2-2023

Covariate adjustment in experiments with matched pairs

Yuehao BAI

Liang JIANG

Joseph P. ROMANO

Azeem M. SHAIKH

Yichong ZHANG
Singapore Management University, yczhang@smu.edu.sg

Covariate Adjustment in Experiments with Matched Pairs*

Yuehao Bai
Department of Economics
University of Michigan
yuehaob@umich.edu

Liang Jiang
Fanhai International School of Finance
Fudan University
jiangliang@fudan.edu.cn

Joseph P. Romano
Departments of Economics & Statistics
Stanford University
romano@stanford.edu

Azeem M. Shaikh
Department of Economics
University of Chicago
amshaikh@uchicago.edu

Yichong Zhang
School of Economics
Singapore Management University
yczhang@smu.edu.sg

February 14, 2023

Abstract

This paper studies inference on the average treatment effect in experiments in which treatment status is determined according to "matched pairs" and it is additionally desired to adjust for observed, baseline covariates to gain further precision. By a "matched pairs" design, we mean that units are sampled i.i.d. from the population of interest, paired according to observed, baseline covariates and finally, within each pair, one unit is selected at random for treatment. Importantly, we presume that not all observed, baseline covariates are used in determining treatment assignment. We study a broad class of estimators based on a "doubly robust" moment condition that permits us to study estimators with both finite-dimensional and high-dimensional forms of covariate adjustment. We find that estimators with finite-dimensional, linear adjustments need not lead to improvements in precision relative to the unadjusted difference-in-means estimator. This phenomenon persists even if the adjustments are interacted with treatment; in fact, doing so leads to no changes in precision. However, gains in precision can be ensured by including fixed effects for each of the pairs. Indeed, we show that this adjustment is the "optimal" finite-dimensional, linear adjustment. We additionally study two estimators with high-dimensional forms of covariate adjustment based on the LASSO. For each such estimator, we show that it leads to improvements in precision relative to the unadjusted difference-in-means estimator and also provides conditions under which it leads to the "optimal' nonparametric, covariate adjustment. A simulation study confirms the practical relevance of our theoretical analysis, and the methods are employed to reanalyze data from an experiment using a "matched pairs" design to study the effect of macroinsurance on microenterprise.

KEYWORDS: Experiment, matched pairs, covariate adjustment, randomized controlled trial, treatment assignment, LASSO

JEL classification codes: C12, C14

*Yichong Zhang acknowledges the financial support from the NSFC under the grant No. 72133002 and a Lee Kong Chian fellowship. Any and all errors are our own.

1 Introduction

This paper studies inference on the average treatment effect in experiments in which treatment status is determined according to "matched pairs." By a "matched pairs" design, we mean that units are sampled i.i.d. from the population of interest, paired according to observed, baseline covariates and finally, within each pair, one unit is selected at random for treatment. This method is used routinely in all parts of the sciences. Indeed, commands to facilitate its implementation are included in popular software packages, such as `sampsi` in Stata. References to a variety of specific examples can be found, for instance, in the following surveys of various field experiments: Donner and Klar (2000), Glennerster and Takavarasha (2013), and Rosenberger and Lachin (2015). See also Bruhn and McKenzie (2009), who, based on a survey of selected development economists, report that 56% of researchers have used such a design at some point. Bai et al. (2021) develop methods for inference on the average treatment effect in such experiments based on the difference-in-means estimator. In this paper, we pursue the goal of improving upon the precision of this estimator by exploiting observed, baseline covariates that are not used in determining treatment status.

To this end, we study a broad class of estimators for the average treatment effect based on a "doubly robust" moment condition. The estimators in this framework are distinguished via different "working models" for the conditional expectations of potential outcomes under treatment and control given the observed, baseline covariates. Importantly, because of the double-robustness, these "working models" need not be correctly specified in order for the resulting estimator to be consistent. In this way, the framework permits us to study both finite-dimensional and high-dimensional forms of covariate adjustment without imposing unreasonable restrictions on the conditional expectations themselves. Under high-level conditions on the "working models" and their corresponding estimators and a requirement that pairs are formed so that units within pairs are suitably "close" in terms of the baseline covariates, we derive the limiting distribution of the covariate-adjusted estimator of the average treatment effect. We further construct an estimator for the variance of the limiting distribution and provide conditions under which it is consistent for this quantity.

Using our general framework, we first consider finite-dimensional, linear adjustments. For this class of estimators, our main findings are summarized as follows. First, we find that such adjustments need not lead to improvements in terms of precision upon the unadjusted difference-in-means estimator. This finding echoes similar findings by Yang and Tsiatis (2001) and Tsiatis et al. (2008) in settings in which treatment is determined by i.i.d. coin flips, and Freedman (2008) in a finite population setting in which treatment is determined according to complete randomization. See Negi and Wooldridge (2021) for a succinct treatment of that literature. More surprisingly, we find that this phenomenon persists even if the adjustments are interacted with treatment. In fact, doing so leads to no changes in precision. In this sense, our results diverge from those in Lin (2013), who found in the same setting studied by Freedman (2008) that such interactions ensured gains in precision relative to the unadjusted difference-in-means estimator. We show, however, that gains in precision can be ensured by including fixed effects for each of the pairs. Similar results have been obtained by Fogarty (2018) in a finite population framework for the estimation of the sample average treatment effect. Our analysis further reveals that the resulting covariate-adjusted estimator is "optimal" among all finite-dimensional, linear adjustments. In particular, further interaction of these adjustments with treatment leads to no further improvements. These results support the simulation-based findings of Bruhn and McKenzie (2009), who advocate for including fixed effects for each of the pairs when analyzing such experiments. We emphasize, however, that the usual heteroskedascity-robust standard errors for the corresponding ordinary least squares estimator that naïvely treats the data (including treatment status) as if it were i.i.d. need not be consistent for the limiting variance derived in our analysis.

We then use our framework to consider high-dimensional adjustments based on the LASSO. We study, in particular, two estimators of this form. The first estimator is motivated by the observation that the finite-dimensional, linear adjustment that includes fixed effects for each of the pairs is identical to the intercept term in the linear regression of the pairwise differences in outcomes on the pairwise differences in covariates. The first estimator we consider is therefore defined as the intercept term in a LASSO-penalized regression of the pairwise difference in the outcomes on the pairwise differences in the covariates. As with its finite-dimensional counterpart, we show that this estimator is more precise than the unadjusted difference-in-means estimator. The second estimator we consider first obtains an intermediate estimator by using the LASSO to estimate the "working model" for the relevant conditional expectations. In a finite population setting in which treatment is determined according to complete randomization, Cohen and Fogarty (2020) show that such an estimator is necessarily more precise than the unadjusted difference-in-means estimator. When treatment is determined according to "matched pairs," however, this intermediate estimator need not be the case. We therefore consider, in an additional step, an estimator based on the finite-dimensional, linear adjustment described above that uses the predicted values for the "working model" as the covariates and includes fixed effects for each of the pairs. We show that the resulting estimator improves upon both the intermediate estimator and the unadjusted difference-in-means estimator in terms of precision. Moreover, we provide conditions under which both of these high-dimensional adjustments attain the relevant semi-parametric efficiency bound derived in Armstrong (2022).

The remainder of our paper is organized as follows. In Section 2, we describe our setup and notation. In particular, there we describe the precise sense in which we require that units in each pair are "close" in terms of their baseline covariates. In Section 3, we introduce our general class of estimators based on a "doubly robust" moment condition. Under certain high-level conditions on the "working models" and their corresponding estimators, we derive the limiting behavior of the covariate-adjusted estimator. In Section 4, we use our general framework to study a variety of estimators with finite-dimensional, linear covariate adjustment. In Section 5, we use our general framework to study two estimators with high-dimensional covariate adjustment based on the LASSO. In Section 6, we examine the finite-sample behavior of tests based on these different estimators via a small simulation study. We find that covariate adjustment can lead to considerable gains in precision. Finally, in Section 7, we apply our methods to reanalyze data from an experiment using a "matched pairs" design to study the effect of macroinsurance on microenterprise.

2 Setup and Notation

Let $Y_i \in \mathbf{R}$ denote the (observed) outcome of interest for the ith unit, $D_i \in \{0, 1\}$ be an indicator for whether the ith unit is treated, and $X_i \in \mathbf{R}^{k_x}$ and $W_i \in \mathbf{R}^{k_w}$ denote observed, baseline covariates for the ith unit; X_i and W_i will be distinguished below through the feature that only the former will be used in determining treatment assignment. Further denote by $Y_i(1)$ the potential outcome of the ith unit if treated and by $Y_i(0)$ the potential outcome of the ith unit if not treated. The (observed) outcome and potential outcomes are related to treatment status by the relationship

$$Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i) \ . \tag{1}$$

For a random variable indexed by i, A_i, it will be useful to denote by $A^{(n)}$ the random vector $(A_1, \ldots, A_{2n})$. Denote by P_n the distribution of the observed data $Z^{(n)}$, where $Z_i = (Y_i, D_i, X_i, W_i)$, and by Q_n the distribution of $U^{(n)}$, where $U_i = (Y_i(1), Y_i(0), X_i, W_i)$. Note that P_n is determined by (1), Q_n, and the mechanism for determining treatment assignment. We assume throughout that $U^{(n)}$ consists of $2n$ i.i.d. observations, i.e., $Q_n = Q^{2n}$, where Q is the marginal distribution of U_i. We therefore state our assumptions below in terms of assumptions on Q and the mechanism for determining treatment assignment. Indeed, we will not make reference to P_n in the sequel, and all operations are understood to be under Q and the mechanism for determining the treatment assignment. Our object of interest is the average effect of the treatment on the outcome of interest, which may be expressed in terms of this notation as

$$\Delta(Q) = E[Y_i(1) - Y_i(0)] \ . \tag{2}$$

We now describe our assumptions on Q. We restrict Q to satisfy the following mild requirement:

Assumption 2.1. The distribution Q is such that

(a) $0 < E[\text{Var}[Y_i(d)|X_i]]$ for $d \in \{0, 1\}$.

(b) $E[Y_i^2(d)] < \infty$ for $d \in \{0, 1\}$.

(c) $E[Y_i(d)|X_i = x]$ and $E[Y_i^2(d)|X_i = x]$ are Lipschitz for $d \in \{0, 1\}$.

Next, we describe our assumptions on the mechanism determining treatment assignment. In order to describe these assumptions more formally, we require some further notation to define the relevant pairs of units. The n pairs may be represented by the sets

$$\{\pi(2j - 1), \pi(2j)\} \text{ for } j = 1, \ldots, n \ ,$$

where $\pi = \pi_n(X^{(n)})$ is a permutation of $2n$ elements. Because of its possible dependence on $X^{(n)}$, π encompasses a broad variety of different ways of pairing the $2n$ units according to the observed, baseline covariates $X^{(n)}$. Given such a π, we assume that treatment status is assigned as described in the following assumption:

Assumption 2.2. Treatment status is assigned so that $(Y^{(n)}(1), Y^{(n)}(0), W^{(n)}) \perp\!\!\!\perp D^{(n)}|X^{(n)}$ and, conditional on $X^{(n)}$, $(D_{\pi(2j-1)}, D_{\pi(2j)}), j = 1, \ldots, n$ are i.i.d. and each uniformly distributed over the values in $\{(0, 1), (1, 0)\}$.

Following Bai et al. (2021), our analysis will additionally require some discipline on the way in which pairs are formed. Let $\|\cdot\|_2$ denote the Euclidean norm. We will require that units in each pair are "close" in the sense described by the following assumption:

Assumption 2.3. The pairs used in determining treatment status satisfy

$$\frac{1}{n} \sum_{1 \le j \le n} \|X_{\pi(2j)} - X_{\pi(2j-1)}\|_2^r \xrightarrow{P} 0$$

for $r \in \{1, 2\}$.

It will at times be convenient to require further that units in consecutive pairs are also "close" in terms of their baseline covariates. One may view this requirement, which is formalized in the following assumption, as "pairing the pairs" so that they are "close" in terms of their baseline covariates.

Assumption 2.4. The pairs used in determining treatment status satisfy

$$\frac{1}{n} \sum_{1 \le j \le \lfloor \frac{n}{2} \rfloor} \|X_{\pi(4j-k)} - X_{\pi(4j-\ell)}\|_2^2 \xrightarrow{P} 0$$

for any $k \in \{2, 3\}$ and $\ell \in \{0, 1\}$.

Bai et al. (2021) provide results to facilitate constructing pairs satisfying Assumptions 2.3–2.4 under weak assumptions on Q. In particular, given pairs satisfying Assumption 2.3, it is frequently possible to "re-order" them so that Assumption 2.4 is satisfied. See Theorem 4.3 in Bai et al. (2021) for further details. As in Bai et al. (2021), we highlight the fact that Assumption 2.4 will only be used to enable consistent estimation of relevant variances.

3 Main Results

To accommodate various forms of covariate-adjusted estimators of $\Delta(Q)$ in a single framework, it is useful to note it follows from Assumption 2.2 that for any $d \in \{0, 1\}$ and any function $m_{d,n} : \mathbf{R}^{k_x} \times \mathbf{R}^{k_w} \to \mathbf{R}$ such that $E[|m_{d,n}(X_i, W_i)|] < \infty$,

$$E\left[2I\{D_i = d\}(Y_i - m_{d,n}(X_i, W_i)) + m_{d,n}(X_i, W_i)\right] = E[Y_i(d)] \ . \tag{3}$$

We note that (3) is just the augmented inverse propensity score weighted (AIPW) moment for $E[Y_i(d)]$ in which the propensity score is $1/2$ and the conditional mean model is $m_{d,n}(X_i, W_i)$. Such a moment is also "double robustness." As the propensity score for the "matched pairs" design is exactly $1/2$, we do not require the conditional mean model to be correctly specified, i.e., $m_{d,n}(X_i, W_i) = E[Y_i(d)|X_i, W_i]$. See, for instance, Robins et al. (1995). Intuitively, $m_{d,n}$ is the "working model" which researchers use to estimate $E[Y_i(d)|X_i, W_i]$, and can be arbitrarily misspecified because of (3). Although $m_{d,n}$ will be identical across $n \ge 1$ for the examples in Section 4, the notation permits $m_{d,n}$ to depend on the sample size n in anticipation of the high-dimensional results in Section 5. Based on the moment condition in (3), our proposed estimator of $\Delta(Q)$ is given by

$$\hat{\Delta}_n = \hat{\mu}_n(1) - \hat{\mu}_n(0) \ , \tag{4}$$

where, for $d \in \{0,1\}$,

$$\hat{\mu}_n(d) = \frac{1}{2n} \sum_{1 \le i \le 2n} (2I\{D_i = d\}(Y_i - \hat{m}_{d,n}(X_i, W_i)) + \hat{m}_{d,n}(X_i, W_i)) \tag{5}$$

and $\hat{m}_{d,n}$ is a suitable estimator of the "working model" $m_{d,n}$ in (3).

We require some disciplines on the behavior of $m_{d,n}$ for $d \in \{0,1\}$ and $n \ge 1$:

Assumption 3.1. The functions $m_{d,n}$ for $d \in \{0,1\}$ and $n \ge 1$ satisfy

(a) For $d \in \{0,1\}$,

$$\liminf_{n\to\infty} E\left[\text{Var}\left[Y_i(d) - \frac{1}{2}(m_{1,n}(X_i,W_i) + m_{0,n}(X_i,W_i))\middle| X_i\right]\right] > 0 \ .$$

(b) For $d \in \{0,1\}$,

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} E[m^2_{d,n}(X_i,W_i) I\{|m_{d,n}(X_i,W_i)| > \lambda\}] = 0 \ .$$

(c) $E[m_{d,n}(X_i,W_i)|X_i = x]$, $E[m^2_{d,n}(X_i,W_i)|X_i = x]$, and $E[m_{d,n}(X_i,W_i)Y_i(d)|X_i = x]$ for $d \in \{0,1\}$, and $E[m_{1,n}(X_i,W_i)m_{0,n}(X_i,W_i)|X_i = x]$ are Lipschitz uniformly over $n \ge 1$.

Assumption 3.1(a) is an assumption to rule out degenerate situations. Assumption 3.1(b) is a mild uniform integrability assumption on the "working models." If $m_{d,n} \equiv m_d$ for $d \in \{0,1\}$, then it is satisfied as long as $E[m^2_d(X_i,W_i)] < \infty$. Assumption 3.1(c) ensures that units that are "close" in terms of the observed covariates are also "close" in terms of potential outcomes, uniformly across $n \ge 1$.

Theorem 3.1 below establishes the limit in distribution of $\hat{\Delta}_n$. We note that the theorem depends on high-level conditions on $m_{d,n}$ and $\hat{m}_{d,n}$. In the sequel, these conditions will be verified in several examples.

Theorem 3.1. *Suppose Q satisfies Assumption 2.1, the treatment assignment mechanism satisfies Assumptions 2.2–2.3, and $m_{d,n}$ for $d \in \{0,1\}$ and $n \ge 1$ satisfy Assumption 3.1. Further suppose $\hat{m}_{d,n}$ satisfies*

$$\frac{1}{\sqrt{2n}} \sum_{1 \le i \le 2n} (2D_i - 1)(\hat{m}_{d,n}(X_i,W_i) - m_{d,n}(X_i,W_i)) \xrightarrow{P} 0 \ . \tag{6}$$

Then, $\hat{\Delta}_n$ defined in (4) satisfies

$$\frac{\sqrt{n}(\hat{\Delta}_n - \Delta(Q))}{\sigma_n(Q)} \xrightarrow{d} N(0,1) \ , \tag{7}$$

where $\sigma^2_n(Q) = \sigma^2_{1,n}(Q) + \sigma^2_2(Q) + \sigma^2_3(Q)$ with

$$\begin{aligned}
\sigma^2_{1,n}(Q) &= \frac{1}{2} E\left[\text{Var}\left[E[Y_i(1) + Y_i(0)|X_i,W_i] - (m_{1,n}(X_i,W_i) + m_{0,n}(X_i,W_i))\middle| X_i\right]\right] \\
\sigma^2_2(Q) &= \frac{1}{2}\text{Var}\left[E\left[Y_i(1) - Y_i(0)\middle| X_i, W_i\right]\right] \\
\sigma^2_3(Q) &= E[\text{Var}[Y_i(1)|X_i,W_i] + \text{Var}[Y_i(0)|X_i,W_i]] \ .
\end{aligned}$$

In order to facilitate the use of Theorem 3.1 for inference about $\Delta(Q)$, we next provide a consistent estimator of $\sigma_n(Q)$. Define

$$\tilde{Y}_i = Y_i - \frac{1}{2}(\hat{m}_{1,n}(X_i, W_i) + \hat{m}_{0,n}(X_i, W_i))$$
$$\hat{\tau}_n^2 = \frac{1}{n} \sum_{1 \le j \le n} (\tilde{Y}_{\pi(2j-1)} - \tilde{Y}_{\pi(2j)})^2$$
$$\hat{\lambda}_n = \frac{2}{n} \sum_{1 \le j \le \lfloor \frac{n}{2} \rfloor} (\tilde{Y}_{\pi(4j-3)} - \tilde{Y}_{\pi(4j-2)})(\tilde{Y}_{\pi(4j-1)} - \tilde{Y}_{\pi(4j)})(D_{\pi(4j-3)} - D_{\pi(4j-2)})(D_{\pi(4j-1)} - D_{\pi(4j)}) \ .$$

The variance estimator is given by

$$\hat{\sigma}_n^2 = \hat{\tau}_n^2 - \frac{1}{2}(\hat{\lambda}_n + \hat{\Delta}_n^2) \ . \tag{8}$$

Note it can be shown similarly as in Remark 3.9 of Bai et al. (2021) that $\hat{\sigma}_n^2$ in (8) is nonnegative.

Theorem 3.2 below establishes the consistency of this estimator and its implications for inference about $\Delta(Q)$. In the statement of the theorem, we make use of the following notation: for any scalars a and b, $[a \pm b]$ is understood to be $[a - b, a + b]$.

Theorem 3.2. *Suppose Q satisfies Assumption 2.1, the treatment assignment mechanism satisfies Assumptions 2.2–2.3, and $m_{d,n}$ for $d \in \{0, 1\}$ and $n \ge 1$ satisfy Assumption 3.1. Further suppose $\hat{m}_{d,n}$ satisfies (6) and*

$$\frac{1}{2n} \sum_{1 \le i \le 2n} (\hat{m}_{d,n}(X_i, W_i) - m_{d,n}(X_i, W_i))^2 \xrightarrow{P} 0 \ . \tag{9}$$

Then,

$$\frac{\hat{\sigma}_n}{\sigma_n(Q)} \xrightarrow{P} 1 \ .$$

Hence, (7) holds with $\hat{\sigma}_n$ in place of $\sigma_n(Q)$. In particular, for any $\alpha \in (0, 1)$,

$$P\left\{\Delta(Q) \in \left[\hat{\Delta}_n \pm \hat{\sigma}_n \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right]\right\} \to 1 - \alpha \ ,$$

where Φ is the standard normal c.d.f.

Remark 3.1. An important and immediate implication of Theorem 3.1 is that $\sigma_n^2(Q)$ is minimized when

$$E[Y_i(0) + Y_i(1)|X_i, W_i] - E[Y_i(0) + Y_i(1)|X_i] = m_{0,n}(X_i, W_i) + m_{1,n}(X_i, W_i) - E[m_{0,n}(X_i, W_i) + m_{1,n}(X_i, W_i)|X_i]$$

with probability one. In other words, the "working model" for $E[Y_i(0)+Y_i(1)|X_i, W_i]$ given by $m_{0,n}(X_i, W_i)+m_{1,n}(X_i, W_i)$, need only be correct "on average" over the variables that are not used in determining the pairs. For such a choice of $m_{0,n}(X_i, W_i)$ and $m_{1,n}(X_i, W_i)$, $\sigma_n^2(Q)$ in Theorem 3.1 becomes simply

$$\frac{1}{2} \operatorname{Var}\left[E\left[Y_i(1) - Y_i(0)\middle|X_i, W_i\right]\right] + E[\operatorname{Var}[Y_i(1)|X_i, W_i] + \operatorname{Var}[Y_i(0)|X_i, W_i]] \ ,$$

which agrees with the variance obtained in Bai et al. (2021) when both X_i and W_i are used in determining the pairs. Such a variance also achieves the efficiency bound derived by Armstrong (2022). ■

Remark 3.2. Following Bai et al. (2022), it is straightforward to extend the analysis in this paper to the case with multiple treatment arms and where treatment status is determined using a "matched tuples" design, but we do not pursue this further in this paper. ■

4 Linear Adjustments

In this section, we consider linearly covariate-adjusted estimators of $\Delta(Q)$ based on a set of regressors generated by $X_i \in \mathbf{R}^{k_x}$ and $W_i \in \mathbf{R}^{k_w}$. To this end, define $\psi_i = \psi(X_i, W_i)$, where $\psi : \mathbf{R}^{k_x} \times \mathbf{R}^{k_w} \to \mathbf{R}^p$. We impose the following assumptions on the function ψ:

Assumption 4.1. The function ψ is such that

(a) no component of ψ is constant and $E[\text{Var}[\psi_i|X_i]]$ is nonsingular.

(b) $\text{Var}[\psi_i] < \infty$.

(c) $E[\psi_i|X_i = x]$, $E[\psi_i\psi_i'|X_i = x]$, and $E[\psi_i Y_i(d)|X_i = x]$ for $d \in \{0, 1\}$ are Lipschitz.

Assumption 4.1 is analogous to Assumption 2.1. Note, in particular, that Assumption 4.1(a) rules out situations where ψ_i is a function of X_i only. See Remark 4.2 for a discussion of the behavior of the covariate-adjusted estimators in such situations.

4.1 Linear Adjustments without Pair Fixed Effects

Consider the following linear regression model:

$$Y_i = \alpha + \Delta D_i + \psi_i'\beta + \epsilon_i \ . \tag{10}$$

Let $\hat{\alpha}_n^{\text{naive}}$, $\hat{\Delta}_n^{\text{naive}}$, and $\hat{\beta}_n^{\text{naive}}$ denote the OLS estimators of α, Δ, and β in (10). It follows from direct calculation that

$$\hat{\Delta}_n^{\text{naive}} = \frac{1}{n} \sum_{1 \le i \le 2n} (Y_i - \psi_i'\hat{\beta}_n^{\text{naive}})(2D_i - 1) \ .$$

Therefore, $\hat{\Delta}_n^{\text{naive}}$ satisfies (4)–(5) with

$$\hat{m}_{d,n}(X_i, W_i) = \psi_i'\hat{\beta}_n^{\text{naive}} \ .$$

Theorem 4.1 establishes (6) and (9) for a suitable choice of $m_{d,n}(X_i, W_i)$ for $d \in \{0, 1\}$ and, as a result, the limiting distribution of $\hat{\Delta}_n^{\text{naive}}$ and the validity of the variance estimator.

