show that \( \frac{1}{n} \sum_{i=1}^{n} V_i p^s(X_i; \delta_n)^l (I_{i,n}^s - I_{i,n}) = o_p(1) \) if \( \delta_n \to 0 \) for \( l \geq 0 \). We have

\[
\begin{align*}
E \left[ \frac{1}{n} \sum_{i=1}^{n} V_i p^s(X_i; \delta_n)^l (I_{i,n}^s - I_{i,n}) \right] &= |E[V_i p^s(X_i; \delta_n)^l (I_{i,n}^s - I_{i,n})]| \\
&\leq E[|E[V_i|X_i]| |E[p^s(X_i; \delta_n)^l (I_{i,n}^s - I_{i,n})|X_i]|] \\
&\leq E[|E[V_i|X_i]|E[|I_{i,n}^s - I_{i,n}|X_i]] \\
\end{align*}
\]

Note that by construction, \( 1 \{p^s(X_i; \delta_n) \in (0,1)\} \leq 1 \{p^A(X_i; \delta_n) \in (0,1)\} \) with probability one conditional on \( X_i = x \), so that \( E[I_{i,n}^s - I_{i,n}|X_i = x] = -E[I_{i,n}^s - I_{i,n}|X_i = x] \). Suppose \( A \) is continuous at \( x \) and \( A(x) \in (0,1) \). Then \( A(x^*) \in (0,1) \) for all \( x^* \in B(x, \delta_n) \) and hence \( p^A(x; \delta_n) \in (0,1) \) and \( p^s(x; \delta_n) \in (0,1) \) for sufficiently small \( \delta_n > 0 \), so that \( E[I_{i,n}^s - I_{i,n}|X_i = x] \to 0 \) as \( n \to \infty \). Suppose \( x \in \text{int}(X_0) \cup \text{int}(X_1) \). Then \( B(x, \delta_n) \subset X_0 \) or \( B(x, \delta_n) \subset X_1 \) for sufficiently small \( \delta_n > 0 \) by the fact that \( \text{int}(X_0) \) and \( \text{int}(X_1) \) are open, and hence \( p^A(x; \delta_n) \in (0,1) \) and \( p^s(x; \delta_n) \in (0,1) \) for sufficiently small \( \delta_n > 0 \), so that \( E[I_{i,n}^s - I_{i,n}|X_i = x] \to 0 \) as \( n \to \infty \). Therefore, \( E[I_{i,n}^s - I_{i,n}|X_i = x] \to 0 \) for almost every \( x \in X \), since \( A \) is continuous at \( x \) for almost every \( x \in X \) by Assumption 1 (a), and either \( A(x) \in (0,1) \) or \( x \in \text{int}(X_0) \cup \text{int}(X_1) \) for almost every \( x \in X \) by Assumption 1 (b). By the Dominated Convergence Theorem, 

\[-E[|E[V_i|X_i]|E[I_{i,n}^s - I_{i,n}|X_i]|] \to 0 \text{ as } n \to \infty. \]

Also,

\[
\text{Var} \left( \frac{1}{n} \sum_{i=1}^{n} V_i p^s(X_i; \delta_n)^l (I_{i,n}^s - I_{i,n}) \right) \leq \frac{1}{n} E[V_i^2 p^s(X_i; \delta_n)^{2l}(I_{i,n}^s - I_{i,n})^2] \leq \frac{1}{n} E[V_i^2],
\]

which converges to zero as \( n \to \infty \).

Next, we show that, for \( l \geq 0 \), \( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i p^s(X_i; \delta_n)^l (I_{i,n}^s - I_{i,n}) = o_p(1) \) if Assumption A.2 holds and \( E[V_i|X_i] \) is bounded. Let \( \eta_n = \gamma \frac{\log n}{S_n}, \) where \( \gamma > 1/2 \). We have

\[
\begin{align*}
E \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i p^s(X_i; \delta_n)^l (I_{i,n}^s - I_{i,n}) \right] &= \sqrt{n} E[|E[V_i|X_i]|E[|I_{i,n}^s - I_{i,n}|X_i]] \\
&= \sqrt{n} E[|E[V_i|X_i]|E[|I_{i,n}^s - 1|X_i]I_{i,n}] \\
&= \sqrt{n} E[|E[V_i|X_i]|E[|I_{i,n}^s - 1|X_i]I_{i,n}] \\
&\leq \sqrt{n} E[|E[V_i|X_i]|(|1 - p^A(X_i; \delta_n)| - p^A(X_i; \delta_n))S_n + p^A(X_i; \delta_n)S_n]I_{i,n}],
\end{align*}
\]

where the first equality follows from the fact that \( I_{i,n}^s \leq I_{i,n} \) with strict inequality only.
if \( I_{i,n} = 1 \), and the second inequality follows from Lemma C.7. The right-hand side equals
\[
\sqrt{n}E[|E[V_i|X_i]|((1 - p^A(X_i; \delta_n))^2 + p^A(X_i; \delta_n)^2) \\
\times (1\{p^A(X_i; \delta_n) \in (0, \eta_n) \cup (1 - \eta_n, 1)\} + 1\{p^A(X_i; \delta_n) \in [\eta_n, 1 - \eta_n]\})] \\
\leq (\sup_{x \in X} |E[V_i|X_i = x]|)(\sqrt{n}Pr(p^A(X_i; \delta_n) \in (0, \eta_n) \cup (1 - \eta_n, 1)) + 2\sqrt{n}(1 - \eta_n)^{S_n}).
\]