Theorem 4.1. *Suppose Q satisfies Assumption 2.1 and the treatment assignment mechanism satisfies Assumptions 2.2–2.3. Further suppose ψ satisfies Assumption 4.1. Then, as $n \to \infty$,*

$$\hat{\beta}_n^{\text{naive}} \xrightarrow{P} \beta^{\text{naive}} = \text{Var}[\psi_i]^{-1} \, \text{Cov}[\psi_i, Y_i(1) + Y_i(0)] \ .$$

Moreover, (6)*,* (9)*, and Assumption* 3.1 *are satisfied with*

$$m_{d,n}(X_i, W_i) = \psi_i' \beta^{\text{naive}}$$

for $d \in \{0, 1\}$ *and* $n \geq 1$*.*

Remark 4.1. Freedman (2008) studies regression adjustment based on (10) when treatment is assigned by complete randomization instead of a "matched pairs" design. In such settings, Lin (2013) proposes adjustment based on the following linear regression model:

$$Y_i = \alpha + \Delta D_i + (\psi_i - \bar{\psi}_n)' \gamma + D_i(\psi_i - \bar{\psi}_n)' \eta + \epsilon_i \ , \tag{11}$$

where

$$\bar{\psi}_n = \frac{1}{2n} \sum_{1 \leq i \leq 2n} \psi_i \ .$$

Let $\hat{\alpha}_n^{\text{int}}, \hat{\Delta}_n^{\text{int}}, \hat{\gamma}_n^{\text{int}}, \hat{\eta}_n^{\text{int}}$ denote the OLS estimators for $\alpha, \Delta, \gamma, \eta$ in (11). It is straightforward to show $\hat{\Delta}_n^{\text{int}}$ satisfies (4)–(5) with

$$\hat{m}_{1,n}(X_i, W_i) = (\psi_i - \hat{\mu}_{\psi,n}(1))'(\hat{\gamma}_n^{\text{int}} + \hat{\eta}_n^{\text{int}})$$
$$\hat{m}_{0,n}(X_i, W_i) = (\psi_i - \hat{\mu}_{\psi,n}(0))' \hat{\gamma}_n^{\text{int}} \ ,$$

where

$$\hat{\mu}_{\psi,n}(d) = \frac{1}{n} \sum_{1 \leq i \leq 2n} I\{D_i = d\} \psi_i \ .$$

It can be shown using similar arguments to those used to establish Theorem 4.1 that (6) and Assumption 3.1 are satisfied with

$$m_{d,n}(X_i, W_i) = (\psi_i - E[\psi_i])' \operatorname{Var}[\psi_i]^{-1} \operatorname{Cov}[\psi_i, Y_i(d)]$$

for $d \in \{0, 1\}$ and $n \geq 1$. It thus follows by inspecting the expression for $\sigma_n^2(Q)$ in Theorem 3.1 that the limiting variance of $\hat{\Delta}_n^{\text{int}}$ is the same as that of $\hat{\Delta}_n^{\text{naive}}$ based on (10). ■

4.2 Linear Adjustments with Pair Fixed Effects

Remark 4.1 implies that in "matched pairs" designs, including interaction terms in the linear regression does not lead to an estimator with lower limiting variance than the one based on the linear regression without interaction terms. It is therefore interesting to study whether there exists a linearly covariate-adjusted estimator with lower limiting variance than the one based on (10) and (11). To that end, consider instead the following linear regression model:

$$Y_i = \Delta D_i + \psi_i' \beta + \sum_{1 \leq j \leq n} \theta_j I\{i \in \{\pi(2j-1), \pi(2j)\}\} + \epsilon_i \ . \tag{12}$$

Let $\hat{\Delta}_n^{\rm pfe}$, $\hat{\beta}_n^{\rm pfe}$, and $\hat{\gamma}_{j,n}$, $1 \le j \le n$ denote the OLS estimators of Δ, β, θ_j, $1 \le j \le n$ in (12). It follows from the Frisch-Waugh-Lovell theorem that

$$\hat{\Delta}_n^{\rm pfe} = \frac{1}{n} \sum_{1 \le i \le 2n} (Y_i - \psi_i' \hat{\beta}_n^{\rm pfe})(2D_i - 1) \ .$$

Therefore, $\hat{\Delta}_n^{\rm pfe}$ satisfies (4)–(5) with

$$\hat{m}_{d,n}(X_i, W_i) = \psi_i' \hat{\beta}_n^{\rm pfe} \ .$$

Theorem 4.2 establishes (6) and (9) for a suitable choice of $m_{d,n}(X_i, W_i), d \in \{0, 1\}$ and, as a result, the limiting distribution of $\hat{\Delta}_n^{\rm pfe}$ and the validity of the variance estimator.

Theorem 4.2. *Suppose Q satisfies Assumption 2.1 and the treatment assignment mechanism satisfies Assumptions 2.2–2.3. Then, as $n \to \infty$,*

$$\hat{\beta}_n^{\rm pfe} \xrightarrow{P} \beta^{\rm pfe} = (2E[\mathrm{Var}[\psi_i | X_i]])^{-1} E[\mathrm{Cov}[\psi_i, Y_i(1) + Y_i(0) | X_i]] \ .$$

Moreover, (6), (9), and Assumption 3.1 are satisfied with

$$m_{d,n}(X_i, W_i) = \psi_i' \beta^{\rm pfe}$$

for $d \in \{0, 1\}$ and $n \ge 1$.

Remark 4.2. When ψ is restricted to be a function of X_i only, then $\hat{\Delta}_n^{\rm pfe}$ coincides to first order with the unadjusted difference-in-means estimator defined as

$$\hat{\Delta}_n^{\rm unadj} = \frac{1}{n} \sum_{1 \le i \le 2n} Y_i D_i - \frac{1}{n} \sum_{1 \le i \le 2n} Y_i (1 - D_i) \ . \tag{13}$$

To see this, suppose further that ψ is Lipschitz and that $\mathrm{Var}[Y_i(d) | X_i = x], d \in \{0, 1\}$ are bounded. The proof of Theorem 4.2 reveals that $\hat{\Delta}_n^{\rm pfe}$ and $\hat{\beta}_n^{\rm pfe}$ coincide with the OLS estimators of the intercept and slope parameters in a linear regression of $(Y_{\pi(2j)} - Y_{\pi(2j-1)})(D_{\pi(2j)} - D_{\pi(2j-1)})$ on a constant and $(\psi_{\pi(2j)} - \psi_{\pi(2j-1)})(D_{\pi(2j)} - D_{\pi(2j-1)})$. Using this observation, it follows by arguing as in Section S.1.1 of Bai et al. (2021) that

$$\sqrt{n}(\hat{\Delta}_n^{\rm pfe} - \Delta(Q)) = \sqrt{n}(\hat{\Delta}_n^{\rm unadj} - \Delta(Q)) + o_P(1) \ .$$

See also Remark 3.8 of Bai et al. (2021). ■

Remark 4.3. Note in the expression of $\sigma_n^2(Q)$ in Theorem 3.1 only depends on $m_{d,n}(X_i, W_i), d \in \{0, 1\}$ through $\sigma_{1,n}^2(Q)$. With this in mind, consider the class of all linearly covariate-adjusted estimators based on ψ_i, i.e., $m_{d,n}(X_i, W_i) = \psi_i' \beta(d)$. For this specification of $m_{d,n}(X_i, W_i), d \in \{0, 1\}$,

$$\sigma_{1,n}^2(Q) = E[(E[Y_i(1) + Y_i(0) | X_i, W_i] - E[Y_i(1) + Y_i(0) | X_i] - (\psi_i - E[\psi_i | X_i])'(\beta(1) + \beta(0)))^2] \ .$$

It follows that among all such linear adjustments, $\sigma_n^2(Q)$ in (7) is minimized when

$$\beta(1) + \beta(0) = 2\beta^{\rm pfe} \ .$$

This observation implies that the linear adjustment with pair fixed effects, i.e., $\hat{\Delta}_n^{\rm pfe}$, yields the optimal linear adjustment in the sense of minimizing $\sigma_n^2(Q)$. Its limiting variance is, in particular, weakly smaller than the limiting variance of the unadjusted difference-in-means estimator defined in (13). The same limiting variance is attained by $m_{d,n}(X_i, W_i) = \psi_i'\beta(d) + h_d(X_i)$ for $d \in \{0, 1\}$. On the other hand, the covariate-adjusted estimators based on (10) or (11), i.e., $\hat{\Delta}_n^{\rm naive}$ and $\hat{\Delta}_n^{\rm int}$, are in general not optimal among all linearly covariate-adjusted estimators based on ψ_i. In fact, the limiting variances of these two estimators may even be larger than that of the unadjusted difference-in-means estimator. Simulation evidence in Section 7 illustrates such a phenomenon in an example. In this sense, these estimators suffer from a counterpart to the critique raised by Freedman (2008). ■

Remark 4.4. Even though $\hat{\Delta}_n^{\rm pfe}$ can be computed via ordinary least squares estimation of (12), we emphasize that the usual heteroskedascity-robust standard errors that naïvely treats the data (including treatment status) as if it were i.i.d. need not be consistent for the limiting variance derived in our analysis. See Bai et al. (2022) for details. ■

Remark 4.5. One can also consider the estimator based on the following linear regression model:

$$Y_i = \Delta D_i + (\psi_i - \bar{\psi}_n)'\gamma + D_i(\psi_i - \hat{\mu}_{\psi,n}(1))'\eta + \sum_{1 \le j \le n} \theta_j I\{i \in \{\pi(2j-1), \pi(2j)\}\} + \epsilon_i \ . \tag{14}$$

Let $\hat{\Delta}_n^{\rm int-pfe}, \hat{\gamma}_n^{\rm int-pfe}, \hat{\eta}_n^{\rm int-pfe}$ denote the OLS estimators for Δ, γ, η in (14). It is straightforward to show $\hat{\Delta}_n^{\rm int-pfe}$ satisfies (4)–(5) with

$$\begin{aligned} \hat{m}_{1,n}(X_i, W_i) &= (\psi_i - \hat{\mu}_{\psi,n}(1))'\hat{\eta}_n^{\rm int-pfe} \\ \hat{m}_{0,n}(X_i, W_i) &= (\psi_i - \hat{\mu}_{\psi,n}(0))'(\hat{\eta}_n^{\rm int-pfe} - \hat{\gamma}_n^{\rm int-pfe}) \ . \end{aligned}$$

Following similar arguments to those used in the proof of Theorem 4.1, we can establish that (6) and Assumption 3.1 are satisfied with

$$\begin{aligned} m_{1,n}(X_i, W_i) &= (\psi_i - E[\psi_i])'\eta^{\rm int-pfe} \\ m_{0,n}(X_i, W_i) &= (\psi_i - E[\psi_i])'(\eta^{\rm int-pfe} - \gamma^{\rm int-pfe}) \ , \end{aligned}$$

where

$$\begin{aligned} \gamma^{\rm int-pfe} &= (E[\mathrm{Var}[\psi_i|X_i]])^{-1} E[\mathrm{Cov}[\psi_i, Y_i(1) - Y_i(0)|X_i]] \\ \eta^{\rm int-pfe} &= (E[\mathrm{Var}[\psi_i|X_i]])^{-1} E[\mathrm{Cov}[\psi_i, Y_i(1)|X_i]] \end{aligned}$$

Because $2\eta^{\rm int-pfe} - \gamma^{\rm int-pfe} = 2\beta^{\rm pfe}$, it follows from Remark 4.3 that the limiting variance of $\hat{\Delta}_n^{\rm int-pfe}$ is identical to the limiting variance of $\hat{\Delta}_n^{\rm pfe}$. ■

5 High-Dimensional Adjustments

In this section, we study covariate adjustments based on high-dimensional regressors. Such settings can arise if the covariates W_i are high-dimensional or if the regressors include many transformations of X_i and W_i. To accommodate situations where the dimension of W_i increases with n, we add a subscript and denote it by $W_{n,i}$ instead. Let $k_{w,n}$ denote the dimension of $W_{n,i}$. For $n \geq 1$, let $\psi_{n,i} = \psi_n(X_i, W_{n,i})$, where $\psi_n : \mathbf{R}^{k_x} \times \mathbf{R}^{k_{w,n}} \to \mathbf{R}^{p_n}$ and p_n will be permitted below to be possibly much larger than n.

In what follows, we propose two distinct LASSO-based high-dimensional counterparts to $\hat{\Delta}_n^{\text{pfe}}$ studied in Section 4.2. The first method is motivated by the observation in Section 4.2 that $\hat{\Delta}_n^{\text{pfe}}$ satisfies (4)–(5) with

$$\hat{m}_{d,n}(X_i, W_i) = \psi_i' \hat{\beta}_n^{\text{pfe}} ,$$

where $\hat{\beta}_n^{\text{pfe}}$ can be obtained, as described in Remark 4.2, through OLS regression of the pairwise differences in the outcomes on a constant and the pairwise differences in the covariates. For our first method, we therefore consider a LASSO-penalized version of this same procedure. As explained further below in Theorem 5.1, when, for $d \in \{0, 1\}$, $m_{d,n}(X_i, W_i)$ is sufficiently well approximated by a sparse linear function of $\psi_{n,i}$, the resulting estimator, $\hat{\Delta}_n^{\text{hd-pd}}$, is optimal in the sense that it minimizes the limiting variance in Theorem 3.1. Moreover, when this is not the case, its limiting variance is still weakly smaller than the limiting variance of the unadjusted difference-in-means estimator.

The second method is a two-step method in the spirit of Fogarty (2018). In the first step, an intermediate estimator, $\hat{\Delta}_n^{\text{hd}}$, is obtained using (4) with a "working model" obtained through a LASSO-based approximation to $m_{d,n}(X_i, W_i)$. As explained further below in Theorem 5.2, when, for $d \in \{0, 1\}$, $m_{d,n}(X_i, W_i)$ is sufficiently well approximated by a sparse linear function of $\psi_{n,i}$, such an estimator is also optimal in the sense that it minimizes the limiting variance in Theorem 3.1. When this is not the case, however, for reasons analogous to those put forward in Remark 4.2, it need not have a limiting variance weakly smaller than the unadjusted difference-in-means estimator. In a second step, we therefore consider an estimator based on OLS estimation of a version of (12) in which the covariates ψ_i are replaced by the LASSO-based estimates of $m_{d,n}(X_i, W_i)$ for $d \in \{0, 1\}$. The resulting estimator, $\hat{\Delta}_n^{\text{hd-f}}$, has limiting variance weakly smaller than that of the intermediate estimator and thus remains optimal in the same sense. Moreover, like $\hat{\Delta}_n^{\text{hd-pd}}$, it too has limiting variance weakly smaller than the unadjusted difference-in-means estimator. Some comparisons between $\hat{\Delta}_n^{\text{hd-pd}}$ and $\hat{\Delta}_n^{\text{hd-f}}$ are described in Remark 5.5.

Before proceeding, we introduce some additional notation that will be required in our formal description of the methods. To this end, define

$$\begin{aligned} \mu_d(X_i) &= E[Y_i(d)|X_i] \\ \Psi(X_i) &= E[\psi_{n,i}|X_i] \\ \tilde{\psi}_{n,i} &= \psi_{n,i} - \Psi(X_i) \ . \end{aligned}$$

We denote by $\psi_{n,i,l}$ and $\tilde{\psi}_{n,i,l}$ the $l$the components of $\psi_{n,i}$ and $\tilde{\psi}_{n,i}$, respectively. For a vector $a \in \mathbf{R}^k$ and

$0 \leq p \leq \infty$, recall that

$$\|a\|_p = \Big(\sum_{1 \leq l \leq k} |a_l|^p \Big)^{1/p} ~,$$

where it is understood that $\|a\|_0 = \sum_{1 \leq l \leq k} I\{a_k \neq 0\}$ and $\|a\|_\infty = \sup_{1 \leq l \leq k} |a_l|$. Using this notation, we further define

$$\Xi_n = \sup_{(x,w) \times \mathrm{supp}(X_i) \times \mathrm{supp}(W_i)} \|\psi_{n,i}(x,w)\|_\infty ~.$$

5.1 First LASSO-based Adjustment

Define

$$(\hat{\alpha}_n^{\mathrm{hd-pd}}, \hat{\beta}_n^{\mathrm{hd-pd}}) \in \operatorname*{argmin}_{a \in \mathbf{R}, b \in \mathbf{R}^{p_n}} \frac{1}{n} \sum_{1 \leq j \leq n} (\delta_{Y,j} - a - \delta_{\psi,j}' b)^2 + \lambda_n^{\mathrm{hd-pd}} \|\hat{\Omega}_n b\|_1 ~, \tag{15}$$

where

$$\begin{aligned} \delta_{Y,j} &= (D_{\pi(2j-1)} - D_{\pi(2j)})(Y_{\pi(2j-1)} - Y_{\pi(2j)}) \\ \delta_{\psi,j} &= (D_{\pi(2j-1)} - D_{\pi(2j)})(\psi_{n,\pi(2j-1)} - \psi_{n,\pi(2j)}) ~, \end{aligned}$$

$\lambda_n^{\mathrm{hd-pd}}$ is a penalty parameter that will be disciplined by the assumptions below, $\hat{\Omega}_n = \mathrm{diag}(\hat{\omega}_{n,1}, \cdots, \hat{\omega}_{n,p_n})$ is a diagonal matrix, and $\hat{\omega}_{n,l}$ is the penalty loading for the lth regressor. For some $\underline{c}$ and $\bar{c}$, we require that

$$0 < \underline{c} \leq \liminf_{n \to \infty} \min_{1 \leq l \leq p_n} \hat{\omega}_{n,l} \leq \limsup_{n \to \infty} \max_{1 \leq l \leq p_n} \hat{\omega}_{n,l} \leq \bar{c} < \infty \tag{16}$$

with probability one. Let $\hat{\Delta}_n^{\mathrm{hd-pd}}$ denote the estimator in (4) with $\hat{m}_{1,n}(X_i, W_{n,i}) = \hat{m}_{0,n}(X_i, W_{n,i}) = \hat{\alpha}_n^{\mathrm{hd-pd}} + \psi_{n,i}' \hat{\beta}_n^{\mathrm{hd-pd}}$. Because there is no penalty term for a in (15), $\hat{\Delta}_n^{\mathrm{hd-pd}} = \hat{\alpha}_n^{\mathrm{hd-pd}}$.

Our analysis of this estimator will require the following assumptions. In our statement of the assumptions, we will make use of the quantity $(\alpha_{d,n}^{\mathrm{hd-pd}}, \beta_{d,n}^{\mathrm{hd-pd}})$, which will be assumed to satisfy

$$s_n^{\mathrm{hd-pd}} = \max_{d \in \{0,1\}} \|\beta_{d,n}^{\mathrm{hd-pd}}\|_0 \tag{17}$$

and

$$\|E[(1, \tilde{\psi}_{n,i}')' \epsilon_{n,i}^{\mathrm{hd-pd}}(d)]\|_\infty = o(\lambda_n^{\mathrm{hd-pd}}) ~, \tag{18}$$

where

$$\epsilon_{n,i}^{\mathrm{hd-pd}}(d) = Y_i(d) - \alpha_{d,n}^{\mathrm{hd-pd}} - \tilde{\psi}_{n,i}' \beta_{d,n}^{\mathrm{hd-pd}} ~.$$

It is instructive to note that (17) requires $\beta_{d,n}^{\mathrm{hd-pd}}$ to be sparse and (18) is the subgradient condition for a ℓ_1-penalized regression of the outcome $Y_i(d)$ on $\tilde{\psi}_{n,i}$ when the penalty is of order $o(\lambda_n^{\mathrm{hd-pd}})$. If $p_n = o(n)$, then both conditions are satisfied for the $\beta_{d,n}^{\mathrm{hd-pd}}$ equal to the coefficients of a linear projection of $Y_i(d)$ onto $\tilde{\psi}_{n,i}$. When $p_n \gg n$, but $E[Y_i(d) | X_i, W_i]$ is approximately sparse in the sense that there exists some sparse $\beta_{d,n}^*$ with $\max_{d \in \{0,1\}} ||\beta_{d,n}^*||_0 \ll n$ such that the approximation error $|E[Y_i(d)|X_i, W_i] - \tilde{\psi}_{n,i}' \beta_{d,n}^*|$ is sufficiently small, then (17) and (18) are satisfied for $\beta_{d,n}^{\mathrm{hd-pd}} = \beta_{d,n}^*$. We emphasize, however, these conditions can still hold when $E[Y_i(d)|X_i, W_i]$ is neither approximately sparse nor linear in $\tilde{\psi}_{n,i}$. We additionally require that

$$\limsup_{n\to\infty} \max_{d\in\{0,1\}} ||\beta_{d,n}^{\mathrm{hd-pd}}||_\infty < \infty \ . \tag{19}$$

Further restrictions on $\beta_{d,n}^{\mathrm{hd-pd}}$ and $\lambda_n^{\mathrm{hd-pd}}$ will be imposed through a combination of the assumptions below.

We now proceed with the statement of our assumptions. The first assumption collects a variety of moment conditions that will be used in our formal analysis:

Assumption 5.1. (a) For some $q > 2$ and constant C_1,

$$\sup_{n\geq 1} \max_{1\leq l\leq p_n} E[|\psi_{n,i,l}^q||X_i] \leq C_1$$

$$\sup_{n\geq 1} |\psi_{n,i}'\beta_{d,n}^{\mathrm{hd-pd}}| \leq C_1$$

$$\sup_{n\geq 1} |E[Y_i(a)|X_i, W_{n,i}]| \leq C_1$$

with probability one.

(b) For some c_0, $\underline{\sigma}$, $\bar{\sigma}$, the following statements hold with probability one:

$$\max_{d\in\{0,1\},1\leq l\leq p_n} \frac{1}{2n} \sum_{1\leq i\leq 2n} E[\epsilon_{n,i}^4(d)|X_i] \leq c_0 < \infty$$

$$\sup_{n\geq 1} \max_{d\in\{0,1\}} E[\epsilon_{n,i}^4(d)] \leq c_0 < \infty$$

$$\min_{d\in\{0,1\}} \mathrm{Var}[Y_i(d) - \psi_{n,i}'(\beta_{1,n}^{\mathrm{hd-pd}} + \beta_{0,n}^{\mathrm{hd-pd}})/2] \geq \underline{\sigma}^2 > 0$$

$$\min_{1\leq l\leq p_n} \frac{1}{n} \sum_{1\leq i\leq 2n} I\{D_i = d\} \mathrm{Var}[\tilde{\psi}_{n,i,l}\epsilon_{n,i}(d)|X_i] \geq \underline{\sigma}^2 > 0$$

$$\min_{1\leq l\leq p_n} \frac{1}{n} \sum_{1\leq j\leq n} E[\tilde{\psi}_{n,\pi(2j),l}^2|D^{(n)}, X^{(n)}]E[\epsilon_{n,\pi(2j-1)}^2 D_{\pi(2j-1)}|D^{(n)}, X^{(n)}] \geq \underline{\sigma}^2 > 0$$

$$\min_{1\leq l\leq p_n} \frac{1}{n} \sum_{1\leq j\leq n} E[\tilde{\psi}_{n,\pi(2j-1),l}^2|D^{(n)}, X^{(n)}]E[\epsilon_{n,\pi(2j)}^2 D_{\pi(2j)}|D^{(n)}, X^{(n)}] \geq \underline{\sigma}^2 > 0$$

$$\min_{1\leq l\leq p_n, d\in\{0,1\}} \mathrm{Var}[E[\tilde{\psi}_{n,i,l}\epsilon_{n,i}(d)|X_i]] \geq \underline{\sigma}^2 > 0 \ .$$

Assumption 5.1(a)–(b) are standard in the high-dimensional estimation literature; see, for instance, Belloni et al. (2017). The last four inequalities in Assumption 5.1(b), in particular, permit us to apply the high-dimensional central limit theorem in Chernozhukov et al. (2017, Theorem 2.1).

As in the preceding sections, we will additionally require some discipline on the way in which pairs are formed. As before, we will require that units in each pair are "close" in the sense described by the first part of the following assumption, but we will additionally require a Lipschitz-like condition that will play the role of Assumption 2.1(c). Bai et al. (2021) provide algorithms ensuring that part (a) is satisfied with $\zeta_n = O(n^{-1/(2k_x)})$ under weak assumptions on the distribution of X_i.

Assumption 5.2. (a) For some ζ_n,

$$\left(\frac{1}{n}\sum_{1\le j\le n}\left\|X_{\pi(2j)}-X_{\pi(2j-1)}\right\|_2^2\right)^{1/2}\le\sqrt{\mathrm{Var}[||X_i||_2]}\zeta_n \quad \text{with probability one} .$$

(b) For some $L>0$ and any x_1 and x_2 in the support of X_i, we have

$$|(\Psi(x_1)-\Psi(x_2))'\beta_{d,n}^{\mathrm{hd-pd}}|\le L||x_1-x_2||_2 .$$

We next specify our restrictions on the penalty parameter $\lambda_n^{\mathrm{hd-pd}}$.

Assumption 5.3. (a) For some $\ell\ell_n\to\infty$,

$$\lambda_n^{\mathrm{hd-pd}}=\ell\ell_n\left(\frac{1}{\sqrt{n}}\Phi^{-1}\left(1-\frac{0.1}{2\log(n)p_n}\right)+\zeta_n\right) ,$$

where ζ_n is as in Assumption 5.2(a).

(b) $\Xi_n^2(\log(p_n\vee n))^7/n\to 0$, $\ell\ell_n s_n^{\mathrm{hd-pd}}\log(p_n\vee n)/\sqrt{n}\to 0$, and $s_n^{\mathrm{hd-pd}}\log^{1/2}(p_n\vee n)\ell\ell_n\zeta_n\to 0$.

We note that the first three requirements in Assumption 5.3(b) allow p_n to be much greater than n. If $\zeta_n=O(n^{-1/(2k_x)})$, then the last requirement in Assumption 5.3(b) implies $s_n^{\mathrm{hd-pd}}=o(n^{1/(2k_x)})$.

Finally, as is common in the analysis of ℓ_1-penalized regression, we require a "restricted eigenvalue" condition. See, for instance, Belloni et al. (2017). This assumption permits us to apply Bickel et al. (2009, Lemma 4.1) and establish the error bounds for

$$|\hat\alpha_n^{\mathrm{hd-pd}}-\alpha_n^{\mathrm{hd-pd}}|+||\hat\beta_n^{\mathrm{hd-pd}}-\beta_n^{\mathrm{hd-pd}}||_1 \quad\text{and}\quad \frac{1}{n}\sum_{1\le j\le n}\left(\hat\alpha_n^{\mathrm{hd-pd}}-\alpha_n^{\mathrm{hd-pd}}+\delta'_{\psi,j}(\hat\beta_n^{\mathrm{hd-pd}}-\beta_n^{\mathrm{hd-pd}})\right)^2 ,$$

where $\alpha_n^{\mathrm{hd-pd}}=\alpha_{1,n}^{\mathrm{hd-pd}}-\alpha_{0,n}^{\mathrm{hd-pd}}$ and $\beta_n^{\mathrm{hd-pd}}=(\beta_{1,n}^{\mathrm{hd-pd}}+\beta_{0,n}^{\mathrm{hd-pd}})/2$.

Assumption 5.4. For some $\kappa_1>0$, κ_2 and $\ell_n\to\infty$, the following statements hold with probability approaching one:

$$\inf_{d\in\{0,1\},v\in\mathbf{R}^{p_n+1}:\|v\|_0\le(s_n^{\mathrm{hd-pd}}+1)\ell_n}(\|v\|_2^2)^{-1}v'\left(\frac{1}{n}\sum_{1\le j\le n}\breve\delta_{\psi,j}\breve\delta'_{\psi,j}\right)v\ge\kappa_1$$
$$\sup_{d\in\{0,1\},v\in\mathbf{R}^{p_n+1}:\|v\|_0\le(s_n^{\mathrm{hd-pd}}+1)\ell_n}(\|v\|_2^2)^{-1}v'\left(\frac{1}{n}\sum_{1\le j\le n}\breve\delta_{\psi,j}\breve\delta'_{\psi,j}\right)v\le\kappa_2 ,$$

where $\breve\delta_{\psi,j}=(1,\delta'_{\psi,j})'$.

Using these assumptions, the following theorem characterizes the behavior of $\hat\Delta_n^{\mathrm{hd-pd}}$:

Theorem 5.1. *Suppose Q satisfies Assumption 2.1 and the treatment assignment mechanism satisfies Assumptions 2.2–2.3. Further suppose Assumptions 5.1–5.4 and (16) – (19) hold. Then, (6), (9), and Assumption 3.1 are satisfied with $\hat m_{d,n}=\hat\alpha_n^{\mathrm{hd-pd}}+\psi'_{n,i}\hat\beta_n^{\mathrm{hd-pd}}$ and $m_{d,n}(X_i,W_{n,i})=\alpha_n^{\mathrm{hd-pd}}+\psi'_{n,i}\beta_n^{\mathrm{hd-pd}}$ for*

$d \in \{0,1\}$ *and* $n \geq 1$. *Moreover, the variance of* $\hat{\Delta}_n^{\mathrm{hd-pd}}$*, denoted by* $\sigma_n^{\mathrm{hd-pd},2}$*, satisfies*

$$\limsup_{n\to\infty}(\sigma_n^{\mathrm{hd-pd},2} - \sigma_n^{\mathrm{na},2}) \leq 0.$$

If we further assume the true specification is approximately sparse, i.e., there exists $\beta^*_{d,n}$ *such that* $||\beta^*_{d,n}||_1 = O(s_n)$, $E[Y_i(d)|X_i, W_i] = \alpha^*_{d,n} + \psi'_{n,i}\beta^*_{d,n} + R_{n,i}$ *and* $E[R^2_{n,i}] = o(1)$*, then* $\sigma_n^{\mathrm{hd-pd},2}$ *achieves the minimum variance, i.e.,*

$$\lim_{n\to\infty}\sigma_n^{\mathrm{hd-pd},2} = \sigma_2^2(Q) + \sigma_3^2(Q)\ .$$

Remark 5.1. If the additional covariates $W_{n,i}$ are fixed-dimensional, and $\psi_{n,i}$ contains sieve bases of $(W_{n,i}, X_i)$, then approximate sparsity holds under appropriate smoothness conditions on the conditional expectation. Under these circumstances, Theorem 5.1 implies the LASSO-based adjustment achieves minimum variance derived in Remark 3.1, which achieves the semiparametric efficiency bound derived by Armstrong (2022). ■

Remark 5.2. In practice, we choose

$$\ell\ell_n = \sqrt{\log\log n}/5$$

and replace ζ_n by

$$\left(\frac{1}{n}\sum_{1\leq j\leq n}\left\|X_{\pi(2j)} - X_{\pi(2j-1)}\right\|_2^2\right)^{1/2}/\hat{\sigma}_X\ ,$$

where $\hat{\sigma}_X$ is the sample standard deviation of $\{||X_i||_2\}_{1\leq i\leq n}$. ■

Remark 5.3. While our theory only requires that $\hat{\omega}_{n,\ell}, \ell = 1, \ldots, p_n$ satisfy (16), we recommend employing an iterative estimation procedure outlined by Belloni et al. (2017) to estimate $(\hat{\alpha}_n^{\mathrm{hd-pd},(m)}, \hat{\beta}_{d,n}^{\mathrm{hd-pd}})$, in which the m-th step's penalty loadings are estimated based the $(m-1)$-th step's LASSO estimates. Formally, this iterative procedure is described by the following algorithm:

Algorithm 5.1.

Step 0: Set $\hat{\epsilon}_{n,j}^{\mathrm{hd-pd},(0)} = \delta_{Y,j}$.

⋮

Step m: Compute $\hat{\omega}_{n,l}^{(m)} = \sqrt{\frac{1}{n}\sum_{1\leq j\leq n}\delta^2_{\psi,j,l}(\hat{\epsilon}_{n,j}^{\mathrm{hd-pd},(m-1)})^2}$ and compute $(\hat{\alpha}_n^{\mathrm{hd-pd},(m)}, \hat{\beta}_n^{\mathrm{hd-pd},(m)})$ following (15) with $\hat{\omega}_{n,l}^{\mathrm{hd-pd},(m)}$ as the penalty loadings, and $\hat{\epsilon}_{n,j}^{\mathrm{hd-pd,(m)}} = \delta_{Y,j} - \hat{\alpha}_n^{\mathrm{hd-pd},(m)} - \delta'_{\psi,j}\hat{\beta}_n^{\mathrm{hd-pd,(m)}}$.

⋮

Step M: ...

Step $M+1$: Set $\hat{\beta}_n^{\mathrm{hd-pd}} = \hat{\beta}_n^{\mathrm{hd-pd,(M)}}$.

As suggested by Belloni et al. (2017), we set M to be 15. We note that `R` package **hdm** has a built-in option for this iterative procedure. For this choice of penalty loadings, arguments similar to those in Belloni et al. (2017) can be used to verify (16) under "matched pairs" designs. ■

5.2 Second LASSO-based Adjustment

For $d \in \{0,1\}$, define

$$(\hat{\alpha}^{\mathrm{hd}}_{d,n}, \hat{\beta}^{\mathrm{hd}}_{d,n}) \in \underset{a\in\mathbf{R}, b\in\mathbf{R}^{p_n}}{\operatorname{argmin}} \frac{1}{n} \sum_{1\le i\le 2n: D_i=d} (Y_i - a - \psi'_{n,i} b)^2 + \lambda^{\mathrm{hd}}_{d,n} \|\hat{\Omega}_n(d) b\|_1 , \tag{20}$$

where $\lambda^{\mathrm{hd}}_{d,n}$ is a penalty parameter that will be disciplined by the assumptions below, $\hat{\Omega}_n(d) = \mathrm{diag}(\hat{\omega}_1(d), \cdots, \hat{\omega}_{p_n}(d))$ is a diagonal matrix, and $\hat{\omega}_{n,l}(d)$ is the penalty loading for the lth regressor. Define $\Omega^*_n(d) = \mathrm{diag}(\omega^*_{n,1}(d), \cdots, \omega^*_{n,p_n}(d))$, where $\omega^{*,2}_{n,l}(d) = \mathrm{Var}[\psi_{n,i,l} v_i]$ and $v_i = Y_i - E[Y_i|X_i, W_i]$. For some $\underline{c}$ and $\bar{c}$, we require that

$$0 < \underline{c} \le \liminf_{n\to\infty} \min_{1\le l\le p_n} \hat{\omega}_{n,l}(d)/\omega^*_{n,l}(d) \le \limsup_{n\to\infty} \max_{1\le l\le p_n} \hat{\omega}_{n,l}(d)/\omega^*_{n,l}(d) \le \bar{c} < \infty \tag{21}$$

with probability one. Let $\hat{\Delta}^{\mathrm{hd}}_n$ denote the estimator in (4) with $\hat{m}_{d,n} = \psi'_{n,i}\hat{\beta}^{\mathrm{hd}}_{d,n}$ for $d \in \{0,1\}$.

Our analysis of this estimator will require the following assumptions. In our statement of the assumptions, we will make use of the quantity $\beta^{\mathrm{hd}}_{d,n}$, which will be assumed to satisfy

$$s^{\mathrm{hd}}_n = \max_{d\in\{0,1\}} \|\beta^{\mathrm{hd}}_{d,n}\|_0 \tag{22}$$

and

$$\|\Omega^*_n(d)^{-1} E[\psi_{n,i}\epsilon^{\mathrm{hd}}_{n,i}(d)]\|_\infty + |E\epsilon^{\mathrm{hd}}_{n,i}(d)| = o\left(\lambda^{\mathrm{hd}}_{d,n}\right) , \tag{23}$$

where

$$\epsilon^{\mathrm{hd}}_{n,i}(d) = Y_i(d) - \alpha^{\mathrm{hd}}_{d,n} - \psi'_{n,i}\beta^{\mathrm{hd}}_{d,n} .$$

Here, it is useful to recall the discussion after equations (17)–(18).

We now proceed with the statement of our assumptions. The first assumption collects a variety of moment conditions that will be used in our formal analysis:

Assumption 5.5. (a) For some $q > 2$ and constant C_1,

$$\sup_{n\ge1} \max_{1\le l\le p_n} E[|\psi^q_{n,i,l}||X_i] \le C_1$$

$$\sup_{n\ge1} |\psi'_{n,i}\beta^{\mathrm{hd-pd}}_{d,n}| \le C_1$$

$$\sup_{n\ge1} |E[Y_i(a)|X_i, W_{n,i}]| \le C_1$$

with probability one.

(b) For some c_0, $\underline{\sigma}$, $\bar{\sigma}$,

$$0 < \underline{\sigma}^2 \le \liminf_{n\to\infty} \min_{d\in\{0,1\}, 1\le l\le p_n} \omega^2_{n,l}(d) \le \limsup_{n\to\infty} \max_{d\in\{0,1\}, 1\le l\le p_n} \omega^2_{n,l}(d) \le \bar{\sigma}^2 < \infty .$$

Moreover, the following statements hold with probability one:

$$\sup_{n\geq 1}\max_{d\in\{0,1\}} E[(\psi_{n,i}'\beta_{d,n}^{\rm hd})^2]\leq c_0<\infty$$

$$\max_{d\in\{0,1\},1\leq l\leq p_n}\frac{1}{2n}\sum_{1\leq i\leq 2n}E[\epsilon_{n,i}^4(d)|X_i]\leq c_0<\infty$$

$$\sup_{n\geq 1}\max_{d\in\{0,1\}}E[\epsilon_{n,i}^4(d)]\leq c_0<\infty$$

$$\min_{d\in\{0,1\}}\mathrm{Var}[Y_i(d)-\psi_{n,i}'(\beta_{1,n}^{\rm hd}+\beta_{0,n}^{\rm hd})/2]\geq \underline{\sigma}^2>0$$

$$\min_{1\leq l\leq p_n}\frac{1}{n}\sum_{1\leq i\leq 2n}I\{D_i=d\}\,\mathrm{Var}[\psi_{n,i,l}\epsilon_{n,i}(d)|X_i]\geq\underline{\sigma}^2>0$$

$$\min_{1\leq l\leq p_n}\mathrm{Var}[E[\psi_{n,i,l}\epsilon_{n,i}(d)|X_i]]\geq\underline{\sigma}^2>0\ .$$

The discussion after Assumption 5.1 applies to the preceding assumption as well. Our analysis will, as before, also require some discipline on the way in which pairs are formed. For this purpose, Assumption 2.3 will suffice, but we will need an additional Lipshitz-like condition, similar to Assumption 5.2(b):

Assumption 5.6. For some $L>0$ and any x_1 and x_2 in the support of X_i, we have

$$|(\Psi(x_1)-\Psi(x_2))'\beta_{d,n}^{\rm hd}|\leq L||x_1-x_2||_2\ .$$

We next specify our restrictions on the penalty parameter $\lambda_n^{\rm hd}$.

Assumption 5.7. (a) For some $\ell\ell_n\to\infty$,

$$\lambda_{d,n}^{\rm hd}=\frac{\ell\ell_n}{\sqrt{n}}\Phi^{-1}\left(1-\frac{0.1}{2\log(n)p_n}\right)\ .$$

(b) $\Xi_n^2(\log p_n)^7/n\to 0$ and $(\ell\ell_n s_n^{\rm hd}\log p_n)/\sqrt{n}\to 0$.

We note that Assumption 5.7(b) permits p_n to be much greater than n.

Finally, as is common in the analysis of ℓ_1-penalized regression, we require a "restricted eigenvalue" condition. This assumption permits us to apply Bickel et al. (2009, Lemma 4.1) and establish the error bounds for $|\hat\alpha_{d,n}^{\rm hd}-\alpha_{d,n}^{\rm hd}|+||\hat\beta_{d,n}^{\rm hd}-\beta_{d,n}^{\rm hd}||_1$ and $\frac{1}{n}\sum_{1\leq i\leq 2n}I\{D_i=d\}\left(\hat\alpha_{d,n}^{\rm hd}-\alpha_{d,n}^{\rm hd}+\psi_{n,i}'(\hat\beta_{d,n}^{\rm hd}-\beta_{d,n}^{\rm hd})\right)^2$.

Assumption 5.8. For some $\kappa_1>0,\kappa_2$ and $\ell_n\to\infty$, the following statements hold with probability approaching one:

$$\inf_{d\in\{0,1\},v\in\mathbf{R}^{p_n+1}:\|v\|_0\leq(s_n^{\rm hd}+1)\ell_n}(\|v\|_2^2)^{-1}v'\left(\frac{1}{n}\sum_{1\leq i\leq 2n}I\{D_i=d\}\breve\psi_{n,i}\breve\psi_{n,i}'\right)v\geq\kappa_1$$

$$\sup_{d\in\{0,1\},v\in\mathbf{R}^{p_n+1}:\|v\|_0\leq(s_n^{\rm hd}+1)\ell_n}(\|v\|_2^2)^{-1}v'\left(\frac{1}{n}\sum_{1\leq i\leq 2n}I\{D_i=d\}\breve\psi_{n,i}\breve\psi_{n,i}'\right)v\leq\kappa_2$$

$$\inf_{d\in\{0,1\},v\in\mathbf{R}^{p_n+1}:\|v\|_0\leq(s_n^{\rm hd}+1)\ell_n}(\|v\|_2^2)^{-1}v'\left(\frac{1}{n}\sum_{1\leq i\leq 2n}I\{D_i=d\}E[\breve\psi_{n,i}\breve\psi_{n,i}'|X_i]\right)v\geq\kappa_1$$

$$\sup_{d\in\{0,1\},v\in\mathbf{R}^{p_n+1}:\|v\|_0\le(s_n^{\mathrm{hd}}+1)\ell_n}(\|v\|_2^2)^{-1}v'\left(\frac{1}{n}\sum_{1\le i\le 2n}I\{D_i=d\}E[\breve{\psi}_{n,i}\breve{\psi}_{n,i}'|X_i]\right)v\le\kappa_2\ ,$$

where $\breve{\psi}_{n,i}=(1,\psi_{n,i}')'$.

Using these assumptions, the following theorem characterizes the behavior of $\hat{\Delta}_n^{\mathrm{hd}}$:

Theorem 5.2. *Suppose Q satisfies Assumption 2.1 and the treatment assignment mechanism satisfies Assumptions 2.2–2.3. Further suppose Assumptions 5.5–5.8 and (21) hold. Then, (6), (9), and Assumption 3.1 are satisfied with $\hat{m}_{d,n}=\hat{\alpha}_{d,n}^{\mathrm{hd}}+\psi_{n,i}'\hat{\beta}_{d,n}^{\mathrm{hd}}$ and*

$$m_{d,n}(X_i,W_{n,i})=\alpha_{d,n}^{\mathrm{hd}}+\psi_{n,i}'\beta_{d,n}^{\mathrm{hd}}$$

for $d\in\{0,1\}$ and $n\ge 1$. Denote the variance of $\hat{\Delta}_n^{\mathrm{hd}}$ by $\sigma_n^{\mathrm{hd},2}$. If the LASSO adjustment is approximately correctly specified, i.e., $E[Y_i(d)|X_i,W_i]=\alpha_{d,n}^{\mathrm{hd}}+\psi_{n,i}'\beta_{d,n}^{\mathrm{hd}}+R_{n,i}(d)$ and $\max_{d\in\{0,1\}}E[R_{n,i}^2(d)]=o(1)$, then $\sigma_n^{\mathrm{hd},2}$ achieves the minimum variance, i.e.,

$$\lim_{n\to\infty}\sigma_n^{\mathrm{hd},2}=\sigma_2^2(Q)+\sigma_3^2(Q).$$

Remark 5.4. As in Remark 5.3, we recommend employing an iterative estimation procedure outlined by Belloni et al. (2017) to estimate $\hat{\beta}_{d,n}^{\mathrm{hd}}$, in which the m-th step's penalty loadings are estimated based the $(m-1)$-th step's LASSO estimates. Formally, this iterative procedure is described by the following algorithm:

Algorithm 5.2.

Step 0: Set $\hat{\epsilon}_{n,i}^{\mathrm{hd},(0)}(d)=Y_i$ if $D_i=d$.

$\vdots$

Step m: Compute $\hat{\omega}_{n,l}^{(m)}(d)=\sqrt{\frac{1}{n}\sum_{1\le i\le 2n}I\{D_i=d\}\psi_{n,i,l}^2(\hat{\epsilon}_{n,i}^{\mathrm{hd},(m-1)}(d))^2}$ and compute $(\hat{\alpha}_{d,n}^{\mathrm{hd},(m)},\hat{\beta}_{d,n}^{\mathrm{hd},(m)})$ following (20) with $\hat{\omega}_{n,l}^{(m)}$ as the penalty loadings, and $\hat{\epsilon}_{n,i}^{\mathrm{hd},(m)}(d)=Y_i-\hat{\alpha}_{d,n}^{\mathrm{hd},(m)}-\psi_i'\hat{\beta}_{d,n}^{\mathrm{hd},(m)}$ if $D_i=d$.

$\vdots$

Step M: ...

Step $M+1$: Set $\hat{\beta}_{d,n}^{\mathrm{hd}}=\hat{\beta}_{d,n}^{\mathrm{hd},(M)}$.

As suggested by Belloni et al. (2017), we set M to be 15. We note that R package **hdm** has a built-in option for this iterative procedure. For this choice of penalty loadings, arguments similar to those in Belloni et al. (2017) can be used to verify (21) under "matched pairs" designs. ■

When the LASSO adjustment is approximately correctly specified, Theorem 5.1 shows $\hat{\Delta}_n^{\mathrm{hd}}$ derived in Remark 3.1, and thus, is guaranteed to be weakly more efficient than the ATE estimator without any adjustments. On the other hand, when the LASSO adjustment is not approximately correctly specified, $\hat{\Delta}_n^{\mathrm{hd}}$ suffers from Freedman (2008)'s critique that it may be less efficient than $\hat{\Delta}_n^{\mathrm{unadj}}$. To overcome this problem, we consider an additional step in which we treat the LASSO adjustments $(\psi'_{n,i}\hat{\beta}^{\text{hd}}_{1,n}, \psi'_{n,i}\hat{\beta}^{\text{hd}}_{0,n})$ as linear covariates and rerun a linear regression with pair fixed effects. Such a procedure has also been studied by Cohen and Fogarty (2020) in the setting with low-dimensional covariates and complete randomization. Theorem 5.3 below shows the new estimator for the ATE is weakly more efficient than both $\hat{\Delta}^{\text{unadj}}_n$ and $\hat{\Delta}^{\text{hd}}_n$. To state the results, define $\Gamma_{n,i} = (\psi'_{n,i}\beta^{\text{hd}}_{1,n}, \psi'_{n,i}\beta^{\text{hd}}_{0,n})'$, $\hat{\Gamma}_{n,i} = (\psi'_{n,i}\hat{\beta}^{\text{hd}}_{1,n}, \psi'_{n,i}\hat{\beta}^{\text{hd}}_{0,n})$, and $\hat{\Delta}^{\text{hd}-\text{f}}_n$ as the estimator in (12) with ψ_i replaced by $\hat{\Gamma}_{n,i}$. Note that $\hat{\Delta}^{\text{hd}-\text{f}}_n$ remains numerically the same if we include the intercept $\hat{\alpha}^{\text{hd}}_{d,n}$ in the definition of $\hat{\Gamma}_{n,i}$. Following Remark 4.2, $\hat{\Delta}^{\text{hd}-\text{f}}_n$ is the intercept in the linear regression of $(D_{\pi(2j-1)} - D_{\pi(2j-1)})(Y_{\pi(2j-1)} - Y_{\pi(2j)})$ on constant and $(D_{\pi(2j-1)} - D_{\pi(2j-1)})(\hat{\Gamma}_{n,\pi(2j-1)} - \hat{\Gamma}_{n,\pi(2j)})$. Replacing $\hat{\Gamma}_{n,i}$ by $\hat{\Gamma}_{n,i} + (\hat{\alpha}^{\text{hd}}_{1,n}, \hat{\alpha}^{\text{hd}}_{0,n})'$ will not change the regression estimators.

The following assumption will be employed to control $\Gamma_{n,i}$ in our subsequent analysis:

Assumption 5.9. For some $\kappa_1 > 0$ and κ_2,

$$\inf_{n \geq 1} \inf_{v \in \mathbf{R}^2} ||v||_2^{-2} v' E[\text{Var}[\Gamma_{n,i}|X_i]]v \geq \kappa_1$$

$$\sup_{n \geq 1} \sup_{v \in \mathbf{R}^2} ||v||_2^{-2} v' E[\text{Var}[\Gamma_{n,i}|X_i]]v \leq \kappa_2 \ .$$

The following theorem characterizes the behavior of $\hat{\Delta}^{\text{hd}-\text{f}}_n$:

Theorem 5.3. *Suppose Q satisfies Assumption 2.1 and the treatment assignment mechanism satisfies Assumptions 2.2–2.3. Further suppose Assumptions 5.5–5.8 and (21) hold. In addition, suppose Assumption 5.9 holds. Then, (6), (9), and Assumption 3.1 are satisfied with $\hat{m}_{d,n}(X_i, W_{n,i}) = \hat{\Gamma}'_{n,i}\hat{\beta}^{\text{hd}-\text{f}}_n$ and*

$$m_{d,n}(X_i, W_{n,i}) = \Gamma'_{n,i}\beta^{\text{hd}-\text{f}}_n$$

for $d \in \{0, 1\}$ and $n \geq 1$, where $\beta^{\text{hd}-\text{f}}_n = (2E[\text{Var}[\Gamma_{n,i}|X_i]])^{-1}E[\text{Cov}[\Gamma_{n,i}, Y_i(1) + Y_i(0)|X_i]]$. In addition, denote the variance of $\hat{\Delta}^{\text{hd}-\text{f}}_n$ as $\sigma^{\text{hd}-\text{f},2}_n$. Then, $\sigma^{\text{na},2}_n \geq \sigma^{\text{hd}-\text{f},2}_n$ and $\sigma^{\text{hd},2}_n \geq \sigma^{\text{hd}-\text{f},2}_n$.

Remark 5.5. We briefly comment on the comparison between the two LASSO-based adjustment methods. First, when the LASSO adjustment is approximately correctly specified, then both methods produce the same adjustment asymptotically, which achieves the minimum variance. Second, when the pseudo-true values in the two methods are different, it is unclear which adjustment is more efficient. However, it is possible to use the regression adjustments obtained from both LASSO estimations as regressors in the refitting step in the second method and produce one regression-adjusted ATE estimator which is more efficient than both $\hat{\Delta}^{\text{hd}-\text{pd}}_n$ and $\hat{\Delta}^{\text{hd}-\text{f}}_n$, provided that the full rank condition in Assumption 5.9 holds. Third, the first method tends to select less regressors when the dimension of X_i is large, as its ℓ_1 penalty depends on ζ_n. Fourth, the ℓ_1 penalty of the second method is well studied in the literature. See, for example, Belloni et al. (2012), Belloni et al. (2014), and Belloni et al. (2017). Finally, it is possible to relax the full rank condition in Assumption 5.9 by running a ridge regression or truncating the minimum eigenvalue of the gram matrix in the refitting step in the second method, which is left for future work. ■

6 Simulations

In this section, we conduct Monte Carlo experiments to assess the finite-sample performance of the inference methods proposed in the paper. In all cases, we follow Bai et al. (2021) to consider tests of the hypothesis that

$$H_0 : \Delta(Q) = \Delta_0 \text{ versus } H_1 : \Delta(Q) \neq \Delta_0.$$

with $\Delta_0 = 0$ at nominal level $\alpha = 0.05$.

6.1 Data Generating Processes

We generate potential outcomes for $d \in \{0, 1\}$ and $1 \leq i \leq 2n$ by the equation

$$Y_i(d) = \mu_d + m_d(X_i, W_i) + \sigma_d(X_i, W_i)\epsilon_{d,i}, \ a = 0, 1, \tag{24}$$

where $\mu_d, m_d(X_i, W_i), \sigma_d(X_i, W_i)$, and $\epsilon_{d,i}$ are specified in each model as follows. In each of the specifications, $(X_i, W_i, \epsilon_{0,i}, \epsilon_{1,i})$ are i.i.d. across i. The number of pairs n is equal to 100 and 200 respectively. The number of replications is 10,000.

Model 1 $(X_i, W_i)^\top = (\Phi(V_{i1}), \Phi(V_{i2}))^\top$, where $\Phi(\cdot)$ is the standard normal distribution function and

$$V_i \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right),$$

$m_0(X_i, W_i) = \gamma\left(W_i - \frac{1}{2}\right)$; $m_1(X_i, W_i) = m_0(X_i, W_i)$; $\epsilon_{d,i} \sim N(0,1)$ for $a = 0, 1$; $\sigma_0(X_i, W_i) = \sigma_0 = 1$ and $\sigma_1(X_i, W_i) = \sigma_1$. We set $\gamma = 4$, $\sigma_1 = 1$, $\rho = 0.2$.

Model 2 $(X_i, W_i)^\top = (\Phi(V_{i1}), V_{1i}V_{i2})^\top$, where V_i is the same as in Model 1. $m_0(X_i, W_i) = m_1(X_i, W_i) = \gamma_1(W_i - \rho) + \gamma_2\left(\Phi^{-1}(X_i)^2 - 1\right)$. $\epsilon_{d,i} \sim N(0, 1)$ for $a = 0, 1$; $\sigma_0(X_i, W_i) = \sigma_0 = 1$ and $\sigma_1(X_i, W_i) = \sigma_1$. $(\gamma_1, \gamma_2)^\top = (1, 2)^\top$, $\sigma_1 = 1$, $\rho = 0.2$.

Model 3 The same as in Model 2, except that $m_0(X_i, W_i) = m_1(X_i, W_i) = \gamma_1(W_i - \rho) + \gamma_2\left(\Phi(W_i) - \frac{1}{2}\right) + \gamma_3\left(\Phi^{-1}(X_i)^2 - 1\right)$ with $(\gamma_1, \gamma_2, \gamma_3)^\top = \left(\frac{1}{4}, 1, 2\right)^\top$.