By Assumption A.2, \( \sqrt{n}Pr(p^A(X_i; \delta_n) \in (0, \eta_n) \cup (1 - \eta_n, 1)) = o(1) \). As for \( \sqrt{n}(1 - \eta_n)^{S_n} \), using the fact that \( e^t \geq 1 + t \) for every \( t \in \mathbb{R} \), we have
\[
\sqrt{n}(1 - \eta_n)^{S_n} \leq \sqrt{n}(e^{-\eta_n})^{S_n} = \sqrt{n}e^{-\eta_n} = \sqrt{n}e^{-\gamma \log n} = \sqrt{n}n^{-\gamma} = n^{1/2 - \gamma} \rightarrow 0,
\]
since \( \gamma > 1/2 \). As for variance,
\[
\text{Var}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i p^s(X_i; \delta_n)^l (I_{i,n}^* - I_{i,n})\right) \leq E[V_i^2 p^s(X_i; \delta_n)^2 (I_{i,n}^* - I_{i,n})^2] \\
\leq E[V_i^2 |I_{i,n}^* - I_{i,n}|] = E[E[V_i^2|X_i]E[I_{i,n}^* - I_{i,n}|X_i]] = o(1).
\]

Now, we show that \( \frac{1}{n} \sum_{i=1}^{n} V_i (p^s(X_i; \delta_n)^l - p^A(X_i; \delta_n)^l)I_{i,n} = o_p(1) \) if \( S_n \to \infty \) for \( l = 0, 1, 2, 3, 4 \). For \( l = 0, 1, 2, 3, 4 \), we can write
\[
\left| E\left[\frac{1}{n} \sum_{i=1}^{n} V_i (p^s(X_i; \delta_n)^l - p^A(X_i; \delta_n)^l)I_{i,n}\right]\right| \\
\leq CE[E[V_i|X_i]|E[p^s(X_i; \delta_n) - p^A(X_i; \delta_n)|X_i]I_{i,n}]
\]
for some constant \( C > 0 \). By Lemma C.7, \( E[E[V_i|X_i]|E[p^s(X_i; \delta_n) - p^A(X_i; \delta_n)|X_i]I_{i,n}] = o(1) \). As for variance, for \( l \geq 0 \),
\[
\text{Var}\left(\frac{1}{n} \sum_{i=1}^{n} V_i (p^s(X_i; \delta_n)^l - p^A(X_i; \delta_n)^l)I_{i,n}\right) \leq \frac{1}{n} E[V_i^2 (p^s(X_i; \delta_n)^l - p^A(X_i; \delta_n)^l)^2 I_{i,n}],
\]
which converges to zero as \( n \to \infty \).

Lastly, we show that, for \( l = 0, 1, 2 \), \( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i (p^s(X_i; \delta_n)^l - p^A(X_i; \delta_n)^l)I_{i,n} = o_p(1) \) if Assumptions A.1 and A.2 hold, \( n\delta_n^2 \to 0 \), and \( E[V_i|X_i] \) is bounded. Let \( C^* \) and \( D^* \) be defined as in Appendix A. We first obtain a bound on \( E[p^s(x; \delta_n)^2 - p^A(x; \delta_n)^2] \).
that holds for every \( x \notin N(D^*, \delta_n) \) and \( n \). Fix \( x \in N(D^*, \delta_n) \), and let \( X_1^*, \ldots, X_{S_n}^* \) be \( S_n \) independent draws from the uniform distribution on \( B(x, \delta_n) \) so that \( p^*(x; \delta_n) = \frac{1}{S_n} \sum_{s=1}^{S_n} A(X_s^*) \). We have

\[
E[p^*(x; \delta_n)^2 - p^A(x; \delta_n)^2] = \text{Var}(p^*(x; \delta_n)) = \frac{1}{S_n} \text{Var}(A(X_s^*)).
\]