Model 4 $(X_i, W_i)^\top = (V_{i1}, V_{1i}V_{i2})^\top$, where V_i is the same as in Model 1. $m_0(X_i, W_i) = m_1(X_i, W_i) = \gamma_1(W_i - \rho) + \gamma_2\left(\Phi(W_i) - \frac{1}{2}\right) + \gamma_3\left(X_i^2 - 1\right)$. $\epsilon_{d,i} \sim N(0, 1)$ for $a = 0, 1$; $\sigma_0(X_i, W_i) = \sigma_0 = 1$ and $\sigma_1(X_i, W_i) = \sigma_1$. $(\gamma_1, \gamma_2, \gamma_3)^\top = (2, 1, 2)^\top$.

Model 5 The same as in Model 4, except that $m_1(X_i, W_i) = m_0(X_i, W_i) + \left(\Phi(X_i) - \frac{1}{2}\right)$.

Model 6 The same as in Model 5, except that $\sigma_0(X_i, W_i) = (\Phi(X_i) + 0.5)$ and $\sigma_1(X_i, W_i) = (\Phi(X_i) + 0.5)\sigma_1$.

Model 7 $X_i = (V_{i1}, V_{i2})^\top$ and $W_i = (V_{i1}V_{i3}, V_{i2}V_{i4})^\top$, where $V_i \sim N(0, \Sigma)$ with $\dim(V_i) = 4$ and Σ consisting of 1 on the diagonal and ρ on all other elements. $m_0(X_i, W_i) = m_1(X_i, W_i) = \gamma_1'(W_i - \rho) +$

$\gamma_2' \left(\Phi\left(W_i\right)-\frac{1}{2}\right)+\gamma_3\left(X_{i1}^2-1\right)$ with $\gamma_1=(2,2)^{\top}, \gamma_2=(1,1)^{\top}, \gamma_3=1$. $\epsilon_{d,i} \sim N(0,1)$ for $a=0,1$; $\sigma_0\left(X_i, W_i\right)=\sigma_0=1$ and $\sigma_1\left(X_i, W_i\right)=\sigma_1$. $\sigma_1=1$, $\rho=0.2$.

Model 8 The same as in Model 7, except that $m_1\left(X_i, W_i\right)=m_0\left(X_i, W_i\right)+\left(\Phi\left(X_{i1}\right)-\frac{1}{2}\right)$.

Model 9 The same as in Model 8, except that $\sigma_0\left(X_i, W_i\right)=\left(\Phi\left(X_{i1}\right)+0.5\right)$ and $\sigma_1\left(X_i, W_i\right)=\left(\Phi\left(X_{i1}\right)+0.5\right)\sigma_1$.

Model 10 $X_i=\left(\Phi\left(V_{i1}\right), \cdots, \Phi\left(V_{i4}\right)\right)^{\top}$ and $W_i=\left(V_{i1}V_{i5}, V_{i2}V_{i6}\right)^{\top}$, where $V_i \sim N(0, \Sigma)$ with $\dim(V_i)=6$ and Σ consisting of 1 on the diagonal and ρ on all other elements. $m_0\left(X_i, W_i\right)=m_1\left(X_i, W_i\right)=\gamma_1'\left(W_i-\rho\right)+\gamma_2'\left(\Phi\left(W_i\right)-\frac{1}{2}\right)+\gamma_3'\left(\left(\Phi^{-1}\left(X_{i1}\right)^2, \Phi^{-1}\left(X_{i2}\right)^2\right)^{\top}-1\right)$ with $\gamma_1=(1,1)^{\top}, \gamma_2=\left(\frac{1}{2}, \frac{1}{2}\right)^{\top}, \gamma_3=\left(\frac{1}{2}, \frac{1}{2}\right)^{\top}$. $\epsilon_{d,i} \sim N(0,1)$ for $a=0,1$; $\sigma_0\left(X_i, W_i\right)=\sigma_0=1$ and $\sigma_1\left(X_i, W_i\right)=\sigma_1$. $\sigma_1=1$, $\rho=0.2$.

Model 11 The same as in Model 10, except that $m_1\left(X_i, W_i\right)=m_0\left(X_i, W_i\right)+\frac{1}{4}\sum_{j=1}^4\left(X_{ij}-\frac{1}{2}\right)$.

Model 12 $X_i=\left(\Phi\left(V_{i1}\right), \cdots, \Phi\left(V_{i4}\right)\right)^{\top}$ and $W_i=\left(V_{i1}V_{i41}, \cdots, V_{i40}V_{i80}\right)^{\top}$, where $V_i \sim N(0, \Sigma)$ with $\dim(V_i)=80$. Σ is the Toeplitz matrix

$$\Sigma=\begin{pmatrix} 1 & 0.5 & 0.5^2 & \cdots & 0.5^{79} \\ 0.5 & 1 & 0.5 & \cdots & 0.5^{78} \\ 0.5^2 & 0.5 & 1 & \cdots & 0.5^{77} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0.5^{79} & 0.5^{78} & 0.5^{77} & \cdots & 1 \end{pmatrix}.$$

$m_0\left(X_i, W_i\right)=m_1\left(X_i, W_i\right)=\gamma_1'\left(W_i-\frac{1}{2}\right)+\gamma_2'\left(\Phi^{-1}\left(X_i\right)^2-1\right)$, $\gamma_1=\left(\frac{1}{1^2}, \frac{1}{2^2}, \cdots, \frac{1}{40^2}\right)^{\top}$ with $\dim(\gamma_1)=40$, and $\gamma_2=\left(\frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}\right)^{\top}$ with $\dim(\gamma_2)=4$. $\epsilon_{d,i} \sim N(0,1)$ for $a=0,1$; $\sigma_0\left(X_i, W_i\right)=\sigma_0=1$ and $\sigma_1\left(X_i, W_i\right)=\sigma_1$ with $\sigma_1=1$

Model 13 The same as in Model 12, except that $m_0\left(X_i, W_i\right)=m_1\left(X_i, W_i\right)=\gamma_1'\left(W_i-\rho\right)+\gamma_2'\left(\Phi\left(W_i\right)-\frac{1}{2}\right)+\gamma_3'\left(\Phi^{-1}\left(X_i\right)^2-1\right)$, $\gamma_1=\left(\frac{1}{1^2}, \cdots, \frac{1}{40^2}\right)^{\top}$, $\gamma_2=\frac{1}{8}\left(\frac{1}{1^2}, \cdots, \frac{1}{40^2}\right)^{\top}$, and $\gamma_3=\left(\frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}\right)^{\top}$ with $\dim(\gamma_1)=\dim(\gamma_2)=40$ and $\dim(\gamma_3)=4$.

Model 14 The same as in Model 13, except that $m_1\left(X_i, W_i\right)=m_0\left(X_i, W_i\right)+\sum_{j=1}^4 \frac{1}{j^2}\left(X_{ij}-\frac{1}{2}\right)$.

Model 15 The same as in Model 14, except that $\sigma_0\left(X_i, W_i\right)=\left(X_{i1}+0.5\right)$ and $\sigma_1\left(X_i, W_i\right)=\left(X_{i1}+0.5\right)\sigma_1$.

It is worth noting that Models 1, 2, 3, 4, 7, 10, 12, and 13 imply homogeneous treatment effects because $m_1\left(X_i, W_i\right)=m_0\left(X_i, W_i\right)$. Among them, $E[Y_i(a)|X_i, W_i]-E[Y_i(a)|X_i]$ is linear in W_i in Models 1, 2, and 12. Models 5, 8, 11, and 14 have heterogeneous but homoscedastic treatment effects. In Models 6, 9, and 15, however, the implied treatment effects are both heterogeneous and heteroscedastic. Models 12-15 contain high-dimensional covariates.

We follow Bai et al. (2021) to match pairs. Specifically, if $\dim\left(X_i\right)=1$, we match pairs by sorting $X_i, i=1, \ldots, 2n$. If $\dim\left(X_i\right)>1$, we match pairs by the permutation π calculated using the `R` package *nbpMatching*. For more details, see Bai et al. (2021, Section 4). After matching the pairs, we flip coins to randomly select one unit within each pair for treatment and another for control.

6.2 Estimation and Inference

We set $\mu_0 = 0$ and $\mu_1 = \Delta$, where $\Delta = 0$ and $1/4$ are used to illustrate the size and power, respectively. Rejection probabilities in percentage points are presented. To further illustrate the efficiency gains obtained by regression adjustments, in Figure 1, we plot the average standard error reduction in percentage relative to the standard error of the estimator without adjustments for various estimation methods.

Specifically, we consider the following adjusted estimators.

(i) NA: the estimator with no adjustments. In this case, our standard error is identical to the adjusted standard error proposed by Bai et al. (2021).

(ii) LA: the linear adjustments with regressors W_i but without pair dummies.

(iii) LA2: the linear adjustments with X_i and W_i regressors but without pair dummies.

(iv) LDA: the linear adjustments with regressors W_i and pair dummies.

(v) HD-PD: the first LASSO-based adjustment.

(vi) HD-F: the second LASSO-based adjustment.

See Section C for the regressors used in the LASSO adjustments.

For Models 1-11, we examine the performance of estimators (i)-(v). For Models 12-15, we assess the performance between estimators (i) and (v) in high-dimensional settings. Note that the adjustments are misspecified for almost all the models. The only exception is Model 1, for which the linear adjustment in W_i is correctly specified because $m_d(X_i, W_i)$ is just a linear function of W_i.

6.3 Simulation Results

Tables 1 and 3 report size at the 0.05 level and power of the different methods for Models 1–11 when n is 100 and 200, respectively. Several patterns emerge. First, for all the estimators, the rejection rates under H_0 are close to the nominal level even when $n = 100$ and with misspecified adjustments. This result is expected because all the estimators take into account the dependence structure arising in MPDs, consistent with the findings in Bai et al. (2021).

Second, in terms of power, "LDA" is higher than "NA", "LA", and "LA2" for all eleven models, as predicted by our theory. This finding confirms that "LDA" is the optimal linear adjustment and will not degrade the precision of the ATE estimator. In contrast, we observe that "LA" and "LA2" in Model 3 are even less powerful than the unadjusted estimator "NA." Figure 1 further confirms that these two methods inflate the estimation standard error. This result echos Freedman's critique (Freedman, 2008) that careless regression adjustments may degrade the estimation precision. Our "LDA" addresses this issue because it has been proven to be weakly more efficient than the unadjusted estimator.

Third, the improvement of power for "LDA" is mainly due to the reduction of estimation standard errors, which can be more than 50% as shown in Figure 1 for Models 4–9. This means that the length of the confidence interval of the "LDA" estimator is just half of that for the "NA" estimator. Note the standard error of the "NA" estimator is the one proposed by Bai et al. (2021), which has already been adjusted to account for the cross-sectional dependence created in pair matching. The extra 50% reduction is therefore produced purely by the regression adjustment. For Models 10-11, the reduction of standard error achieved by "LDA" is more than 40% as well. For Model 1, the correct specification in the adjustments leads to all three methods achieving the global minimum asymptotic variance and maximum power. For Model 2, $m_d(X_i, W_i) - E[m_d(X_i, W_i)|X_i] = \gamma(W_i - E[W_i|X_i])$ so that the linear adjustment γW_i satisfies the conditions in Theorem 3.1. Therefore, "LDA", as the best linear adjustment, is also the best adjustment globally, achieving the global minimum asymptotic variance and maximum power. In contrast, "LA" and "LA2" are not the best linear adjustment and therefore less powerful than "LDA" because of the omitted pair dummies.

Finally, the LASSO-based adjustments have the best power for most models as they automatically achieve the global minimum asymptotic variance. Compared to "HD-PD", "HD-F" has slightly better power.

Tables 2 and 4 report the size and power for LASSO-based adjustments when both W_i and X_i are high-dimensional. We see that the size under the null is close to the nominal 5% while the power for the adjusted estimator is higher than the unadjusted one. Figure 1 further illustrates the reduction of the standard error is more than 30% for all high-dimensional models.

Table 1: Rejection probabilities for Models 1-11 when $n = 100$

	H_0: $\Delta = 0$						H_1: $\Delta = 1/4$					
Model	NA	LA	LA2	LDA	HD-PD	HD-F	NA	LA	LA2	LDA	HD-PD	HD-F
1	5.47	5.57	5.63	5.76	6.12	5.84	22.48	43.89	43.95	43.91	44.69	43.92
2	4.96	5.26	5.30	5.47	5.74	5.32	23.32	28.02	27.96	37.21	39.00	33.12
3	4.99	5.28	5.24	5.48	5.78	5.27	32.19	27.88	27.96	37.34	38.59	36.29
4	5.31	5.28	5.28	5.48	5.93	5.79	11.78	27.88	28.03	37.34	42.21	43.28
5	5.43	5.09	5.08	5.49	5.84	5.78	11.87	27.72	27.88	36.69	41.24	43.08
6	5.28	5.43	5.41	5.58	5.90	5.79	11.78	26.67	26.72	34.71	38.76	40.29
7	5.64	5.63	5.62	5.98	6.45	6.04	9.24	34.55	34.65	37.96	37.72	42.08
8	5.63	5.54	5.51	6.03	6.26	6.17	9.28	34.11	34.42	37.22	36.78	41.29
9	5.74	5.69	5.76	6.19	6.32	5.89	8.99	32.39	32.30	35.42	34.66	38.75
10	5.24	5.78	5.73	6.05	6.07	6.04	14.27	30.80	30.75	32.02	28.37	32.51
11	5.19	5.78	5.72	6.07	6.01	5.95	14.36	30.60	30.49	32.21	27.92	32.81

Table 2: Rejection probabilities for Models 12-15 when $n = 100$

	H_0: $\Delta = 0$			H_1: $\Delta = 1/4$		
	NA	HD-PD	HD-F	NA	HD-PD	HD-F
12	5.35	6.15	6.12	22.01	39.59	42.56
13	5.31	6.21	6.11	21.47	39.62	42.47
14	5.24	6.04	6.07	21.39	38.11	41.14
15	5.31	6.05	6.23	20.73	35.90	38.67

Table 3: Rejection probabilities for Models 1-11 when $n = 200$

	H_0: $\Delta = 0$						H_1: $\Delta = 1/4$					
Model	NA	LA	LA2	LDA	HD-PD	HD-F	NA	LA	LA2	LDA	HD-PD	HD-F
1	5.08	5.04	5.10	5.21	5.38	5.31	38.94	70.35	70.36	70.32	70.53	70.30
2	5.69	5.28	5.28	5.24	5.42	5.40	40.31	49.25	49.32	65.36	65.71	57.87
3	5.44	5.29	5.30	5.35	5.60	5.41	56.89	49.43	49.51	64.96	65.34	62.42
4	5.45	5.29	5.29	5.35	5.42	5.20	18.55	49.43	49.67	64.96	67.93	69.96
5	5.45	5.24	5.18	5.19	5.44	5.29	18.41	48.65	48.80	64.11	66.83	69.09
6	5.62	5.32	5.31	5.35	5.50	5.43	18.19	46.71	46.67	61.09	63.95	65.98
7	5.24	5.51	5.46	5.34	5.78	5.49	11.86	60.73	60.63	65.14	64.88	69.24
8	5.23	5.49	5.47	5.35	6.00	5.65	11.84	60.00	60.10	64.93	64.02	68.02
9	5.30	5.58	5.57	5.66	5.73	5.81	11.90	57.25	57.28	61.61	60.98	64.88
10	5.34	5.19	5.15	5.25	5.33	5.31	23.95	55.49	55.44	56.64	52.05	56.43
11	5.41	5.36	5.32	5.34	5.53	5.41	23.88	55.01	55.05	56.31	51.87	56.18

Table 4: Rejection probabilities for Models 12-15 when $n = 200$

	H_0: $\Delta = 0$			H_1: $\Delta = 1/4$		
	NA	HD-PD	HD-F	NA	HD-PD	HD-F
12	4.97	5.22	5.22	38.91	65.28	68.10
13	4.95	5.24	5.19	38.04	65.29	68.06
14	5.01	5.20	5.24	37.65	63.92	66.69
15	5.15	5.27	5.40	36.61	61.11	63.79

7 Empirical Illustration

In this section, we revisit the randomized experiment with a matched pairs design conducted in Groh and McKenzie (2016). In the paper, they examined the impact of macroinsurance on microenterprises. Here, we apply the covariate adjustment methods developed in this paper to their data and investigate the average effect of macroinsurance on three outcome variables: the microenterprise owners' loan renewal, their firms' monthly profits, and revenues.

The subjects in the experiment are microenterprise owners, who were the clients of the largest microfinance institution in Egypt. In the randomization, after an exact match of gender and the institution's branch code, those clients were grouped into pairs by applying an optimal greedy algorithm to additional 13 matching variables. Within each pair, a macroinsurance product was then offered to one randomly assigned client, and the other acted as a control. Based on the pair identities and all the matching variables, we re-order the pairs in our sample according to the procedure described in Section 5.1 of Jiang et al. (2022). The resulting sample contains 2824 microenterprise owners, that's, 1412 pairs of them.[1]

Table 5 reports the ATEs with the standard errors (in parentheses) estimated by different methods. Among them, "GM" corresponds to the method used in Groh and McKenzie (2016).[2] The description

[1]See Groh and McKenzie (2016) and Jiang et al. (2022) for more details.

[2]Groh and McKenzie (2016) estimated the effect by regression with regressors including some baseline covariates and dummies for the pairs. Specifically, for loan renewal, the regressors include a variable "high chance of renewing loan" and its interaction with treatment status. For the other two outcome variables, the regressor is the baseline value for the outcome of interest.

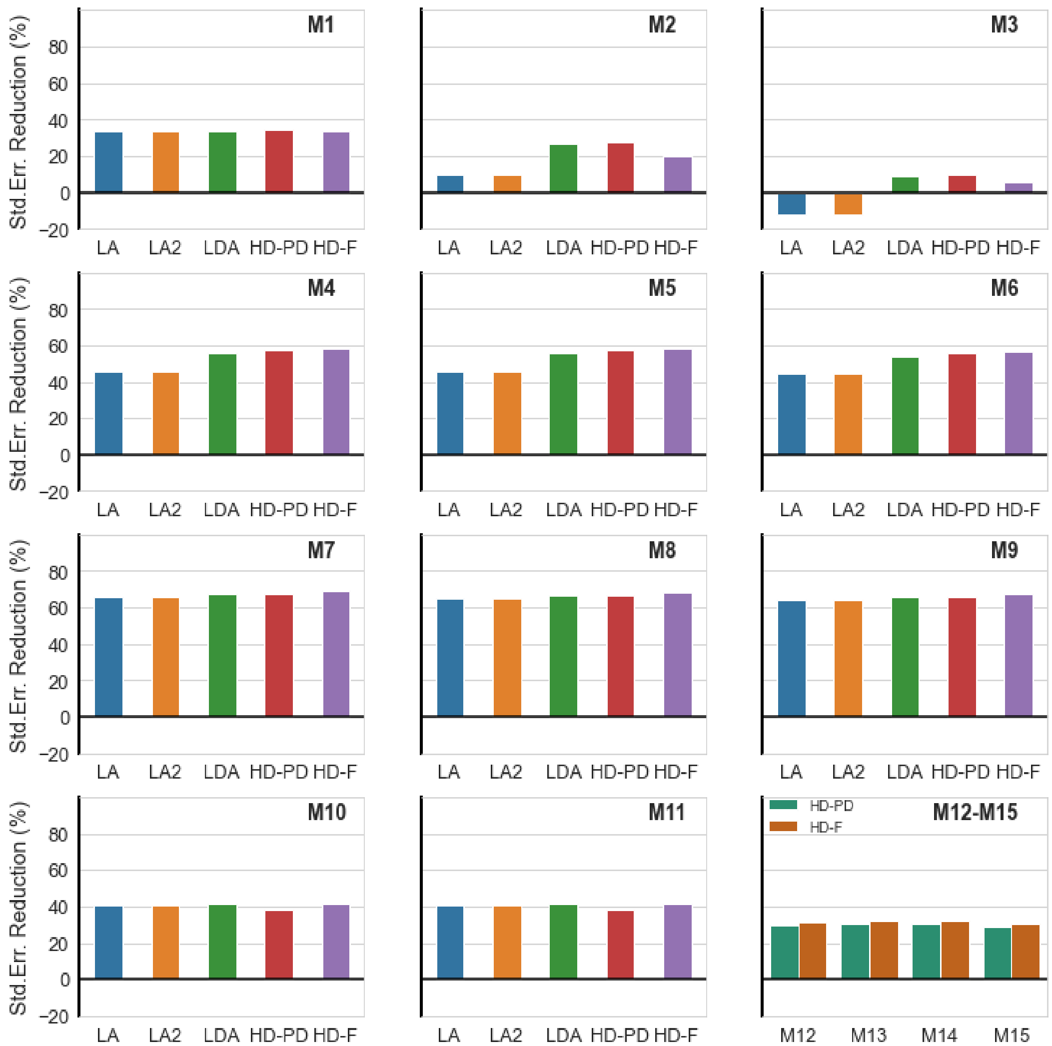

Figure 1: Average Standard Error Reduction in Percentage under H_1 when $n = 200$

Notes: The figure plots average standard error reduction in percentage achieved by regression adjustments relative to "NA" under H_1 for Models 1-15 when $n = 200$.

of other methods is similar to that in Section 6.2.[3] The results in this table prompt the following four observations.

First, in line with the theoretical and simulation results, the standard errors for the covariate-adjusted

The standard errors for the "GM" ATE estimate are calculated by the usual heteroskedastity-consistent estimator. The "GM" results in Table 5 were obtained by applying the Stata code provided by Groh and McKenzie (2016).

[3]To maintain comparability, we keep X_i and W_i the same in all the adjustments for each outcome variable. Specifically,

(i) X_i include gender and 13 additional matching variables for all the adjustments. Three of the matching variables are continuous and others are dummies.

(ii) For loan renewal, W_i include baseline value of loan amount, high chance of renewing loan, the interaction between the high chance of renewing loan and treatment status, and the interaction of these three variables with three continuous variables and the first three discrete variables in X_i. For the other two outcome variables, W_i only includes the baseline value for the outcome of interest and its interaction with three continuous variables and the first three discrete variables in X_i. All the continuous variables in X_i, the baseline values of loan amount, and the baseline value for the other three outcome variables are standardized at first when the regression-adjusted estimators are used.

ATEs are generally lower than those for the ATE estimate without adjustment. This observation holds for almost all the outcome variables and adjustment methods. For example, when the outcome variable is revenue, the standard errors for the covariate-adjusted ATE estimates are at least 9.7% less than that for the ATE estimate without adjustment.

Second, the standard errors for the ATE estimates obtained by the "GM" method are mostly higher than those for the ATE estimates obtained by the covariate adjustments. Especially, when the outcome variable is loan renewal, the standard errors for the "GM" ATE estimates are at least 16.7% higher than those for all other estimates. This observation may imply that the "GM" method is not the most efficient way to estimate the ATE of macrofinance on loan renewal.

Third, the size of the standard errors is mostly similar for all the covariate-adjustment ATEs. Among them, the standard errors for the "LA2" and "LDA" estimates are slightly less than those for the other regression-adjusted estimates.

Finally, between the two LASSO-based adjustments, "HD-F" achieves the smaller size of the standard errors. Surprisingly, "HD-PD" has the same estimates as "NA", which means it selects none of the variables in the adjustments. This result is caused by using a large rule-of-thumb penalty. There are more than 10 matching variables in this application, which leads to low matching quality and then produces a large penalty for the adjustments.

Table 5: Impacts of Macronsurance for Microenterprises

Y	n	NA	GM	LA	LA2	LDA	HD-PD	HD-F
Loan	1350	-0.007	0.004	-0.004	-0.006	0.006	-0.007	-0.003
renewal		(0.0180)	(0.0212)	(0.0178)	(0.0177)	(0.0177)	(0.0180)	(0.0177)
Profits	1322	-85.6	-50.9	-35.6	-46.8	-40.6	-85.6	-55.1
		(49.4)	(46.4)	(45.7)	(45.3)	(45.6)	(49.4)	(45.7)
Revenue	1318	-838.6	-657.6	-666.8	-664.7	-671.3	-838.6	-590.1
		(319.0)	(283.4)	(283.5)	(279.8)	(281.4)	(319.0)	(285.2)

Notes: The table reports the ATE estimates of the effect of macroinsurance for microenterprises. Standard errors are in parentheses.

A Proofs of Main Results

In the appendix, we use $a_n \lesssim b_n$ to denote there exists $c > 0$ such that $a_n \leq c b_n$.

A.1 Proof of Theorem 3.1

To begin, note

$$\begin{aligned}
\hat{\mu}_n(1) &= \frac{1}{2n} \sum_{1 \leq i \leq 2n} (2D_i(Y_i(1) - \hat{m}_{1,n}(X_i, W_i)) + \hat{m}_{1,n}(X_i, W_i)) \\
&= \frac{1}{2n} \sum_{1 \leq i \leq 2n} (2D_i Y_i(1) - (2D_i - 1)\hat{m}_{1,n}(X_i, W_i)) \\
&= \frac{1}{2n} \sum_{1 \leq i \leq 2n} (2D_i Y_i(1) - (2D_i - 1)m_{1,n}(X_i, W_i)) + o_P(n^{-1/2}) \\
&= \frac{1}{2n} \sum_{1 \leq i \leq 2n} (2D_i Y_i(1) - D_i m_{1,n}(X_i, W_i) - (1 - D_i)m_{1,n}(X_i, W_i)) + o_P(n^{-1/2}) ,
\end{aligned} \tag{25}$$

where the third equality follows from (6). Similarly,

$$\hat{\mu}_n(0) = \frac{1}{2n} \sum_{1 \leq i \leq 2n} (2(1 - D_i)Y_i(0) - D_i m_{0,n}(X_i, W_i) - (1 - D_i)m_{0,n}(X_i, W_i)) + o_P(n^{-1/2}) . \tag{26}$$

It follows from (25)–(26) that

$$\hat{\Delta}_n = \frac{1}{n} \sum_{1 \leq i \leq 2n} D_i \phi_{1,n,i} - \frac{1}{n} \sum_{1 \leq i \leq 2n} (1 - D_i)\phi_{0,n,i} + o_P(n^{-1/2}) , \tag{27}$$

where

$$\begin{aligned}
\phi_{1,n,i} &= Y_i(1) - \frac{1}{2}(m_{1,n}(X_i, W_i) + m_{0,n}(X_i, W_i)) \\
\phi_{0,n,i} &= Y_i(0) - \frac{1}{2}(m_{1,n}(X_i, W_i) + m_{0,n}(X_i, W_i)) .
\end{aligned}$$

Next, consider

$$\mathbb{L}_n = \frac{1}{2\sqrt{n}} \sum_{1 \leq i \leq 2n} (2D_i - 1)E[m_{1,n}(X_i, W_i) + m_{0,n}(X_i, W_i)|X_i] .$$

For simplicity, define $M_{d,n}(X_i) = E[m_{d,n}(X_i, W_i)|X_i]$ for $d \in \{0, 1\}$. It follows from Assumption 2.2 that $E[\mathbb{L}_n|X^{(n)}] = 0$. On the other hand,

$$\begin{aligned}
\text{Var}[\mathbb{L}_n|X^{(n)}] &= \frac{1}{4n} \sum_{1 \leq j \leq n} \big(M_{1,n}(X_{\pi(2j-1)}) + M_{0,n}(X_{\pi(2j-1)}) - (M_{1,n}(X_{\pi(2j)}) + M_{0,n}(X_{\pi(2j)}))\big)^2 \\
&\lesssim \frac{1}{n} \sum_{1 \leq j \leq n} |M_{1,n}(X_{\pi(2j-1)}) - M_{1,n}(X_{\pi(2j)})|^2 + \frac{1}{n} \sum_{1 \leq j \leq n} |M_{0,n}(X_{\pi(2j-1)}) - M_{0,n}(X_{\pi(2j)})|^2 \\
&\xrightarrow{P} 0 ,
\end{aligned}$$

where the inequality follows from $(a+b)^2 \leq 2(a^2+b^2)$ and the convergence follows from Assumptions 2.3 and 3.1(c). By Markov's inequality and the fact that $E[\mathbb{L}_n|X^{(n)}] = 0$, for any $\epsilon > 0$,

$$P\{|\mathbb{L}_n| > \epsilon | X^{(n)}\} \leq \frac{\mathrm{Var}[\mathbb{L}_n|X^{(n)}]}{\epsilon^2} \xrightarrow{P} 0 \ .$$

Since probabilities are bounded, we have $\mathbb{L}_n = o_P(1)$. This fact, together with (27), imply

$$\sqrt{n}(\hat{\Delta}_n - \Delta(Q)) = A_n - B_n + C_n - D_n \ ,$$

where

$$\begin{aligned}
A_n &= \frac{1}{\sqrt{n}} \sum_{1 \leq i \leq 2n} \Big(D_i \phi_{1,n,i} - E[D_i \phi_{1,n,i} | X^{(n)}, D^{(n)}] \Big) \\
B_n &= \frac{1}{\sqrt{n}} \sum_{1 \leq i \leq 2n} \Big((1-D_i) \phi_{0,n,i} - E[(1-D_i) \phi_{0,n,i} | X^{(n)}, D^{(n)}] \Big) \\
C_n &= \frac{1}{\sqrt{n}} \sum_{1 \leq i \leq 2n} D_i (E[Y_i(1)|X_i] - E[Y_i(1)]) \\
D_n &= \frac{1}{\sqrt{n}} \sum_{1 \leq i \leq 2n} (1-D_i)(E[Y_i(0)|X_i] - E[Y_i(0)]) \ .
\end{aligned}$$

Note that conditional on $X^{(n)}$ and $D^{(n)}$, A_n and B_n are independent while C_n and D_n are constants.

We first analyze the limiting behavior of A_n. Define

$$s_n^2 = \sum_{1 \leq i \leq 2n} D_i \, \mathrm{Var}[\phi_{1,n,i}|X_i] \ .$$

Note by Assumption 2.2 that $s_n^2 = n \, \mathrm{Var}[A_n|X^{(n)}, D^{(n)}]$. We proceed verify the Lindeberg condition for A_n conditional on $X^{(n)}$ and $D^{(n)}$, i.e., we show that for every $\epsilon > 0$,

$$\frac{1}{s_n^2} \sum_{1 \leq i \leq 2n} E[|D_i(\phi_{1,n,i} - E[\phi_{1,n,i}|X_i])|^2 I\{|D_i(\phi_{1,n,i} - E[\phi_{1,n,i}|X_i])| > \epsilon s_n\} | X^{(n)}, D^{(n)}] \xrightarrow{P} 0 \ . \tag{28}$$

To that end, first note Lemma B.2 implies

$$\frac{s_n^2}{nE[\mathrm{Var}[\phi_{1,n,i}|X_i]]} \xrightarrow{P} 1 \ . \tag{29}$$

(29) and Assumption 3.1(a) imply that for all $\lambda > 0$,

$$P\{\epsilon s_n > \lambda\} \xrightarrow{P} 1 \ . \tag{30}$$

Furthermore, for some $c > 0$,

$$P\left\{ \frac{s_n^2}{n} > c \right\} \to 1 \ . \tag{31}$$

Next, note for any $\lambda > 0$ and $\delta_1 > 0$, the left-hand side of (28) can be written as

$$
\begin{aligned}
&\frac{1}{s_n^2/n}\frac{1}{n}\sum_{1\le i\le 2n: D_i=1} E[|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]|^2 I\{|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]| > \epsilon s_n\}|X^{(n)}, D^{(n)}] \\
&\le \frac{1}{s_n^2/n}\frac{1}{n}\sum_{1\le i\le 2n} E[|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]|^2 I\{|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]| > \epsilon s_n\}|X^{(n)}, D^{(n)}] \\
&\le \frac{1}{c}\frac{1}{n}\sum_{1\le i\le 2n} E[|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]|^2 I\{|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]| > \lambda\}|X^{(n)}, D^{(n)}] + o_P(1) \\
&\le \frac{2}{c}\frac{1}{2n}\sum_{1\le i\le 2n} E[|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]|^2 I\{|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]| > \lambda\}|X_i] + o_P(1) \ , \qquad (32)
\end{aligned}
$$

where the first inequality follows by inspection, the second follows from (30)–(31), and the last follows from Assumption 2.2. We then argue

$$
\begin{aligned}
&\frac{1}{2n}\sum_{1\le i\le 2n} E[|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]|^2 I\{|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]| > \lambda\}|X_i] \\
&\qquad = E[|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]|^2 I\{|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]| > \lambda\}] + o_P(1) \ . \qquad (33)
\end{aligned}
$$

To this end, we once again verify the Lindeberg condition in Lemma 11.4.2 of Lehmann and Romano (2005). Note

$$
|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]|^2 I\{|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]| > \lambda\} \le |\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]|^2 \ .
$$

Therefore, in light of Lemma B.1, we only need to verify

$$
\lim_{\gamma\to\infty}\limsup_{n\to\infty} E[|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]|^2 I\{|\phi_{1,n,i} - E[\phi_{1,n,i}|X_i]|^2 > \gamma\}] = 0 \ , \qquad (34)
$$

which follows immediately from Lemma B.3.

Another application of (34) implies (28). Lindeberg's central limit theorem and (29) then imply that

$$
\sup_{t\in\mathbf{R}} |P\{A_n/\sqrt{E[\mathrm{Var}[\phi_{1,n,i}|X_i]]} \le t|X^{(n)}, D^{(n)}\} - \Phi(t)| \xrightarrow{P} 0 \ .
$$

Similar arguments lead to

$$
\sup_{t\in\mathbf{R}} |P\{B_n/\sqrt{E[\mathrm{Var}[\phi_{0,n,i}|X_i]]} \le t|X^{(n)}, D^{(n)}\} - \Phi(t)| \xrightarrow{P} 0 \ .
$$

Meanwhile, it follows from the same arguments as those in (S.22)–(S.25) of Bai et al. (2021) that

$$
C_n - D_n \xrightarrow{d} N\left(0, \frac{1}{2}E\left[(E[Y_i(1)|X_i] - E[Y_i(1)] - (E[Y_i(0)|X_i] - E[Y_i(0)]))^2\right]\right) \ .
$$

To establish (7), define $\nu_n^2 = \nu_{1,n}^2 + \nu_{0,n}^2 + \nu_2^2$, where

$$
\begin{aligned}
\nu_{1,n}^2 &= E[\mathrm{Var}[\phi_{1,n,i}|X_i]] \\
\nu_{0,n}^2 &= E[\mathrm{Var}[\phi_{0,n,i}|X_i]]
\end{aligned}
$$

$$\nu^2 = \frac{1}{2} E\left[(E[Y_i(1)|X_i] - E[Y_i(1)] - (E[Y_i(0)|X_i] - E[Y_i(0)]))^2\right]$$

Note

$$\frac{\sqrt{n}(\hat{\Delta}_n - \Delta(Q))}{\nu_n} = \frac{A_n}{\nu_{1,n}} \frac{\nu_{1,n}}{\nu_n} - \frac{B_n}{\nu_{0,n}} \frac{\nu_{0,n}}{\nu_n} + \frac{C_n - D_n}{\nu_2} \frac{\nu_2}{\nu_n} .$$

Further note $\nu_n, \nu_{1,n}, \nu_{0,n}, \nu_2$ are all constants conditional on $X^{(n)}$ and $D^{(n)}$. Suppose by contradiction that $\frac{\sqrt{n}(\hat{\Delta}_n - \Delta(Q))}{\nu_n}$ does not converge in distribution to $N(0,1)$. Then, there exists $\epsilon > 0$ and a subsequence $\{n_k\}$ such that

$$\sup_{t \in \mathbf{R}} |P\{\sqrt{n}_k(\hat{\Delta}_{n_k} - \Delta(Q))/\nu_{n_k} \le t\} - \Phi(t)| \to \epsilon . \tag{35}$$

Because the sequence ν_{1,n_k} and ν_{0,n_k} are bounded by Assumptions 3.1(b), there is a further subsequence, which with some abuse of notation we still denote by $\{n_k\}$, along which $\nu_{1,n_k} \to \nu_1^*$ and $\nu_{0,n_k} \to \nu_0^*$ for some $\nu_1^*, \nu_0^* \ge 0$. Then, $\nu_{1,n_k}/\nu_{n_k}, \nu_{0,n_k}/\nu_{n_k}, \nu_2/\nu_{n_k}$ all converge to constants. Therefore, it follows from Lemma S.1.2 of Bai et al. (2021) that

$$\sqrt{n}_k(\hat{\Delta}_{n_k} - \Delta(Q))/\nu_{n_k} \xrightarrow{d} N(0,1) ,$$

a contradiction to (35). Therefore, the desired convergence in Theorem 3.1 follows. The form of the variance formula as stated in the theorem can be obtained using the arguments in the proof of Theorem 3.1 almost verbatim.

It then follows from Assumption 3.1(a) and similar arguments to those in the proof of Lemma S.1.4 of Bai et al. (2021) that

$$\sup_{t \in \mathbf{R}} |P\{A_n \le t | X^{(n)}, D^{(n)}\} - \Phi(t/\sqrt{\operatorname{Var}[\phi_{1,n,i}|X_i]})| \xrightarrow{P} 0 , \tag{36}$$

where Φ is the distribution function of the standard normal distribution. Similarly,

$$\sup_{t \in \mathbf{R}} |P\{A_n \le t | X^{(n)}, D^{(n)}\} - \Phi(t/\sqrt{\operatorname{Var}[\phi_{0,n,i}|X_i]})| \xrightarrow{P} 0 , \tag{37}$$

Meanwhile, it follows from the same arguments as those in (S.22)–(S.25) of Bai et al. (2021) that

$$C_n - D_n \xrightarrow{d} N\left(0, \frac{1}{2} E\left[(E[Y_i(1)|X_i] - E[Y_i(1)] - (E[Y_i(0)|X_i] - E[Y_i(0)]))^2\right]\right) .$$

A subsequencing argument similar to the one in the proof of Lemma S.1.4 of Bai et al. (2021) implies $\sqrt{n}(\hat{\Delta}_n - \Delta(Q)) \xrightarrow{d} N(0, \sigma_n^2(Q))$, where

$$\sigma_n^2(Q) = E[\operatorname{Var}[\phi_{1,n,i}|X_i]] + E[\operatorname{Var}[\phi_{0,n,i}|X_i]] + \frac{1}{2} E\left[(E[Y_i(1)|X_i] - E[Y_i(1)] - (E[Y_i(0)|X_i] - E[Y_i(0)]))^2\right] .$$

To conclude the proof with the the variance formula as stated in the theorem, note

$$\begin{aligned}
&\operatorname{Var}\left[Y_i(0) - \frac{1}{2}(m_{1,n}(X_i, W_i) + m_{0,n}(X_i, W_i))\Big| X_i\right] \\
&= \operatorname{Var}\left[E\Big[Y_i(0) - \frac{1}{2}(m_{1,n}(X_i, W_i) + m_{0,n}(X_i, W_i))\Big| X_i, W_i\Big]\Big| X_i\right] \\
&\quad + E\Big[\operatorname{Var}\left[Y_i(0) - \frac{1}{2}(m_{1,n}(X_i, W_i) + m_{0,n}(X_i, W_i))\Big| X_i, W_i\right]\Big| X_i\Big]
\end{aligned}$$

$$
\begin{aligned}
&= \mathrm{Var}\left[E\Big[\frac{Y_i(1)+Y_i(0)}{2}\Big|X_i,W_i\Big] - \frac{1}{2}(m_{1,n}(X_i,W_i)+m_{0,n}(X_i,W_i)) - E\Big[\frac{Y_i(1)-Y_i(0)}{2}\Big|X_i,W_i\Big]\Big|X_i\right] \\
&\quad + E[\mathrm{Var}[Y_i(0)|X_i,W_i]|X_i] \\
&= \mathrm{Var}\left[E\Big[\frac{Y_i(1)+Y_i(0)}{2}\Big|X_i,W_i\Big] - \frac{1}{2}(m_{1,n}(X_i,W_i)+m_{0,n}(X_i,W_i))\Big|X_i\right] \\
&\quad + \mathrm{Var}\left[E\Big[\frac{Y_i(1)-Y_i(0)}{2}\Big|X_i,W_i\Big]\Big|X_i\right] \\
&\quad - 2\mathrm{Cov}\Big[E\Big[\frac{Y_i(1)+Y_i(0)}{2}\Big|X_i,W_i\Big] - \frac{1}{2}(m_{1,n}(X_i,W_i)+m_{0,n}(X_i,W_i)), E\Big[\frac{Y_i(1)-Y_i(0)}{2}\Big|X_i,W_i\Big]\Big|X_i\Big] \\
&\quad + E[\mathrm{Var}[Y_i(0)|X_i,W_i]|X_i]\ , \qquad (38)
\end{aligned}
$$

where the first equality follows from the law of total variance, the second one follows by direct calculation, and the last one follows by expanding the variance of the sum. Similarly,

$$
\begin{aligned}
&\mathrm{Var}\left[Y_i(1) - \frac{1}{2}(m_{1,n}(X_i,W_i)+m_{0,n}(X_i,W_i))\Big|X_i\right] \\
&= \mathrm{Var}\left[E\Big[\frac{Y_i(1)+Y_i(0)}{2}\Big|X_i,W_i\Big] - \frac{1}{2}(m_{1,n}(X_i,W_i)+m_{0,n}(X_i,W_i))\Big|X_i\right] \\
&\quad + \mathrm{Var}\left[E\Big[\frac{Y_i(1)-Y_i(0)}{2}\Big|X_i,W_i\Big]\Big|X_i\right] \\
&\quad + 2\mathrm{Cov}\Big[E\Big[\frac{Y_i(1)+Y_i(0)}{2}\Big|X_i,W_i\Big] - \frac{1}{2}(m_{1,n}(X_i,W_i)+m_{0,n}(X_i,W_i)), E\Big[\frac{Y_i(1)-Y_i(0)}{2}\Big|X_i,W_i\Big]\Big|X_i\Big] \\
&\quad + E[\mathrm{Var}[Y_i(1)|X_i,W_i]|X_i]\ . \qquad (39)
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\sigma_n^2(Q) &= \frac{1}{2}E[\mathrm{Var}[E[Y_i(1)+Y_i(0)|X_i,W_i] - (m_{1,n}(X_i,W_i)+m_{0,n}(X_i,W_i))|X_i]] \\
&\quad + \frac{1}{2}E[\mathrm{Var}[E[Y_i(1)-Y_i(0)|X_i,W_i]|X_i]] + \frac{1}{2}\mathrm{Var}[E[Y_i(1)-Y_i(0)|X_i]] \\
&\quad + E[\mathrm{Var}[Y_i(0)|X_i,W_i]|X_i] + E[\mathrm{Var}[Y_i(1)|X_i,W_i]|X_i] \\
&= \frac{1}{2}E[\mathrm{Var}[E[Y_i(1)+Y_i(0)|X_i,W_i] - (m_{1,n}(X_i,W_i)+m_{0,n}(X_i,W_i))|X_i]] \\
&\quad + \frac{1}{2}E[(E[Y_i(1)-Y_i(0)|X_i,W_i] - E[Y_i(1)-Y_i(0)|X_i])^2] \\
&\quad + \frac{1}{2}E[(E[Y_i(1)-Y_i(0)|X_i] - E[Y_i(1)-Y_i(0)])^2] \\
&\quad + E[(Y_i(0)-E[Y_i(0)|X_i,W_i])^2] + E[(Y_i(1)-E[Y_i(1)|X_i,W_i])^2] \\
&= \frac{1}{2}E[\mathrm{Var}[E[Y_i(1)+Y_i(0)|X_i,W_i] - (m_{1,n}(X_i,W_i)+m_{0,n}(X_i,W_i))|X_i]] \\
&\quad + \frac{1}{2}\mathrm{Var}[E[Y_i(1)-Y_i(0)|X_i,W_i]] + E[\mathrm{Var}[Y_i(0)|X_i,W_i]] + E[\mathrm{Var}[Y_i(1)|X_i,W_i]]\ ,
\end{aligned}
$$

where the first equality follows by definition, the second one follows from (38)–(39), the third one again follows by definition, and the last one follows because by the law of iterated expectations,

$$
E[(E[Y_i(1)-Y_i(0)|X_i,W_i] - E[Y_i(1)-Y_i(0)|X_i])(E[Y_i(1)-Y_i(0)|X_i] - E[Y_i(1)-Y_i(0)])] = 0\ .
$$

The conclusion then follows. ■

A.2 Proof of Theorem 3.2

Theorem 3.1 implies $\hat{\Delta}_n \xrightarrow{P} \Delta(Q)$. Next, we show

$$\hat{\tau}_n^2 - E[\text{Var}[\phi_{1,n,i}|X_i]] + E[\text{Var}[\phi_{0,n,i}|X_i]] + E[(E[Y_i(1)|X_i] - E[Y_i(0)|X_i])^2] \xrightarrow{P} 0 \ . \tag{40}$$

To that end, define

$$\mathring{Y}_i = Y_i - \frac{1}{2}(m_{1,n}(X_i, W_i) + m_{0,n}(X_i, W_i)) \ .$$

Note

$$\begin{aligned}\hat{\tau}_n^2 &= \frac{1}{n}\sum_{1\le j\le n}\Big(\mathring{Y}_{\pi(2j-1)} - \mathring{Y}_{\pi(2j)} + (\tilde{Y}_{\pi(2j-1)} - \tilde{Y}_{\pi(2j)} - (\mathring{Y}_{\pi(2j-1)} - \mathring{Y}_{\pi(2j)}))\Big)^2 \\ &= \frac{1}{n}\sum_{1\le j\le n}(\mathring{Y}_{\pi(2j-1)} - \mathring{Y}_{\pi(2j)})^2 + \frac{1}{n}\sum_{1\le j\le n}(\tilde{Y}_{\pi(2j-1)} - \tilde{Y}_{\pi(2j)} - (\mathring{Y}_{\pi(2j-1)} - \mathring{Y}_{\pi(2j)}))^2 \\ &\quad + \frac{2}{n}\sum_{1\le j\le n}(\tilde{Y}_{\pi(2j-1)} - \tilde{Y}_{\pi(2j)} - (\mathring{Y}_{\pi(2j-1)} - \mathring{Y}_{\pi(2j)}))(\mathring{Y}_{\pi(2j-1)} - \mathring{Y}_{\pi(2j)}) \ .\end{aligned}$$

Therefore, to establish (40), we first show

$$\frac{1}{n}\sum_{1\le j\le n}(\mathring{Y}_{\pi(2j-1)} - \mathring{Y}_{\pi(2j)})^2 - E[\text{Var}[\phi_{1,n,i}|X_i]] + E[\text{Var}[\phi_{0,n,i}|X_i]] + E[(E[Y_i(1)|X_i] - E[Y_i(0)|X_i])^2] \xrightarrow{P} 0 \tag{41}$$

and

$$\frac{1}{n}\sum_{1\le j\le n}(\tilde{Y}_{\pi(2j-1)} - \tilde{Y}_{\pi(2j)} - (\mathring{Y}_{\pi(2j-1)} - \mathring{Y}_{\pi(2j)}))^2 \xrightarrow{P} 0 \ . \tag{42}$$

(42) immediately follows from repeated applications of the inequality $(a-b)^2 \le 2(a^2+b^2)$ and (9). To verify (41), note

$$\frac{1}{n}\sum_{1\le j\le n}(\mathring{Y}_{\pi(2j-1)} - \mathring{Y}_{\pi(2j)})^2 = \frac{1}{n}\sum_{1\le i\le 2n}\mathring{Y}_i^2 - \frac{2}{n}\sum_{1\le j\le n}\mathring{Y}_{\pi(2j-1)}\mathring{Y}_{\pi(2j)} \ .$$

It follows from similar arguments to those in the proof of Lemma B.2 below that

$$\frac{1}{n}\sum_{1\le i\le 2n}\mathring{Y}_i^2 - E[\phi_{1,n,i}^2] + E[\phi_{0,n,i}^2] \xrightarrow{P} 0 \ .$$

Similarly, it follows from the proof of the same lemma that

$$\frac{2}{n}\sum_{1\le j\le n}\mathring{Y}_{\pi(2j-1)}\mathring{Y}_{\pi(2j)} - 2E[E[\phi_{1,n,i}|X_i]E[\phi_{0,n,i}|X_i]] \xrightarrow{P} 0 \ .$$

To establish (41), note

$$\begin{aligned}&E[\phi_{1,n,i}^2] + E[\phi_{0,n,i}^2] - 2E[E[\phi_{1,n,i}|X_i]E[\phi_{0,n,i}|X_i]] \\ &= E[\text{Var}[\phi_{1,n,i}|X_i]] + E[\text{Var}[\phi_{0,n,i}|X_i]] + E[E[\phi_{1,n,i}|X_i]^2] + E[E[\phi_{0,n,i}|X_i]^2] - 2E[E[\phi_{1,n,i}|X_i]E[\phi_{0,n,i}|X_i]] \\ &= E[\text{Var}[\phi_{1,n,i}|X_i]] + E[\text{Var}[\phi_{0,n,i}|X_i]] + E[(E[\phi_{1,n,i}|X_i] - E[\phi_{0,n,i}|X_i])^2] \\ &= E[\text{Var}[\phi_{1,n,i}|X_i]] + E[\text{Var}[\phi_{0,n,i}|X_i]] + E[(E[Y_i(1)|X_i] - E[Y_i(0)|X_i])^2] \ ,\end{aligned}$$

where the last equality follows from the definition of $\phi_{1,n,i}$ and $\phi_{0,n,i}$. It then follows from the Cauchy-Schwarz inequality that

$$\left|\frac{1}{n}\sum_{1\le j\le n}(\tilde{Y}_{\pi(2j-1)}-\tilde{Y}_{\pi(2j)}-(\mathring{Y}_{\pi(2j-1)}-\mathring{Y}_{\pi(2j)}))(\mathring{Y}_{\pi(2j-1)}-\mathring{Y}_{\pi(2j)})\right|$$
$$\le\left(\frac{1}{n}\sum_{1\le j\le n}(\mathring{Y}_{\pi(2j-1)}-\mathring{Y}_{\pi(2j)})^2\right)\left(\frac{1}{n}\sum_{1\le j\le n}(\tilde{Y}_{\pi(2j-1)}-\tilde{Y}_{\pi(2j)}-(\mathring{Y}_{\pi(2j-1)}-\mathring{Y}_{\pi(2j)}))^2\right)\xrightarrow{P}0\ ,$$

which, together with (41)–(42) as well as Assumptions 2.1(b) and 3.1(b), imply (40).