We compute a bound on \( \text{Var}(A(X_s^*)) \). Since \( x \notin N(D^*, \delta_n) \), \( B(x, \delta_n) \cap D^* = \emptyset \), so \( A \) is continuously differentiable on \( B(x, \delta_n) \). By the mean value theorem, for every \( a \in B(0, \delta_n) \),

\[
A(x + a) = A(x) + \nabla A(y(x, a))'a
\]

for some point \( y(x, a) \) on the line segment connecting \( x \) and \( x + a \). Hence,

\[
p^A(x; \delta_n) = \frac{\delta_n^p \int_{B(0,1)} A(x + \delta_n u)du}{\delta_n^p \int_{B(0,1)} du} = A(x) + \delta_n \frac{\int_{B(0,1)} \nabla A(y(x, \delta_n u))'u du}{\int_{B(0,1)} du}.
\]

For every \( v \in B(0,1) \),

\[
|A(x + \delta_n v) - p^A(x; \delta_n)| = \delta_n \left| \nabla A(y(x, \delta_n v))'v - \frac{\int_{B(0,1)} \nabla A(y(x, \delta_n u))'u du}{\int_{B(0,1)} du} \right|
\]

\[
\leq \delta_n \left( \sum_{k=1}^p \left| \frac{\partial A(y(x, \delta_n v))}{\partial x_k} \right| |v_k| + \frac{\int_{B(0,1)} \sum_{k=1}^p \left| \frac{\partial A(y(x, \delta_n u))}{\partial x_k} \right| |u_k| du}{\int_{B(0,1)} du} \right)
\]

\[
\leq \delta_n \left( \sum_{k=1}^p \sup_{x^* \in C^*} \left| \frac{\partial A(x^*)}{\partial x_k} \right| + \sum_{k=1}^p \sup_{x^* \in C^*} \left| \frac{\partial A(x^*)}{\partial x_k} \right| \frac{\int_{B(0,1)} |u_k| du}{\int_{B(0,1)} du} \right) = C\delta_n,
\]

where \( C = \sum_{k=1}^p \sup_{x^* \in C^*} \left| \frac{\partial A(x^*)}{\partial x_k} \right| (1 + \frac{\int_{B(0,1)} |u_k| du}{\int_{B(0,1)} du}) \) is finite under Assumption A.1.

Note that \( C \) is independent of \( x \) and \( n \). It follows that

\[
\text{Var}(A(X_s^*)) = E[(A(X_s^*) - p^A(x; \delta_n))^2] = \frac{\delta_n^p \int_{B(0,1)} (A(x + \delta_n u) - p^A(x; \delta_n))^2 du}{\delta_n^p \int_{B(0,1)} du}
\]

\[
\leq C^2 \delta_n^2,
\]
so that \( E[p^*(x; \delta_n)^2 - p^A(x; \delta_n)^2] \leq C^2 \frac{\delta_n^2}{S_n} \). Now, for \( l = 0, 1, 2 \),

\[
E \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i(p^*(X_i; \delta_n)^l - p^A(X_i; \delta_n)^l)I_{i,n} \right] = \sqrt{n}E[V_i(p^*(X_i; \delta_n)^l - p^A(X_i; \delta_n)^l)I_{i,n}]
\]

\[
= \sqrt{n}E[V_i(p^*(X_i; \delta_n)^l - p^A(X_i; \delta_n)^l)1\{X_i \notin N(D^*, \delta_n)\}]
+ \sqrt{n}E[V_i(p^*(X_i; \delta_n)^l - p^A(X_i; \delta_n)^l)1\{X_i \in N(D^*, \delta_n)\}]
\leq \sqrt{n}E[|E[V_i|X_i]|E[(p^*(X_i; \delta_n)^l - p^A(X_i; \delta_n)^l)|X_i]|1\{X_i \notin N(D^*, \delta_n)\}]
+ \sqrt{n}E[|E[V_i|X_i]|1\{X_i \in N(D^*, \delta_n)\}]
\leq \sqrt{n}(\sup_{x \in X} |E[V_i|X_i = x]|) \left( C^2 \frac{\delta_n^2}{S_n} + \Pr(X_i \in N(D^*, \delta_n)) \right) = O \left( \frac{\sqrt{n} \delta_n^2}{S_n} \right) + O(\sqrt{n} \delta_n),
\]

where the last equality follows from Assumption A.1 (a). Since \( (n\delta_n)^{-1/2}S_n \to \infty \) and \( n\delta_n^2 \to 0 \), \( O \left( \frac{\sqrt{n} \delta_n^2}{S_n} \right) + O(\sqrt{n} \delta_n) = o(1) \).

As for variance, by Lemma C.7, for \( l = 0, 1, 2 \),

\[
\text{Var} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i(p^*(X_i; \delta_n)^l - p^A(X_i; \delta_n)^l) \right) \leq E[V_i^2(p^*(X_i; \delta_n)^l - p^A(X_i; \delta_n)^l)^2 I_{i,n}]
\]

\[
\leq E[E[V_i^2|X_i]E[(p^*(X_i; \delta_n)^l - p^A(X_i; \delta_n)^l)^2|X_i]|I_{i,n}] \leq \frac{4}{S_n} E[E[V_i^2|X_i]|I_{i,n}] = o(1).
\]

\[\square\]

References