Next, we show

$$\hat{\lambda}_n\xrightarrow{P}E[(E[Y_i(1)|X_i]-E[Y_i(0)|X_i])^2]\ . \tag{43}$$

Note

$$\hat{\lambda}_n-\frac{2}{n}\sum_{1\le j\le\lfloor\frac{n}{2}\rfloor}(\mathring{Y}_{\pi(4j-3)}-\mathring{Y}_{\pi(4j-2)})(\mathring{Y}_{\pi(4j-1)}-\mathring{Y}_{\pi(4j)})(D_{\pi(4j-3)}-D_{\pi(4j-2)})(D_{\pi(4j-1)}-D_{\pi(4j)}) \tag{44}$$
$$=\frac{2}{n}\sum_{1\le j\le\lfloor\frac{n}{2}\rfloor}(\tilde{Y}_{\pi(4j-3)}-\mathring{Y}_{\pi(4j-3)}-(\tilde{Y}_{\pi(4j-2)}-\mathring{Y}_{\pi(4j-2)}))(\mathring{Y}_{\pi(4j-1)}-\mathring{Y}_{\pi(4j)})$$
$$\times(D_{\pi(4j-3)}-D_{\pi(4j-2)})(D_{\pi(4j-1)}-D_{\pi(4j)})$$
$$+\frac{2}{n}\sum_{1\le j\le\lfloor\frac{n}{2}\rfloor}(\mathring{Y}_{\pi(4j-3)}-\mathring{Y}_{\pi(4j-2)})(\tilde{Y}_{\pi(4j-1)}-\mathring{Y}_{\pi(4j-1)})-(\tilde{Y}_{\pi(4j)}-\mathring{Y}_{\pi(4j)}))$$
$$\times(D_{\pi(4j-3)}-D_{\pi(4j-2)})(D_{\pi(4j-1)}-D_{\pi(4j)})$$
$$+\frac{2}{n}\sum_{1\le j\le\lfloor\frac{n}{2}\rfloor}(\tilde{Y}_{\pi(4j-3)}-\mathring{Y}_{\pi(4j-3)}-(\tilde{Y}_{\pi(4j-2)}-\mathring{Y}_{\pi(4j-2)}))$$
$$\times(\tilde{Y}_{\pi(4j-1)}-\mathring{Y}_{\pi(4j-1)})-(\tilde{Y}_{\pi(4j)}-\mathring{Y}_{\pi(4j)}))(D_{\pi(4j-3)}-D_{\pi(4j-2)})(D_{\pi(4j-1)}-D_{\pi(4j)})\ .$$

In what follows, we show

$$\frac{2}{n}\sum_{1\le j\le\lfloor\frac{n}{2}\rfloor}(\mathring{Y}_{\pi(4j-3)}-\mathring{Y}_{\pi(4j-2)})^2=O_P(1) \tag{45}$$

$$\frac{2}{n}\sum_{1\le j\le\lfloor\frac{n}{2}\rfloor}(\mathring{Y}_{\pi(4j-1)}-\mathring{Y}_{\pi(4j)})^2=O_P(1) \tag{46}$$

$$\frac{2}{n}\sum_{1\le j\le\lfloor\frac{n}{2}\rfloor}(\tilde{Y}_{\pi(4j-3)}-\mathring{Y}_{\pi(4j-3)}-(\tilde{Y}_{\pi(4j-2)}-\mathring{Y}_{\pi(4j-2)}))^2=o_P(1) \tag{47}$$

$$\frac{2}{n}\sum_{1\le j\le\lfloor\frac{n}{2}\rfloor}(\tilde{Y}_{\pi(4j-1)}-\mathring{Y}_{\pi(4j-1)})-(\tilde{Y}_{\pi(4j)}-\mathring{Y}_{\pi(4j)}))^2=o_P(1) \tag{48}$$

$$\frac{2}{n}\sum_{1\le j\le\lfloor\frac{n}{2}\rfloor}(\mathring{Y}_{\pi(4j-3)}-\mathring{Y}_{\pi(4j-2)})(\mathring{Y}_{\pi(4j-1)}-\mathring{Y}_{\pi(4j)})(D_{\pi(4j-3)}-D_{\pi(4j-2)})(D_{\pi(4j-1)}-D_{\pi(4j)})$$
$$\xrightarrow{P}E[(E[Y_i(1)|X_i]-E[Y_i(0)|X_i])^2]\ . \tag{49}$$

To establish (45)–(46), note they follow directly from (41) and Assumptions 2.1(b) and 3.1(b). Next, note

(47) follows from repeated applications of the inequality $(a+b)^2 \leq 2(a^2+b^2)$ and (9). (48) can be established by similar arguments. (49) follows from similar arguments to those in the proof of Lemma S.1.7 of Bai et al. (2021), with the uniform integrability arguments replaced by arguments similar to those in the proof of Lemma B.2, together with Assumptions 2.1–2.4 and 3.1. (44)–(49) imply (43) immediately.

Finally, note we have shown

$$\hat{\sigma}_n^2 - \sigma_n^2 \xrightarrow{P} 0 \ .$$

Assumption 3.1(a) implies σ_n^2 is bounded away from zero, so

$$\frac{\hat{\sigma}_n}{\sigma_n} \xrightarrow{P} 1 \ .$$

The conclusion of the theorem then follows. ■

A.3 Proof of Theorem 4.1

We will apply the Frisch-Waugh-Lovell theorem to obtain an expression for $\hat{\beta}_n^{\text{naive}}$. Consider the linear regression of ψ_i on 1 and D_i. Define

$$\hat{\mu}_{\psi,n}(d) = \frac{1}{n} \sum_{1 \leq i \leq 2n} \psi_i I\{D_i = d\}$$

for $d \in \{0, 1\}$ and

$$\hat{\Delta}_{\psi,n} = \hat{\mu}_{\psi,n}(1) - \hat{\mu}_{\psi,n}(0) \ .$$

The ith residual based on the OLS estimation of this linear regression model is given by

$$\tilde{\psi}_i = \psi_i - \hat{\mu}_{\psi,n}(0) - \hat{\Delta}_{\psi,n} D_i \ .$$

$\hat{\beta}_n^{\text{naive}}$ is then given by the OLS estimator of the coefficient in the linear regression of Y_i on $\tilde{\psi}_i$. Note

$$\begin{aligned} \frac{1}{2n} \sum_{1 \leq i \leq 2n} \tilde{\psi}_i \tilde{\psi}_i' &= \frac{1}{2n} \sum_{1 \leq i \leq 2n} (\psi_i - \hat{\mu}_{\psi,n}(1))(\psi_i - \hat{\mu}_{\psi,n}(1))' D_i + \frac{1}{2n} \sum_{1 \leq i \leq 2n} (\psi_i - \hat{\mu}_{\psi,n}(0))(\psi_i - \hat{\mu}_{\psi,n}(0))'(1 - D_i) \\ &= \frac{1}{2n} \sum_{1 \leq i \leq 2n} \psi_i \psi_i' - \frac{1}{2} \hat{\mu}_{\psi,n}(1) \hat{\mu}_{\psi,n}(1)' - \frac{1}{2} \hat{\mu}_{\psi,n}(0) \hat{\mu}_{\psi,n}(0)' \ . \end{aligned}$$

It follows from Assumption 4.1(b) and the weak law of large number that

$$\frac{1}{2n} \sum_{1 \leq i \leq 2n} \psi_i \psi_i' \xrightarrow{P} E[\psi_i \psi_i'] \ .$$

On the other hand, it follows from Assumptions 2.2–2.3 and 4.1(b)–(c) as well as similar arguments to those in the proof of Lemma S.1.5 of Bai et al. (2021) that

$$\hat{\mu}_{\psi,n}(d) \xrightarrow{P} E[\psi_i]$$

for $d \in \{0, 1\}$. Therefore,

$$\frac{1}{2n} \sum_{1 \le i \le 2n} \tilde{\psi}_i \tilde{\psi}_i' \xrightarrow{P} \mathrm{Var}[\psi_i] \ .$$

Next,

$$\frac{1}{2n} \sum_{1 \le i \le 2n} \tilde{\psi}_i Y_i = \frac{1}{2n} \sum_{1 \le i \le 2n} (\psi_i - \hat{\mu}_{\psi,n}(1)) Y_i(1) D_i + \frac{1}{2n} \sum_{1 \le i \le 2n} (\psi_i - \hat{\mu}_{\psi,n}(0)) Y_i(0)(1 - D_i)$$

It follows from similar arguments as above as well as Assumptions 2.1(b), 2.2–2.3, and 4.1(b)–(c) that

$$\frac{1}{2n} \sum_{1 \le i \le 2n} \tilde{\psi}_i Y_i \xrightarrow{P} \mathrm{Cov}[\psi_i, Y_i(1) + Y_i(0)] \ .$$

The convergence of $\hat{\beta}_n^{\text{naive}}$ therefore follows from the continuous mapping theorem and Assumption 4.1(a).

To see (9) is satisfied, note

$$\frac{1}{2n} \sum_{1 \le i \le 2n} (\hat{m}_{d,n}(X_i, W_i) - m_{d,n}(X_i, W_i))^2 = (\hat{\beta}_n^{\text{naive}} - \beta^{\text{naive}})' \left(\frac{1}{2n} \sum_{1 \le i \le 2n} \psi_i \psi_i' \right) (\hat{\beta}_n^{\text{naive}} - \beta^{\text{naive}}) \ .$$

(9) then follows from the fact that $\hat{\beta}_n^{\text{naive}} \xrightarrow{P} \beta$, Assumption 4.1(b), and the weak law of large numbers. To establish (6), first note

$$\frac{1}{\sqrt{2n}} \sum_{1 \le i \le 2n} (2D_i - 1)(\hat{m}_{d,n}(X_i, W_i) - m_{d,n}(X_i, W_i)) = \frac{1}{\sqrt{2}} \sqrt{n} \hat{\Delta}_{\psi,n}' (\hat{\beta}_n^{\text{naive}} - \beta^{\text{naive}}) \ .$$

In what follows, we establish

$$\sqrt{n} \hat{\Delta}_{\psi,n} = O_P(1) \ , \tag{50}$$

from which (6) follows immediately because $\hat{\beta}_n^{\text{naive}} - \beta^{\text{naive}} = o_P(1)$. Note by Assumption 2.2 that $E[\sqrt{n} \hat{\Delta}_{\psi,n} | X^{(n)}] = 0$. Also note

$$\sqrt{n} \hat{\Delta}_{\psi,n} = F_n - G_n + H_n \ ,$$

where

$$\begin{aligned}
F_n &= \frac{1}{\sqrt{n}} \sum_{1 \le i \le 2n} (\psi_i - E[\psi_i | X_i]) D_i \ , \\
G_n &= \frac{1}{\sqrt{n}} \sum_{1 \le i \le 2n} (\psi_i - E[\psi_i | X_i])(1 - D_i) \ , \quad \text{and} \\
H_n &= \frac{1}{\sqrt{n}} \sum_{1 \le j \le n} (E[\psi_{\pi(2j-1)} | X_{\pi(2j-1)}] - E[\psi_{\pi(2j)} | X_{\pi(2j)}])(D_{\pi(2j-1)} - D_{\pi(2j)}) \ .
\end{aligned}$$

We will argue F_n, G_n, H_n are all $O_P(1)$. Since this could be carried out separately for each entry of F_n and G_n, we assume without loss of generality that $k_\psi = 1$. First, it follows from Assumptions 2.2–2.3 and 4.1(c) as well as similar arguments to those in the proof of Lemma S.1.4 of Bai et al. (2021) that

$$\mathrm{Var}[F_n | X^{(n)}, D^{(n)}] = \frac{1}{n} \sum_{1 \le i \le 2n} \mathrm{Var}[\psi_i | X_i] D_i \xrightarrow{P} E[\mathrm{Var}[\psi_i | X_i]] > 0 \ .$$

It then follows from similar arguments using the Lindeberg central limit theorem as in the proof of Lemma S.1.4 of Bai et al. (2021) that $F_n = O_P(1)$. Similar arguments establish $G_n = O_P(1)$. Finally, we show $H_n = O_P(1)$. Note that $E[H_n | X^{(n)}] = 0$ and by Assumptions 2.2–2.3 and 4.1(c),

$$\mathrm{Var}[H_n | X^{(n)}] = \frac{1}{n} \sum_{1 \le j \le n} (E[\psi_{\pi(2j-1)} | X_{\pi(2j-1)}] - E[\psi_{\pi(2j)} | X_{\pi(2j)}])^2 \xrightarrow{P} 0 \ .$$

Therefore, for any fixed $\epsilon > 0$, Markov's inequality implies

$$P\{|H_n - E[H_n | X^{(n)}]| > \epsilon | X^{(n)}\} \le \frac{\mathrm{Var}[H_n | X^{(n)}]}{\epsilon^2} \xrightarrow{P} 0 \ .$$

Since probabilities are bounded and therefore uniformly integrable, we have that

$$P\{|H_n - E[H_n | X^{(n)}]| > \epsilon\} \to 0 \ .$$

Therefore, (50) follows. Finally, it is straightforward to see Assumption 3.1 is implied by Assumption 4.1. ■

A.4 Proof of Theorem 4.2

By the Frisch-Waugh-Lovell theorem, $\hat{\beta}_n^{\mathrm{pfe}}$ is equal to the OLS estimator in the linear regression of $\{(Y_{\pi(2j-1)} - Y_{\pi(2j)}, Y_{\pi(2j)} - Y_{\pi(2j-1)}) : 1 \le j \le n\}$ on $\{(2D_{\pi(2j-1)} - 1, 2D_{\pi(2j)} - 1) : 1 \le j \le n\}$ and $\{(\psi_{\pi(2j-1)} - \psi_{\pi(2j)}, \psi_{\pi(2j)} - \psi_{\pi(2j-1)}) : 1 \le j \le n\}$. To apply the Frisch-Waugh-Lovell theorem again, we study the linear regression of $\{(\psi_{\pi(2j-1)} - \psi_{\pi(2j)}, \psi_{\pi(2j)} - \psi_{\pi(2j-1)}) : 1 \le j \le n\}$ on $\{(2D_{\pi(2j-1)} - 1, 2D_{\pi(2j)} - 1) : 1 \le j \le n\}$. The OLS estimator of the regression coefficient in such a regression equals

$$\hat{\Delta}_{\psi,n} = \frac{1}{n} \sum_{1 \le j \le n} (D_{\pi(2j-1)} - D_{\pi(2j)})(\psi_{\pi(2j-1)} - \psi_{\pi(2j)}) \ .$$

The residual is therefore $\{(\psi_{\pi(2j-1)} - \psi_{\pi(2j)} - (2D_{\pi(2j-1)} - 1)\hat{\Delta}_{\psi,n}, \psi_{\pi(2j)} - \psi_{\pi(2j-1)} - (2D_{\pi(2j)} - 1)\hat{\Delta}_{\psi,n}) : 1 \le j \le n\}$. $\hat{\beta}_n^{\mathrm{pfe}}$ equals the OLS estimator of the coefficient in the linear regression of $\{(Y_{\pi(2j-1)} - Y_{\pi(2j)}, Y_{\pi(2j)} - Y_{\pi(2j-1)}) : 1 \le j \le n\}$ on those residuals. Define

$$\begin{aligned} \delta_{Y,j} &= (D_{\pi(2j-1)} - D_{\pi(2j)})(Y_{\pi(2j-1)} - Y_{\pi(2j)}) \quad \text{and} \\ \delta_{\psi,j} &= (D_{\pi(2j-1)} - D_{\pi(2j)})(\psi_{\pi(2j-1)} - \psi_{\pi(2j)}) \end{aligned}$$

Apparently $\hat{\Delta}_{\psi,n} = \frac{1}{n} \sum_{1 \le j \le n} \delta_{\psi,j}$. A moment's thought reveals that $\hat{\beta}_n^{\mathrm{pfe}}$ further equals the coefficient estimate using least squares in the linear regression of $\delta_{Y,j}$ on $\delta_{\psi,j} - \hat{\Delta}_{\psi,n}$ for $1 \le j \le n$. It follows from Assumptions 2.1(b)–(c), 2.2–2.3, and 4.1(b)–(c) as well as similar arguments to those in the proof of Lemma

S.1.5 of Bai et al. (2021) that

$$\hat{\Delta}_{\psi,n} \xrightarrow{P} 0 \quad \text{and} \tag{51}$$
$$\frac{1}{n}\sum_{1\le j\le n}\delta_{Y,j} \xrightarrow{P} \Delta(Q)\ .$$

Next, note that

$$\begin{aligned}
&\frac{1}{n}\sum_{1\le j\le n}\delta_{\psi,j}\delta_{\psi,j}' \\
&= \frac{1}{n}\sum_{1\le j\le n}(\psi_{\pi(2j-1)}-\psi_{\pi(2j)})(\psi_{\pi(2j-1)}-\psi_{\pi(2j)})' \\
&= \frac{1}{n}\sum_{1\le i\le 2n}\psi_i\psi_i' - \frac{1}{n}\sum_{1\le j\le n}(\psi_{\pi(2j-1)}\psi_{\pi(2j)}' + \psi_{\pi(2j)}\psi_{\pi(2j-1)}')\ .
\end{aligned} \tag{52}$$

For convenience, we introduce the following notation:

$$\begin{aligned}
\mu_d(X_i) &= E[Y_i(d)|X_i] \\
\Psi(X_i) &= E[\psi_i|X_i] \\
\xi_d(X_i) &= E[\psi_i Y_i(d)|X_i]\ .
\end{aligned}$$

The first term in (52) converges in probability to $2E[\psi_i\psi_i']$ by the weak law of large numbers. For the second term, we have that

$$\begin{aligned}
&E\Big[\frac{1}{n}\sum_{1\le j\le n}(\psi_{\pi(2j-1)}\psi_{\pi(2j)}' + \psi_{\pi(2j)}\psi_{\pi(2j-1)}')\Big|X^{(n)}\Big] \\
&= \frac{1}{n}\sum_{1\le i\le 2n}\Psi(X_i)\Psi(X_i)' - \frac{1}{n}\sum_{1\le j\le n}(\Psi(X_{\pi(2j-1)})-\Psi(X_{\pi(2j)}))(\Psi(X_{\pi(2j-1)})-\Psi(X_{\pi(2j)}))' \\
&\xrightarrow{P} 2E[\Psi(X_i)\Psi(X_i)']\ ,
\end{aligned}$$

where the convergence in probability holds because of Assumptions 2.2–2.3 and 4.1(c). It follows from Assumptions 2.2–2.3 and 4.1(b)–(c) as well as similar arguments to those in the proof of Lemma S.1.6 of Bai et al. (2021) that

$$\Big|\frac{1}{n}\sum_{1\le j\le n}(\psi_{\pi(2j-1)}\psi_{\pi(2j)}' + \psi_{\pi(2j)}\psi_{\pi(2j-1)}') - E\Big[\frac{1}{n}\sum_{1\le j\le n}(\psi_{\pi(2j-1)}\psi_{\pi(2j)}' + \psi_{\pi(2j)}\psi_{\pi(2j-1)}')\Big|X^{(n)}\Big]\Big| \xrightarrow{P} 0\ .$$

Therefore,

$$\frac{1}{n}\sum_{1\le j\le n}\delta_{\psi,j}\delta_{\psi,j}' \xrightarrow{P} 2E[\operatorname{Var}[\psi_i|X_i]]\ .$$

We now turn to

$$\frac{1}{n}\sum_{1\le j\le n}\delta_{\psi,j}\delta_{Y,j} = \frac{1}{n}\sum_{1\le j\le n}(\psi_{\pi(2j-1)}-\psi_{\pi(2j)})(Y_{\pi(2j-1)}-Y_{\pi(2j)})\ .$$

Note that

$$E[\psi_{\pi(2j-1)}Y_{\pi(2j-1)}|X^{(n)}] = \frac{1}{2}\xi_1(X_{\pi(2j-1)}) + \frac{1}{2}\xi_0(X_{\pi(2j-1)})$$
$$E[\psi_{\pi(2j-1)}Y_{\pi(2j)}|X^{(n)}] = \frac{1}{2}\Psi(X_{\pi(2j-1)})(\mu_1(X_{\pi(2j)}) + \mu_0(X_{\pi(2j)}))\ .$$

It follows from Assumptions 2.1(b)–(c), 2.2–2.3, 4.1(b)–(c) as well as similar arguments to those in the proof of Lemma S.1.6 of Bai et al. (2021) that

$$\frac{1}{n}\sum_{1\le j\le n}\delta_{\psi,j}\delta_{Y,j} \xrightarrow{P} E[\psi_i(Y_i(1)+Y_i(0))] - E[\Psi(X_i)(\mu_1(X_i)+\mu_0(X_i))]\ .$$

The convergence in probability of $\hat{\beta}_n^{\rm pfe}$ now follows from Assumption 4.1(a) and the continuous mapping theorem. (6)–(9) can be established using similar arguments to those in the proof of Theorem 4.1. Finally, it is straightforward to see Assumption 3.1 is implied by Assumption 4.1. ■

A.5 Proof of Theorem 5.1

We first show

$$|\hat{\alpha}_n^{\rm hd-pd} - \alpha_n^{\rm hd-pd}| + \|\hat{\beta}_n^{\rm hd-pd} - \beta_n^{\rm hd-pd}\|_1 = O_P\left(s_n^{\rm hd-pd}\lambda_n^{\rm hd-pd}\right)\ . \tag{53}$$

Note that

$$\frac{1}{n}\sum_{1\le j\le n}(\delta_{Y,j} - \hat{\alpha}_n^{\rm hd-pd} - \delta'_{\psi,j}\hat{\beta}_n^{\rm hd-pd})^2 + \lambda_n^{\rm hd-pd}\|\hat{\Omega}_n\hat{\beta}_n^{\rm hd-pd}\|_1$$
$$\le \frac{1}{n}\sum_{1\le j\le n}(\delta_{Y,j} - \alpha_n^{\rm hd-pd} - \delta'_{\psi,j}\beta_n^{\rm hd-pd})^2 + \lambda_n^{\rm hd-pd}\|\hat{\Omega}_n\beta_n^{\rm hd-pd}\|_1\ .$$

Rearranging the terms, we then have

$$\frac{1}{n}\sum_{1\le j\le n}(\hat{\alpha}_n^{\rm hd-pd} - \alpha_n^{\rm hd-pd} + \delta'_{\psi,j}(\hat{\beta}_n^{\rm hd-pd} - \beta_n^{\rm hd-pd}))^2 + \lambda_n^{\rm hd-pd}\|\hat{\Omega}_n\hat{\beta}_n^{\rm hd-pd}\|_1$$
$$\le \left(\frac{2}{n}\sum_{1\le j\le n}\delta_{\epsilon,j}\right)(\hat{\alpha}_n^{\rm hd-pd} - \alpha_n^{\rm hd-pd}) + \left(\frac{2}{n}\sum_{1\le j\le n}\delta_{\epsilon,j}\delta'_{\psi,j}\right)(\hat{\beta}_n^{\rm hd-pd} - \beta_n^{\rm hd-pd}) + \lambda_n^{\rm hd-pd}\|\hat{\Omega}_n\beta_n^{\rm hd-pd}\|_1\ , \tag{54}$$

where $\alpha_n^{\rm hd-pd} = \alpha_{n,1}^{\rm hd-pd} - \alpha_{n,0}^{\rm hd-pd}$ and

$$\delta_{\epsilon,j} = (D_{\pi(2j)} - D_{\pi(2j-1)})(Y_{\pi(2j)} - Y_{\pi(2j-1)} - (\psi_{n,\pi(2j)} - \psi_{n,\pi(2j-1)})\beta_n^{\rm hd-pd}) - \alpha_n^{\rm hd-pd}$$

Next, define

$$\mathbb{U}_n = \hat{\Omega}_n^{-1}\frac{1}{n}\sum_{1\le j\le n}\delta_{\epsilon,j}\delta_{\psi,j}$$

and

$$\mathcal{E}_n(d) = \left\{ \|\mathbb{U}_n\|_\infty \leq \ell\ell_n^{1/2} \left(\sqrt{\frac{\log(2np_n)}{n}} + \zeta_n \right), \left| \frac{1}{n} \sum_{1 \leq j \leq n} \delta_{\epsilon,j} \right| \leq \ell\ell_n^{1/2} \left(\sqrt{\frac{\log(2np_n)}{n}} + \zeta_n \right) \right\} .$$

Lemma B.4 implies $P\{\mathcal{E}_n(d)\} \to 1$ for $d \in \{0,1\}$.

On the event $\mathcal{E}_n(d)$, we have

$$\begin{aligned}
&\left| \left(\frac{2}{n} \sum_{1 \leq j \leq n} \delta_{\epsilon,j} \delta'_{\psi,j} \right) (\hat{\beta}_n^{\mathrm{hd-pd}} - \beta_n^{\mathrm{hd-pd}}) \right| \\
&\leq \left\| \hat{\Omega}_n^{-1} \frac{2}{n} \sum_{1 \leq j \leq n} \delta_{\epsilon,j} \delta'_{\psi,j} \right\|_\infty \|\hat{\Omega}_n (\hat{\beta}_n^{\mathrm{hd-pd}} - \beta_n^{\mathrm{hd-pd}})\|_1 \\
&\leq 2\|\mathbb{U}_n\|_\infty \|\hat{\Omega}_n (\hat{\beta}_n^{\mathrm{hd-pd}} - \beta_n^{\mathrm{hd-pd}})\|_1 \\
&\leq 2\ell\ell_n^{-1/2} \lambda_n^{\mathrm{hd-pd}} \|\hat{\Omega}_n (\hat{\beta}_n^{\mathrm{hd-pd}} - \beta_n^{\mathrm{hd-pd}})\|_1 ,
\end{aligned}$$

where the last inequality follows the fact that

$$\lambda_n^{\mathrm{hd-pd}} \geq \ell\ell_n \left(\sqrt{\frac{\log(2np_n)}{n}} + \zeta_n \right) .$$

Next, define

$$\hat{\delta}_{d,n} = \hat{\beta}_n^{\mathrm{hd-pd}} - \beta_n^{\mathrm{hd-pd}}$$

and let $S_{d,n}$ be the support of $\beta_n^{\mathrm{hd-pd}}$. Then, we have

$$\begin{aligned}
&2\ell\ell_n^{-1/2} \|\hat{\Omega}_n \hat{\delta}_{d,n}\|_1 + \|\hat{\Omega}_n \beta_n^{\mathrm{hd-pd}}\|_1 - \|\hat{\Omega}_n \hat{\beta}_n^{\mathrm{hd-pd}}\|_1 \\
&= 2\ell\ell_n^{-1/2} \|(\hat{\Omega}_n \hat{\delta}_{d,n})_{S_{d,n}}\|_1 + 2\ell\ell_n^{-1/2} \|(\hat{\Omega}_n \hat{\delta}_{d,n})_{S^c_{d,n}}\|_1 + \|\hat{\Omega}_n \beta_n^{\mathrm{hd-pd}}\|_1 - \|\hat{\Omega}_n \hat{\beta}_n^{\mathrm{hd-pd}}\|_1 ,
\end{aligned}$$

$$\|\hat{\Omega}_n \hat{\beta}_n^{\mathrm{hd-pd}}\|_1 = \|(\hat{\Omega}_n \hat{\beta}_n^{\mathrm{hd-pd}})_{S_{d,n}}\|_1 + \|(\hat{\Omega}_n \hat{\beta}_n^{\mathrm{hd-pd}})_{S^c_{d,n}}\|_1 = \|(\hat{\Omega}_n \hat{\beta}_n^{\mathrm{hd-pd}})_{S_{d,n}}\|_1 + \|(\hat{\Omega}_n \hat{\delta}_{d,n})_{S^c_{d,n}}\|_1 ,$$

and

$$\|\hat{\Omega}_n \beta_n^{\mathrm{hd-pd}}\|_1 = \|(\hat{\Omega}_n \beta_n^{\mathrm{hd-pd}})_{S_{d,n}}\|_1 \leq \|(\hat{\Omega}_n \hat{\beta}_n^{\mathrm{hd-pd}})_{S_{d,n}}\|_1 + \|(\hat{\Omega}_n \hat{\delta}_{d,n}^{\mathrm{hd-pd}})_{S_{d,n}}\|_1 .$$

Denote $\breve{\delta}_{d,n} = (\hat{\alpha}_n^{\mathrm{hd-pd}} - \alpha_n^{\mathrm{hd-pd}}, \hat{\delta}'_{d,n})'$, $\breve{\delta}_{\psi,j} = (1, \delta'_{\psi,j})'$, and $\breve{S}_{d,n} = \{1, S_{d,n} + 1\}$.[4] Together with (54), we have

$$\begin{aligned}
0 &\leq \frac{1}{n} \sum_{1 \leq j \leq n} (\breve{\delta}'_{\psi,j} \breve{\delta}_{d,n})^2 \\
&\leq \lambda_n^{\mathrm{hd-pd}} \left[\left(1 + 2\ell\ell_n^{-1/2}\right) \|(\hat{\Omega}_n \hat{\delta}_{d,n})_{S_{d,n}}\|_1 - \left(1 - 2\ell\ell_n^{-1/2}\right) \|(\hat{\Omega}_n \hat{\delta}_{d,n})_{S^c_{d,n}}\|_1 \right] + \ell\ell_n^{-1/2} \left(\lambda_n^{\mathrm{hd-pd}}\right) |\hat{\alpha}_n^{\mathrm{hd-pd}} - \alpha_n^{\mathrm{hd-pd}}|
\end{aligned}$$

[4]Suppose $S_{d,n} = \{1, 4, 10\}$, then $\breve{S}_{d,n} = \{1, 2, 5, 10\}$.

$$\leq \lambda_n^{\rm hd-pd} \left[\left(1 + 2\ell\ell_n^{-1/2}\right) \bar{c} \|(\breve{\delta}_{d,n})_{\breve{S}_{d,n}}\|_1 - \left(1 - 2\ell\ell_n^{-1/2}\right) \underline{c} \|(\breve{\delta}_{d,n})_{\breve{S}^c_{d,n}}\|_1 \right] . \tag{55}$$

Define

$$\mathcal{C}_n = \left\{ u \in \mathbf{R}^{p_n+1} : \|u_{\breve{S}^c_{d,n}}\|_1 \leq \frac{2\bar{c}}{\underline{c}} \|u_{\breve{S}_{d,n}}\|_1 \right\} .$$

Then, we have $\breve{\delta}_{d,n} \in \mathcal{C}_n$. It follows from Bickel et al. (2009, Lemma 4.1) and Assumption 5.4 that

$$\inf_{u \in \mathcal{C}_n} (\|u_{\breve{S}_{d,n}}\|_1)^{-2} (s_n^{\rm hd-pd} + 1) u' \left(\frac{1}{n} \sum_{1 \leq j \leq n} \breve{\delta}_{\psi,j} \breve{\delta}'_{\psi,j} \right) u \geq 0.25 \kappa_1^2 .$$

Therefore, we have

$$0.25\kappa_1^2 \|(\breve{\delta}_{d,n})_{\breve{S}_{d,n}}\|_1^2 \leq \left(1 + 2\ell\ell_n^{-1/2}\right) \bar{c} \lambda_n^{\rm hd-pd} (s_n^{\rm hd-pd} + 1) \|(\breve{\delta}_{d,n})_{\breve{S}_{d,n}}\|_1 ,$$

which implies

$$\|(\breve{\delta}_{d,n})_{\breve{S}_{d,n}}\|_1 \leq 4 \left(1 + 2\ell\ell_n^{-1/2}\right) (s_n^{\rm hd-pd} + 1) \lambda_n^{\rm hd-pd} \bar{c} / \kappa_1^2 .$$

We then have

$$\begin{aligned}
&|\hat{\alpha}_n^{\rm hd-pd} - \alpha_n^{\rm hd-pd}| + \|\hat{\beta}_n^{\rm hd-pd} - \beta_n^{\rm hd-pd}\|_1 \\
&= \|(\breve{\delta}_{d,n})_{\breve{S}_{d,n}}\|_1 + \|(\breve{\delta}_{d,n})_{\breve{S}^c_{d,n}}\|_1 \\
&\leq (1 + 2\bar{c}/\underline{c}) \|(\breve{\delta}_{d,n})_{\breve{S}_{d,n}}\|_1 \leq 4(1 + 2\bar{c}/\underline{c}) \left(1 + 2\ell\ell_n^{-1/2}\right) (s_n^{\rm hd-pd} + 1) \lambda_n^{\rm hd-pd} \bar{c} / \kappa_1^2.
\end{aligned}$$

Then, (53) holds because $P\{\mathcal{E}_n(d)\} \to 1$. (9) also follows follows (55) because

$$\begin{aligned}
&\frac{1}{n} \sum_{1 \leq i \leq 2n} I\{D_i = d\} (\hat{\alpha}_n^{\rm hd-pd} - \alpha_n^{\rm hd-pd} + \psi'_{n,i}(\hat{\beta}_{d,n} - \beta_{d,n}))^2 \\
&\leq \lambda_n^{\rm hd-pd} \left(1 + 2\ell\ell_n^{-1/2}\right) \bar{c} \|(\breve{\delta}_{d,n})_{\breve{S}_{d,n}}\|_1 \\
&\leq \lambda_n^{\rm hd-pd} \left(1 + 2\ell\ell_n^{-1/2}\right) \bar{c} \|(\breve{\delta}_{d,n})\|_1 \\
&= O_P\left(s_n^{\rm hd-pd} (\lambda_n^{\rm hd-pd})^2\right) = o_P(1) .
\end{aligned}$$

Next, we show (6) for $\hat{\beta}_n^{\rm hd-pd}$. First note

$$\begin{aligned}
&\left| \frac{1}{\sqrt{2n}} \sum_{1 \leq i \leq 2n} (2D_i - 1)(\hat{m}_{d,n}(X_i, W_{n,i}) - m_{d,n}(X_i, W_{n,i})) \right| \\
&= \left| \frac{1}{\sqrt{2n}} \sum_{1 \leq i \leq 2n} (2D_i - 1) \psi'_{n,i} (\hat{\beta}_n^{\rm hd-pd} - \beta_n^{\rm hd-pd}) \right| \\
&\leq \left\| \frac{1}{\sqrt{2n}} \sum_{1 \leq i \leq n} (2D_i - 1) \psi_{n,i} \right\|_\infty \|\hat{\beta}_n^{\rm hd-pd} - \beta_n^{\rm hd-pd}\|_1 .
\end{aligned}$$

Note that it follows from Assumption 2.2 that conditional on $X^{(n)}$ and $W_n^{(n)}$,

$$\{D_{\pi(2j-1)} - D_{\pi(2j)} : 1 \le j \le n\}$$

is a sequence of independent Rademacher random variables. Therefore, Hoeffding's inequality implies

$$\begin{aligned}
& P\left\{\left\|\frac{1}{\sqrt{2n}}\sum_{1\le i\le 2n}(2D_i-1)\psi_{n,i}\right\|_\infty > t\middle| X^{(n)}, W_n^{(n)}\right\} \\
& \le \sum_{1\le l\le p_n} P\left\{\left|\frac{1}{\sqrt{2n}}\sum_{1\le j\le n}(\psi_{n,\pi(2j-1)}-\psi_{n,\pi(2j)})(D_{\pi(2j-1)}-D_{\pi(2j)})\right| > t\middle| X^{(n)}, W_n^{(n)}\right\} \\
& \le \sum_{1\le l\le p_n} 2\exp\left(-\frac{t^2}{\frac{1}{n}\sum_{1\le j\le n}(\psi_{n,\pi(2j-1)}-\psi_{n,\pi(2j)})^2}\right) .
\end{aligned}$$

Define

$$\nu_n^2 = \max_{1\le l\le p_n}\frac{1}{n}\sum_{1\le i\le 2n}\psi_{n,i,l}^2 .$$

We then have

$$P\left\{\left\|\frac{1}{\sqrt{2n}}\sum_{1\le i\le 2n}(2D_i-1)\psi_{n,i}\right\|_\infty > \nu_n\sqrt{2\log(p_n\vee n)}\middle| X^{(n)}, W_n^{(n)}\right\} \le (p_n\vee n)^{-1} . \tag{56}$$

Next, we determine the order of ν_n^2. Note

$$\begin{aligned}
E[\nu_n^2] & \le \max_{1\le l\le p_n} 2E[\psi_{n,i,l}^2] + 2E\left[\frac{1}{2n}\sum_{1\le i\le 2n}(\psi_{n,i,l}^2 - E[\psi_{n,i,l}^2])\right] \\
& \lesssim 1 + E\left[\max_{1\le l\le p_n}\left|\frac{1}{2n}\sum_{1\le i\le 2n}e_i\psi_{n,i,l}^2\right|\right] \\
& \lesssim 1 + \Xi_n E\left[\max_{1\le l\le p_n}\left|\frac{1}{2n}\sum_{1\le i\le 2n}e_i\psi_{n,i,l}\right|\right] \\
& \lesssim 1 + \Xi_n E\left[\max_{1\le l\le p_n}\left|\frac{1}{2n}\sum_{1\le i\le 2n}e_i(\psi_{n,i,l}-E[\psi_{n,i,l}])\right|\right] + \max_{1\le l\le p_n}|E[\psi_{n,i,l}]| \\
& \lesssim 1 + \Xi_n E\left[\sup_{f\in\mathcal{F}_n}\left|\frac{1}{2n}\sum_{1\le i\le 2n}f(e_i,\psi_{n,i,l})\right|\right]
\end{aligned}$$

where $\{e_i : 1 \le i \le n\}$ is an i.i.d. sequence of Rademacher random variables,

$$\mathcal{F}_n = \{f : \mathbf{R}\times\mathbf{R}^{p_n} \mapsto \mathbf{R}, f(e,\psi) = e\psi_l, 1 \le l \le p_n\} ,$$

and ψ_l is the lth element of ψ. Note the second inequality follows from Lemma 2.3.1 of van der Vaart and Wellner (1996), the third inequality follows from Theorem 4.12 of Ledoux and Talagrand (1991) and the definition of Ξ_n, and the last follows from Assumption 5.1. Note also $\mathcal{F}_n$ has an envelope $F = \Xi_n$ and

$$\sup_{n\geq 1}\sup_{f\in\mathcal{F}_n} E[f^2] < \infty$$

because of Assumption 5.1. Because the cardinality of $\mathcal{F}_n$ is p_n, for any $\epsilon < 1$ we have that

$$\sup_{Q:Q \text{ is a discrete distribution with finite support}} \mathcal{N}(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)) \leq \frac{p_n}{\epsilon} ,$$

where $\mathcal{N}(\epsilon, \mathcal{F}, L_2(Q))$ is the covering number for class $\mathcal{F}$ under the metric $L_2(Q)$ using balls of radius ϵ. Therefore, Corollary 5.1 of Chernozhukov et al. (2014) implies

$$E\left[\sup_{f\in\mathcal{F}_n}\left|\frac{1}{2n}\sum_{1\leq i\leq 2n} e_i(\psi_{n,i,l} - E[\psi_{n,i,l}])\right|\right] \lesssim \sqrt{\frac{\log p_n}{n}} + \frac{\Xi_n \log p_n}{n} = o(\Xi_n^{-1}) .$$

Therefore, $\nu_n = O_P(1)$. Together with (56), they imply

$$\left\|\frac{1}{\sqrt{2n}}\sum_{1\leq i\leq 2n}(2D_i - 1)\psi_{n,i}\right\|_\infty = O_P\left(\sqrt{\log(p_n \vee n)}\right) .$$

In light of (53) and Assumption 5.3, we have

$$\left\|\frac{1}{\sqrt{2n}}\sum_{1\leq i\leq 2n}(2D_i - 1)\psi_{n,i}\right\|_\infty \|\hat{\beta}_n^{\text{hd-pd}} - \beta_n^{\text{hd-pd}}\|_1 = O_P\left(s_n^{\text{hd-pd}}\log^{1/2}(p_n \vee n)\lambda_n^{\text{hd-pd}}\right) = o_P(1) .$$

Next, we turn to the variance $\sigma_n^{\text{hd-pd},2}$. Let $\tilde{Y}_i(d) = Y_i(d) - \mu_d(X_i)$ for $d \in \{0,1\}$. By Theorem 3.1, we have

$$\begin{aligned}
\sigma_n^{\text{hd-pd},2} - \sigma_n^{\text{na},2} &= \frac{1}{2}E\left[\text{Var}\left[E[Y_i(1) + Y_i(0)|X_i, W_{n,i}] - (m_{1,n}(X_i, W_{n,i}) + m_{0,n}(X_i, W_{n,i}))\Big|X_i\right]\right] \\
&\quad - \frac{1}{2}E\left[\text{Var}\left[E[Y_i(1) + Y_i(0)|X_i, W_{n,i}]\Big|X_i\right]\right] \\
&= \frac{1}{2}E\left[E[\tilde{Y}_i(1) + \tilde{Y}_i(0)|W_{n,i}, X_i] - \tilde{\psi}_{n,i}'(\beta_{1,n}^{\text{hd-pd}} + \beta_{0,n}^{\text{hd-pd}})\right]^2 \\
&\quad - \frac{1}{2}E\left[E[\tilde{Y}_i(1) + \tilde{Y}_i(0)|W_{n,i}, X_i]\right]^2 \\
&= -E\left[(\epsilon_{n,i}(1) + \epsilon_{n,i}(0))(\tilde{\psi}_{n,i}'(\beta_{1,n}^{\text{hd-pd}} + \beta_{0,n}^{\text{hd-pd}}))\right] - \frac{1}{2}E\left[(\tilde{\psi}_{n,i}'(\beta_{1,n}^{\text{hd-pd}} + \beta_{0,n}^{\text{hd-pd}}))^2\right] \\
&\leq -E\left[(\epsilon_{n,i}(1) + \epsilon_{n,i}(0))(\tilde{\psi}_{n,i}'(\beta_{1,n}^{\text{hd-pd}} + \beta_{0,n}^{\text{hd-pd}}))\right] .
\end{aligned}$$

It suffices to show $E\left[(\epsilon_{n,i}(1) + \epsilon_{n,i}(0))(\tilde{\psi}_{n,i}'(\beta_{1,n}^{\text{hd-pd}} + \beta_{0,n}^{\text{hd-pd}}))\right] = o(1)$. We have

$$\begin{aligned}
\left|E\left[(\epsilon_{n,i}(1) + \epsilon_{n,i}(0))(\tilde{\psi}_{n,i}'(\beta_{1,n}^{\text{hd-pd}} + \beta_{0,n}^{\text{hd-pd}}))\right]\right| &\leq \left\|E\left[(\epsilon_{n,i}(1) + \epsilon_{n,i}(0))\tilde{\psi}_{n,i}\right]\right\|_\infty \|\beta_{1,n}^{\text{hd-pd}} + \beta_{0,n}^{\text{hd-pd}}\|_1 \\
&= o(\lambda_n^{\text{hd-pd}} s_n^{\text{hd-pd}}) = o(1) ,
\end{aligned}$$

where the second last equality is by (18) and (19) and the last equality is by Assumption 5.3(f). This leads to the desired result that

$$\limsup_{n\geq 1}(\sigma_n^{\text{hd}-\text{pd},2} - \sigma_n^{\text{na},2}) \leq 0.$$

Last, we assume the true specification is approximately sparse as specified in Theorem 5.1. Then, it suffices to show

$$E\left[(E[\tilde{Y}_i(1) + \tilde{Y}_i(0)|W_{n,i}, X_i] - \tilde{\psi}'_{n,i}(\beta_{1,n}^{\text{hd}-\text{pd}} + \beta_{0,n}^{\text{hd}-\text{pd}}))^2\right] = o(1).$$

Note that

$$\begin{aligned}
&E\left[(E[\tilde{Y}_i(1) + \tilde{Y}_i(0)|W_{n,i}, X_i] - \tilde{\psi}'_{n,i}(\beta_{1,n}^{\text{hd}-\text{pd}} + \beta_{0,n}^{\text{hd}-\text{pd}}))^2\right] \\
&\lesssim E\left[(\tilde{\psi}'_{n,i}(\beta_{1,n}^{*} + \beta_{1,n}^{*} - (\beta_{1,n}^{\text{hd}-\text{pd}} + \beta_{0,n}^{\text{hd}-\text{pd}})))^2\right] + o(1) \\
&\lesssim E\left[(\tilde{\psi}'_{n,i}(\beta_{1,n}^{*} - \beta_{1,n}^{\text{hd}-\text{pd}}))^2\right] + E\left[(\tilde{\psi}'_{n,i}(\beta_{0,n}^{*} - \beta_{0,n}^{\text{hd}-\text{pd}}))^2\right] + o(1)
\end{aligned}$$

In addition, we have

$$\begin{aligned}
E\tilde{\psi}_{n,i}\tilde{\psi}'_{n,i}(\beta_{1,n}^{*} - \beta_{1,n}^{\text{hd}-\text{pd}}) &= E\left\{\tilde{\psi}_{n,i}\left[E[Y_i(1)|W_{n,i}, X_i] - R_i - \Psi(X_i)'\beta_{1,n}^{*} - \alpha_{1,n}^{*} - \tilde{\psi}'_{n,i}\beta_{1,n}^{\text{hd}-\text{pd}}\right]\right\} \\
&= E\tilde{\psi}_{n,i}(E[Y_i(1)|W_{n,i}, X_i] - Y_i(1)) - E\tilde{\psi}_{n,i}R_i + E\tilde{\psi}_{n,i}\epsilon_{n,i}(1) \\
&= -E\tilde{\psi}_{n,i}R_i + E\tilde{\psi}_{n,i}\epsilon_{n,i}(1).
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
E\left[(\tilde{\psi}'_{n,i}(\beta_{1,n}^{*} - \beta_{1,n}^{\text{hd}-\text{pd}}))^2\right] &\leq \{E[(\tilde{\psi}_{n,i}(\beta_{1,n}^{*} - \beta_{1,n}^{\text{hd}-\text{pd}}))^2]\}^{1/2}\{ER_{n,i}^2\}^{1/2} + ||\beta_{1,n}^{*} - \beta_{1,n}^{\text{hd}-\text{pd}}||_1 \left\|E\tilde{\psi}_{n,i}\epsilon_{n,i}(1)\right\|_\infty \\
&= \{E[(\tilde{\psi}_{n,i}(\beta_{1,n}^{*} - \beta_{1,n}^{\text{hd}-\text{pd}}))^2]\}^{1/2} o(1) + o(1)\ ,
\end{aligned}$$

where we use the facts that $\{ER_{n,i}^2\}^{1/2} = o(1)$ and

$$||\beta_{1,n}^{*} - \beta_{1,n}^{\text{hd}-\text{pd}}||_1 \left\|E\tilde{\psi}_{n,i}\epsilon_{n,i}(1)\right\|_\infty = o(s_n^{\text{hd}-\text{pd}}\lambda_n) = o(1).$$

This implies $E\left(\tilde{\psi}'_{n,i}(\beta_{1,n}^{*} - \beta_{1,n}^{\text{hd}-\text{pd}}))^2\right] = o(1)$. Similarly, we can show $E\left[(\tilde{\psi}'_{n,i}(\beta_{0,n}^{*} - \beta_{0,n}^{\text{hd}-\text{pd}}))^2\right] = o(1)$, which implies

$$E\left[(E[\tilde{Y}_i(1) + \tilde{Y}_i(0)|W_{n,i}, X_i] - \tilde{\psi}'_{n,i}(\beta_{1,n}^{\text{hd}-\text{pd}} + \beta_{0,n}^{\text{hd}-\text{pd}}))^2\right] = o(1).$$

Last, we note that Assumption 3.1(a) and 3.1(b) follow Assumption 5.1. Assumption 3.1(c) follows Assumptions 5.1 and 5.2. This concludes the proof. ■

A.6 Proof of Theorem 5.2

We first show

$$|\hat{\alpha}^{\rm hd}_{d,n} - \alpha^{\rm hd}_{d,n}| + \|\hat{\beta}^{\rm hd}_{d,n} - \beta^{\rm hd}_{d,n}\|_1 = O_P\left(s^{\rm hd}_n \lambda^{\rm hd}_n\right) \tag{57}$$

To that end, note

$$\begin{aligned}
&\frac{1}{n}\sum_{1\le i\le 2n} I\{D_i = d\}(Y_i(d) - \hat{\alpha}^{\rm hd}_{d,n} - \psi'_{n,i}\hat{\beta}^{\rm hd}_{d,n})^2 + \lambda^{\rm hd}_{d,n}\|\hat{\Omega}_n(d)\hat{\beta}^{\rm hd}_{d,n}\|_1 \\
&\le \frac{1}{n}\sum_{1\le i\le 2n} I\{D_i = d\}(Y_i(d) - \alpha^{\rm hd}_{d,n} - \psi'_{n,i}\beta^{\rm hd}_{d,n})^2 + \lambda^{\rm hd}_{d,n}\|\hat{\Omega}_n(d)\beta^{\rm hd}_{d,n}\|_1 \ .
\end{aligned}$$

Rearranging the terms, we then have

$$\begin{aligned}
&\frac{1}{n}\sum_{1\le i\le 2n} I\{D_i = d\}(\hat{\alpha}^{\rm hd}_{d,n} - \alpha^{\rm hd}_{d,n} + \psi'_{n,i}(\hat{\beta}^{\rm hd}_{d,n} - \beta^{\rm hd}_{d,n}))^2 + \lambda^{\rm hd}_{d,n}\|\hat{\Omega}_n(d)\hat{\beta}^{\rm hd}_{d,n}\|_1 \\
&\le \left(\frac{2}{n}\sum_{1\le i\le 2n} I\{D_i = d\}\epsilon_{n,i}(d)\psi'_{n,i}\right)(\hat{\beta}^{\rm hd}_{d,n} - \beta^{\rm hd}_{d,n}) + \left(\frac{2}{n}\sum_{1\le i\le 2n} I\{D_i = d\}\epsilon_{n,i}(d)\right)(\hat{\alpha}^{\rm hd}_{d,n} - \alpha^{\rm hd}_{d,n}) \\
&+ \lambda^{\rm hd}_{d,n}\|\hat{\Omega}_n(d)\beta^{\rm hd}_{d,n}\|_1
\end{aligned} \tag{58}$$

Next, define

$$\mathbb{U}_n = \Omega_n^{-1}(d)\frac{1}{n}\sum_{1\le i\le 2n} I\{D_i = d\}(\psi_{n,i}\epsilon_{n,i}(d) - E[\psi_{n,i}\epsilon_{n,i}(d)])$$

and

$$\mathcal{E}_n(d) = \left\{\|\mathbb{U}_n\|_\infty \le \frac{6\bar{\sigma}}{\underline{\sigma}}\sqrt{\frac{\log(2np_n)}{n}}, \left|\frac{1}{n}\sum_{i\in[2n]} I\{D_i = d\}\epsilon_{n,i}(d) - E[\epsilon_{n,i}(d)]\right| \le \sqrt{\frac{\log(2np_n)}{n}}\right\} \ .$$

Lemma B.6 implies $P\{\mathcal{E}_n(d)\} \to 1$ for $d \in \{0, 1\}$.

On the event $\mathcal{E}_n(d)$, we have

$$\begin{aligned}
&\left|\left(\frac{2}{n}\sum_{1\le i\le 2n} I\{D_i = d\}\epsilon_{n,i}(d)\psi'_{n,i}\right)(\hat{\beta}^{\rm hd}_{d,n} - \beta^{\rm hd}_{d,n})\right| \\
&\le \left\|\Omega_n^{-1}(d)\frac{2}{n}\sum_{1\le i\le 2n} I\{D_i = d\}\epsilon_{n,i}(d)\psi_{n,i}\right\|_\infty \|\Omega_n(d)(\hat{\beta}^{\rm hd}_{d,n} - \beta^{\rm hd}_{d,n})\|_1 \\
&\le 2\|\mathbb{U}_n\|_\infty\|\Omega_n(d)(\hat{\beta}^{\rm hd}_{d,n} - \beta^{\rm hd}_{d,n})\|_1 + \left\|\Omega_n^{-1}(d)\frac{2}{n}\sum_{1\le i\le 2n} I\{D_i = d\}E[\epsilon_{n,i}(d)\psi_{n,i}]\right\|_\infty \|\Omega_n(d)(\hat{\beta}^{\rm hd}_{d,n} - \beta^{\rm hd}_{d,n})\|_1 \\
&\le 2\|\mathbb{U}_n\|_\infty\|\Omega_n(d)(\hat{\beta}^{\rm hd}_{d,n} - \beta^{\rm hd}_{d,n})\|_1 + \left\|\Omega_n^{*,-1}(d)2E[\epsilon_{n,i}(d)\psi_{n,i}]\right\|_\infty \|\Omega_n(d)(\hat{\beta}^{\rm hd}_{d,n} - \beta^{\rm hd}_{d,n})\|_1 \\
&\le \left(\frac{12\bar{\sigma}}{\underline{\sigma}\ell\ell_n} + d_n\right)\lambda^{\rm hd}_{d,n}\|\Omega_n(d)(\hat{\beta}^{\rm hd}_{d,n} - \beta^{\rm hd}_{d,n})\|_1 \ ,
\end{aligned}$$

where the second last inequality is by $\omega^*_{n,l}(d) \le \omega_{n,l}(d)$, $d_n = o(1)$, and the last inequality follows from (23) and the fact that

$$\lambda_{d,n}^{\mathrm{hd}} \geq \ell\ell_n \sqrt{\frac{\log(2np_n)}{n}} \ .$$

Next, define

$$\hat{\delta}_{d,n} = \hat{\beta}_{d,n}^{\mathrm{hd}} - \beta_{d,n}^{\mathrm{hd}}$$

and let $S_{d,n}$ be the support of $\beta_{d,n}^{\mathrm{hd}}$. Then, we have

$$\begin{aligned}
&\left(\frac{12\bar{\sigma}}{\underline{\sigma}\ell\ell_n} + d_n\right) \|\Omega_n(d)\hat{\delta}_{d,n}\|_1 + \|\hat{\Omega}_n(d)\beta_{d,n}^{\mathrm{hd}}\|_1 - \|\hat{\Omega}_n(d)\hat{\beta}_{d,n}^{\mathrm{hd}}\|_1 \\
&= \left(\frac{12\bar{\sigma}}{\underline{\sigma}\ell\ell_n} + d_n\right) \|(\Omega_n(d)\hat{\delta}_{d,n})_{S_{d,n}}\|_1 + \left(\frac{12\bar{\sigma}}{\underline{\sigma}\ell\ell_n} + d_n\right) \|(\Omega_n(d)\hat{\delta}_{d,n})_{S_{d,n}^c}\|_1 + \|\hat{\Omega}_n(d)\beta_{d,n}^{\mathrm{hd}}\|_1 - \|\hat{\Omega}_n(d)\hat{\beta}_{d,n}^{\mathrm{hd}}\|_1
\end{aligned}$$

and

$$\begin{aligned}
&\|\hat{\Omega}_n(d)\hat{\beta}_{d,n}^{\mathrm{hd}}\|_1 = \|(\hat{\Omega}_n(d)\hat{\beta}_{d,n}^{\mathrm{hd}})_{S_{d,n}}\|_1 + \|(\hat{\Omega}_n(d)\hat{\beta}_{d,n}^{\mathrm{hd}})_{S_{d,n}^c}\|_1 = \|(\hat{\Omega}_n(d)\hat{\beta}_{d,n}^{\mathrm{hd}})_{S_{d,n}}\|_1 + \|(\hat{\Omega}_n(d)\hat{\delta}_{d,n})_{S_{d,n}^c}\|_1 \\
&\|\hat{\Omega}_n(d)\beta_{d,n}^{\mathrm{hd}}\|_1 = \|(\hat{\Omega}_n(d)\beta_{d,n}^{\mathrm{hd}})_{S_{d,n}}\|_1 \leq \|(\hat{\Omega}_n(d)\hat{\beta}_{d,n}^{\mathrm{hd}})_{S_{d,n}}\|_1 + \|(\hat{\Omega}_n(d)\hat{\delta}_{d,n}^{\mathrm{hd}})_{S_{d,n}}\|_1
\end{aligned}$$

Further define $\breve{\delta}_{d,n} = (\hat{\alpha}_{d,n}^{\mathrm{hd}} - \alpha_{d,n}^{\mathrm{hd}}, \hat{\delta}_{d,n}')'$ and $\breve{S}_{d,n} = \{1, S_{d,n}+1\}$ as illustrated in the proof of Theorem 5.1 above and recall $\breve{\psi}_{n,i} = (1, \psi_{n,i}')'$. Then, together with (58), we have

$$\begin{aligned}
0 &\leq \frac{1}{n} \sum_{1 \leq i \leq 2n} I\{D_i = d\}(\breve{\psi}_{n,i}'\breve{\delta}_{d,n})^2 \\
&\leq \lambda_{d,n}^{\mathrm{hd}} \left[\left(\frac{12\bar{\sigma}}{\underline{\sigma}\ell\ell_n} + d_n + \bar{c}\right) \|(\Omega_n(d)\hat{\delta}_{d,n})_{S_{d,n}}\|_1 - \left(\underline{c} - \frac{12\bar{\sigma}}{\underline{\sigma}\ell\ell_n} - d_n\right) \|(\Omega_n(d)\hat{\delta}_{d,n})_{S_{d,n}^c}\|_1\right] \\
&+ (1/\ell_n + d_n)|\hat{\alpha}_{d,n}^{\mathrm{hd}} - \alpha_{d,n}^{\mathrm{hd}}| \\
&\leq \lambda_{d,n}^{\mathrm{hd}} \left[\left(\frac{12\bar{\sigma}}{\underline{\sigma}\ell\ell_n} + d_n + \bar{c}\right) \bar{\sigma}\|(\hat{\delta}_{d,n})_{S_{d,n}}\|_1 - \left(\underline{c} - \frac{12\bar{\sigma}}{\underline{\sigma}\ell\ell_n} - d_n\right) \underline{\sigma}\|(\hat{\delta}_{d,n})_{S_{d,n}^c}\|_1\right] \\
&+ (1/\ell_n + d_n)|\hat{\alpha}_{d,n}^{\mathrm{hd}} - \alpha_{d,n}^{\mathrm{hd}}| \\
&\leq \lambda_{d,n}^{\mathrm{hd}} \left[\left(\frac{12\bar{\sigma}}{\underline{\sigma}\ell\ell_n} + d_n + \bar{c}\right) \bar{\sigma}\|(\breve{\delta}_{d,n})_{\breve{S}_{d,n}}\|_1 - \left(\underline{c} - \frac{12\bar{\sigma}}{\underline{\sigma}\ell\ell_n} - d_n\right) \underline{\sigma}\|(\breve{\delta}_{d,n})_{\breve{S}_{d,n}^c}\|_1\right] \ .
\end{aligned}$$

Define

$$\mathcal{C}_n = \left\{u \in \mathbf{R}^{p_n+1} : \|u_{\breve{S}_{d,n}^c}\|_1 \leq \frac{2\bar{\sigma}\bar{c}}{\underline{\sigma}\underline{c}}\|u_{\breve{S}_{d,n}}\|_1\right\} \ .$$

For sufficiently large n, we have $\breve{\delta}_{d,n} \in \mathcal{C}_n$. It follows from Bickel et al. (2009) and Assumption 5.8 that

$$\inf_{u \in \mathcal{C}_n} (\|u_{\breve{S}_{d,n}}\|_1)^{-2}(s_n^{\mathrm{hd}} + 1)u' \left(\frac{1}{n} \sum_{1 \leq i \leq 2n} I\{D_i = d\}\breve{\psi}_{n,i}\breve{\psi}_{n,i}'\right) u \geq 0.25\kappa_1^2 \ .$$

Therefore, we have

$$0.25\kappa_1^2\|(\breve{\delta}_{d,n})_{\breve{S}_{d,n}}\|_1^2 \leq \left(\frac{12\bar{\sigma}}{\underline{\sigma}\ell\ell_n} + d_n + \bar{c}\right) \lambda_{d,n}^{\mathrm{hd}}(s_n^{\mathrm{hd}} + 1)\|(\breve{\delta}_{d,n})_{S_{d,n}}\|_1 \ ,$$

which implies

$$\|(\breve{\delta}_{d,n})_{S_{d,n}}\|_1 \le 4\left(\frac{12\bar{\sigma}}{\underline{\sigma}\ell\ell_n} + d_n + \bar{c}\right)(s_n^{\text{hd}}+1)\lambda_{d,n}^{\text{hd}}/\kappa_1^2 \ .$$

We then have

$$\begin{aligned}|\hat{\alpha}_{d,n}^{\text{hd}} - \alpha_{d,n}^{\text{hd}}| + \|\hat{\beta}_{d,n}^{\text{hd}} - \beta_{d,n}^{\text{hd}}\|_1 \le \|(\breve{\delta}_{d,n})_{\breve{S}_{d,n}}\|_1 + \|(\breve{\delta}_{d,n})_{\breve{S}_{d,n}^c}\|_1 \le 4\|(\breve{\delta}_{d,n})_{\breve{S}_{d,n}}\|_1 \\ \le 16\left(\frac{12\bar{\sigma}}{\underline{\sigma}\ell\ell_n} + d_n + \bar{c}\right)(s_n^{\text{hd}}+1)\lambda_{d,n}^{\text{hd}}/\kappa_1^2\end{aligned}$$

and

$$\frac{1}{n}\sum_{1\le i\le 2n} I\{D_i = d\}(\hat{\alpha}_{d,n}^{\text{hd}} - \alpha_{d,n}^{\text{hd}} + \psi_{n,i}'\hat{\delta}_{d,n})^2 \le 6\left(\frac{12\bar{\sigma}}{\underline{\sigma}\ell\ell_n} + d_n + \bar{c}\right)s_n^{\text{hd}}(\lambda_{d,n}^{\text{hd}})^2/\kappa_1^2 \ .$$

(57) follows because $P\{\mathcal{E}_n(d)\} \to 1$. (9) also follows because

$$\frac{1}{n}\sum_{1\le i\le 2n} I\{D_i = d\}(\hat{\alpha}_{d,n}^{\text{hd}} - \alpha_{d,n}^{\text{hd}} + \psi_{n,i}'(\hat{\beta}_{d,n} - \beta_{d,n}))^2 = O_P\left(s_n^{\text{hd}}(\lambda_n^{\text{hd}})^2\right) = o_P(1) \ .$$

Next, we show (6) for $\hat{\beta}_{d,n}^{\text{hd}}$. First note

$$\begin{aligned}&\left|\frac{1}{\sqrt{2n}}\sum_{1\le i\le 2n}(2D_i - 1)(\hat{m}_{d,n}(X_i, W_{n,i}) - m_{d,n}(X_i, W_{n,i}))\right| \\ &\left|\frac{1}{\sqrt{2n}}\sum_{1\le i\le 2n}(2D_i - 1)\psi_{n,i}'(\hat{\beta}_{d,n}^{\text{hd}} - \beta_{d,n}^{\text{hd}})\right| \\ &\le \left\|\frac{1}{\sqrt{2n}}\sum_{1\le i\le n}(2D_i - 1)\psi_{n,i}\right\|_\infty \|\hat{\beta}_{d,n}^{\text{hd}} - \beta_{d,n}^{\text{hd}}\|_1 \ .\end{aligned}$$

Next, note that it follows from Assumption 2.2 that conditional on $X^{(n)}$ and $W_n^{(n)}$,

$$\{D_{\pi(2j-1)} - D_{\pi(2j)} : 1 \le j \le n\}$$

is a sequence of independent Rademacher random variables. Therefore, Hoeffding's inequality implies

$$\begin{aligned}&P\left\{\left\|\frac{1}{\sqrt{2n}}\sum_{1\le i\le 2n}(2D_i - 1)\psi_{n,i}\right\|_\infty > t \,\middle|\, X^{(n)}, W_n^{(n)}\right\} \\ &\le \sum_{1\le l\le p_n} P\left\{\left|\frac{1}{\sqrt{2n}}\sum_{1\le j\le n}(\psi_{n,\pi(2j-1)} - \psi_{n,\pi(2j)})(D_{\pi(2j-1)} - D_{\pi(2j)})\right| > t \,\middle|\, X^{(n)}, W_n^{(n)}\right\} \\ &\le \sum_{1\le l\le p_n} 2\exp\left(-\frac{t^2}{\frac{1}{n}\sum_{1\le j\le n}(\psi_{n,\pi(2j-1)} - \psi_{n,\pi(2j)})^2}\right) \ .\end{aligned}$$

Define

$$\nu_n^2 = \max_{1\le l\le p_n}\frac{1}{n}\sum_{1\le i\le 2n}\psi_{n,i,l}^2 \ .$$

We then have

$$P\left\{\left\|\frac{1}{\sqrt{2n}}\sum_{1\le i\le 2n}(2D_i-1)\psi_{n,i}\right\|_\infty > \nu_n\sqrt{2\log(p_n\vee n)}\middle|X^{(n)},W_n^{(n)}\right\}\le (p_n\vee n)^{-1}\ . \tag{59}$$

Next, we determine the order of ν_n^2. Note

$$\begin{aligned}
E[\nu_n^2] &\le \max_{1\le l\le p_n} 2E[\psi_{n,i,l}^2] + 2E\left[\frac{1}{2n}\sum_{1\le i\le 2n}(\psi_{n,i,l}^2 - E[\psi_{n,i,l}^2])\right] \\
&\lesssim 1 + E\left[\max_{1\le l\le p_n}\left|\frac{1}{2n}\sum_{1\le i\le 2n} e_i\psi_{n,i,l}^2\right|\right] \\
&\lesssim 1 + \Xi_n E\left[\max_{1\le l\le p_n}\left|\frac{1}{2n}\sum_{1\le i\le 2n} e_i\psi_{n,i,l}\right|\right] \\
&\lesssim 1 + \Xi_n E\left[\max_{1\le l\le p_n}\left|\frac{1}{2n}\sum_{1\le i\le 2n} e_i(\psi_{n,i,l} - E[\psi_{n,i,l}])\right|\right] + \max_{1\le l\le p_n}|E[\psi_{n,i,l}]| \\
&\lesssim 1 + \Xi_n E\left[\sup_{f\in\mathcal{F}_n}\left|\frac{1}{2n}\sum_{1\le i\le 2n} f(e_i,\psi_{n,i,l})\right|\right]
\end{aligned}$$

where $\{e_i : 1\le i\le n\}$ is an i.i.d. sequence of Rademacher random variables,

$$\mathcal{F}_n = \{f : \mathbf{R}\times\mathbf{R}^{p_n}\mapsto\mathbf{R}, f(e,\psi) = e\psi_l, 1\le l\le p_n\}\ ,$$

and ψ_l is the lth element of ψ. Note the second inequality follows from Lemma 2.3.1 of van der Vaart and Wellner (1996), the third inequality follows from Theorem 4.12 of **?** and the definition of Ξ_n, and the last follows from Assumption 5.5. Note also $\mathcal{F}_n$ has an envelope $F = \Xi_n$ and

$$\sup_{n\ge 1}\sup_{f\in\mathcal{F}_n} E[f^2] < \infty$$

because of Assumption 5.5. Because the cardinality of $\mathcal{F}_n$ is p_n, for any $\epsilon < 1$ we have that

$$\sup_{Q:Q\text{ is a discrete distribution with finite support}} \mathcal{N}(\epsilon\|F\|_{Q,2},\mathcal{F},L_2(Q)) \le \frac{p_n}{\epsilon}\ ,$$

where $\mathcal{N}(\epsilon,\mathcal{F},L_2(Q))$ is the covering number for class $\mathcal{F}$ under the metric $L_2(Q)$ using balls of radius ϵ. Therefore, Corollary 5.1 of Chernozhukov et al. (2014) implies

$$E\left[\sup_{f\in\mathcal{F}_n}\left|\frac{1}{2n}\sum_{1\le i\le 2n} e_i(\psi_{n,i,l} - E[\psi_{n,i,l}])\right|\right] \lesssim \sqrt{\frac{\log p_n}{n}} + \frac{\Xi_n\log p_n}{n} = o(\Xi_n^{-1})\ .$$

Therefore, $\nu_n = O_p(1)$. Together with (59), they imply

$$\left\| \frac{1}{\sqrt{2n}} \sum_{1 \le i \le 2n} (2D_i - 1)\psi_{n,i} \right\|_\infty = O_P\left(\sqrt{\log(p_n \vee n)}\right) .$$

In light of (57) and Assumption 5.7, we have

$$\left\| \frac{1}{\sqrt{2n}} \sum_{1 \le i \le 2n} (2D_i - 1)\psi_{n,i} \right\|_\infty \|\hat{\beta}^{\rm hd}_{d,n} - \beta^{\rm hd}_{d,n}\|_1 = O_P\left(\frac{s_n^{\rm hd}\ell\ell_n \log(p_n \vee n)}{\sqrt{n}}\right) = o_P(1) .$$

Next, note that Assumption 3.1(a) and 3.1(b) follow Assumption 5.5, and Assumption 3.1(c) follows Assumptions 5.5 and 5.6.

Last, suppose the true specification is approximately sparse as specified in Theorem 5.2. Let $\tilde{Y}_i(d) = Y_i(d) - \mu_d(X_i)$, $\tilde{\psi}_{n,i} = \psi_{n,i} - E[\psi_{n,i}|X_i]$, and $\tilde{R}_{n,i}(d) = R_{n,i}(d) - E[R_{n,i}(d)|X_i]$. Then, we have

$$E\left[(E[\tilde{Y}_i(1) + \tilde{Y}_i(0)|W_{n,i}, X_i] - \tilde{\psi}'_{n,i}(\beta^{\rm hd}_{1,n} + \beta^{\rm hd}_{0,n}))^2\right] = E[(\tilde{R}_{n,i}(1) + \tilde{R}_{n,i}(0))^2] = o(1) .$$

This concludes the proof. ■

A.7 Proof of Theorem 5.3

Further denote $\tilde{\Delta}^{\rm hd}_n$ as the estimator in (12) with ψ_i replaced by $\Gamma_{n,i}$. We first show

$$\hat{\beta}^{\rm hd-f}_n - \beta^{\rm hd-f}_n = o_P(1) . \tag{60}$$

Let

$$\begin{aligned}
\hat{\Delta}_{\Gamma,n} &= \frac{1}{n} \sum_{1 \le j \le n} (D_{\pi(2j-1)} - D_{\pi(2j)})(\Gamma_{n,\pi(2j-1)} - \Gamma_{n,\pi(2j)}) , \\
\hat{\Delta}_{\hat{\Gamma},n} &= \frac{1}{n} \sum_{1 \le j \le n} (D_{\pi(2j-1)} - D_{\pi(2j)})(\hat{\Gamma}_{n,\pi(2j-1)} - \hat{\Gamma}_{n,\pi(2j)}) , \\
\delta_{\Gamma,j} &= (D_{\pi(2j-1)} - D_{\pi(2j)})(\Gamma_{n,\pi(2j-1)} - \Gamma_{n,\pi(2j)}) , \\
\delta_{\hat{\Gamma},j} &= (D_{\pi(2j-1)} - D_{\pi(2j)})(\hat{\Gamma}_{n,\pi(2j-1)} - \hat{\Gamma}_{n,\pi(2j)}) .
\end{aligned}$$

Then, by the proof of Theorem 4.2, we have $\hat{\beta}^{\rm hd-f}_n$ equals the coefficient estimate using least squares in the linear regression of $\delta_{Y,j}$ on $\delta_{\hat{\Gamma},j} - \hat{\Delta}_{\hat{\Gamma},n}$. Then, for any $u \in \mathbf{R}^2$ such that $||u||_2 = 1$, we have

$$\begin{aligned}
&\left| \left(\frac{1}{n} \sum_{1 \le j \le n} ((\delta_{\hat{\Gamma},j} - \hat{\Delta}_{\hat{\Gamma},n})'u)^2\right)^{1/2} - \left(\frac{1}{n} \sum_{1 \le j \le n} ((\delta_{\Gamma,j} - \hat{\Delta}_{\Gamma,n})'u)^2\right)^{1/2} \right| \\
&\le \left(\frac{1}{n} \sum_{1 \le j \le n} ((\delta_{\hat{\Gamma},j} - \delta_{\Gamma,j})'u)^2 - ((\hat{\Delta}_{\hat{\Gamma},n} - \hat{\Delta}_{\Gamma,n})'u)^2\right)^{1/2}
\end{aligned}$$

$$
\leq \left(\frac{2}{n} \sum_{1 \leq i \leq 2n} \left\| \hat{\Gamma}_{n,i} + (\hat{\alpha}_{1,n}^{\mathrm{hd}}, \hat{\alpha}_{0,n}^{\mathrm{hd}})' - \Gamma_{n,i} - (\alpha_{1,n}^{\mathrm{hd}}, \alpha_{0,n}^{\mathrm{hd}})' \right\|_2^2 \right)^{1/2}
$$

$$
\lesssim \sum_{d \in \{0,1\}} \frac{1}{2n} \sum_{1 \leq i \leq 2n} (\hat{\alpha}_{d,n}^{\mathrm{hd}} - \alpha_{d,n}^{\mathrm{hd}} + \psi_{n,i}'(\hat{\beta}_{d,n}^{\mathrm{hd}} - \beta_{d,n}^{\mathrm{hd}}))^2 = o_P(1) ,
$$

where the second inequality is by the fact that

$$
\delta_{\Gamma,j} = (D_{\pi(2j-1)} - D_{\pi(2j)})(\Gamma_{n,\pi(2j-1)} + (\alpha_{1,n}^{\mathrm{hd}}, \alpha_{0,n}^{\mathrm{hd}})' - \Gamma_{n,\pi(2j)} - (\alpha_{1,n}^{\mathrm{hd}}, \alpha_{0,n}^{\mathrm{hd}})') ,
$$

$$
\delta_{\hat{\Gamma},j} = (D_{\pi(2j-1)} - D_{\pi(2j)})(\hat{\Gamma}_{n,\pi(2j-1)} + (\hat{\alpha}_{1,n}^{\mathrm{hd}}, \hat{\alpha}_{0,n}^{\mathrm{hd}})' - \hat{\Gamma}_{n,\pi(2j)} - (\hat{\alpha}_{1,n}^{\mathrm{hd}}, \hat{\alpha}_{0,n}^{\mathrm{hd}})') ,
$$

and the last equality is by the proof of Theorem 5.2. This implies

$$
\begin{aligned}
&\frac{1}{n} \sum_{1 \leq j \leq n} (\delta_{\hat{\Gamma},j} - \hat{\Delta}_{\hat{\Gamma},n})(\delta_{\hat{\Gamma},j} - \hat{\Delta}_{\hat{\Gamma},n})' - 2E[\mathrm{Var}[\Gamma_{n,i}|X_i]] \\
&= \frac{1}{n} \sum_{1 \leq j \leq n} (\delta_{\hat{\Gamma},j} - \hat{\Delta}_{\hat{\Gamma},n})(\delta_{\hat{\Gamma},j} - \hat{\Delta}_{\hat{\Gamma},n})' - \frac{1}{n} \sum_{1 \leq j \leq n} (\delta_{\Gamma,j} - \hat{\Delta}_{\Gamma,n})(\delta_{\Gamma,j} - \hat{\Delta}_{\Gamma,n})' \\
&+ \frac{1}{n} \sum_{1 \leq j \leq n} (\delta_{\Gamma,j} - \hat{\Delta}_{\Gamma,n})(\delta_{\Gamma,j} - \hat{\Delta}_{\Gamma,n})' - 2E[\mathrm{Var}[\Gamma_{n,i}|X_i]] = o_P(1) ,
\end{aligned}
$$

where the last equality holds due to the same argument as used in the proof of Theorem 4.2. Similarly, we can show that

$$
\frac{1}{n} \sum_{1 \leq j \leq n} \delta_{Y,j}(\delta_{\hat{\Gamma},j} - \hat{\Delta}_{\hat{\Gamma},n}) - E[\mathrm{Cov}[\Gamma_{n,i}, Y_i(1) + Y_i(0)|X_i]] = o_P(1) ,
$$

which leads to (60).

Next, we show (6). We have

$$
\begin{aligned}
&\frac{1}{\sqrt{2n}} \sum_{1 \leq i \leq 2n} (2D_i - 1)(\hat{m}_{d,n}(X_i, W_{n,i}) - m_{d,n}(X_i, W_{n,i})) \\
&= \frac{1}{\sqrt{2n}} \sum_{1 \leq i \leq 2n} (2D_i - 1)(\hat{\Gamma}_{n,i} - \Gamma_{n,i})' \hat{\beta}_n^{\mathrm{hd-f}} + \frac{1}{\sqrt{2n}} \sum_{1 \leq i \leq 2n} (2D_i - 1)\Gamma_{n,i}'(\hat{\beta}_n^{\mathrm{hd-f}} - \beta_n^{\mathrm{hd-f}}) \\
&= \left(\tfrac{1}{\sqrt{2n}} \textstyle\sum_{1 \leq i \leq 2n} (2D_i - 1)\psi_{n,i}'(\hat{\beta}_{1,n}^{\mathrm{hd}} - \beta_{1,n}^{\mathrm{hd}}), \tfrac{1}{\sqrt{2n}} \sum_{1 \leq i \leq 2n} (2D_i - 1)\psi_{n,i}'(\hat{\beta}_{0,n}^{\mathrm{hd}} - \beta_{0,n}^{\mathrm{hd}}) \right) \hat{\beta}_n^{\mathrm{hd-f}} \\
&+ \left(\tfrac{1}{\sqrt{2n}} \textstyle\sum_{1 \leq i \leq 2n} (2D_i - 1)\psi_{n,i}'\beta_{1,n}^{\mathrm{hd}}, \tfrac{1}{\sqrt{2n}} \sum_{1 \leq i \leq 2n} (2D_i - 1)\psi_{n,i}'\beta_{0,n}^{\mathrm{hd}} \right) (\hat{\beta}_{1,n}^{\mathrm{hd-f}} - \beta_{1,n}^{\mathrm{hd-f}}) \\
&= o_P(1) ,
\end{aligned}
$$

where the last equality holds by (60) and the facts that

$$
\left(\tfrac{1}{\sqrt{2n}} \textstyle\sum_{1 \leq i \leq 2n} (2D_i - 1)\psi_{n,i}'(\hat{\beta}_{1,n}^{\mathrm{hd}} - \beta_{1,n}^{\mathrm{hd}}), \tfrac{1}{\sqrt{2n}} \sum_{1 \leq i \leq 2n} (2D_i - 1)\psi_{n,i}'(\hat{\beta}_{0,n}^{\mathrm{hd}} - \beta_{0,n}^{\mathrm{hd}}) \right) = o_P(1)
$$

as shown in Theorem 5.2 and

$$
\left(\tfrac{1}{\sqrt{2n}} \textstyle\sum_{1 \leq i \leq 2n} (2D_i - 1)\psi_{n,i}'\beta_{1,n}^{\mathrm{hd}}, \tfrac{1}{\sqrt{2n}} \sum_{1 \leq i \leq 2n} (2D_i - 1)\psi_{n,i}'\beta_{0,n}^{\mathrm{hd}} \right) = O_P(1) .
$$

For (9), we note that

$$
\begin{aligned}
&\frac{1}{2n}\sum_{1\le i\le 2n}(\hat{m}_{d,n}(X_i,W_{n,i})-m_{d,n}(X_i,W_{n,i}))^2\\
&\lesssim \frac{1}{2n}\sum_{1\le i\le 2n}((\hat{\Gamma}_{n,i}-\Gamma_{n,i})'\hat{\beta}_n^{\text{hd}-\text{f}})^2+\frac{1}{2n}\sum_{1\le i\le 2n}(\Gamma_{n,i}'(\hat{\beta}_n^{\text{hd}-\text{f}}-\beta_n^{\text{hd}-\text{f}}))^2\\
&\lesssim \frac{1}{2n}\sum_{1\le i\le 2n}((\psi_{n,i}'(\beta_{1,n}^{\text{hd}}-\hat{\beta}_{1,n}^{\text{hd}}))^2+(\psi_{n,i}'(\beta_{0,n}^{\text{hd}}-\hat{\beta}_{0,n}^{\text{hd}}))^2)||\hat{\beta}_n^{\text{hd}-\text{f}}||_2^2\\
&+\frac{1}{2n}\sum_{1\le i\le 2n}[(\psi_{n,i}'\beta_{1,n}^{\text{hd}})^2+(\psi_{n,i}'\beta_{0,n}^{\text{hd}})^2]||\hat{\beta}_n^{\text{hd}-\text{f}}-\beta_n^{\text{hd}-\text{f}}||_2^2\\
&\lesssim \sum_{d=0,1}\frac{1}{2n}\sum_{1\le i\le 2n}\left[(\alpha_{d,n}^{\text{hd}}-\hat{\alpha}_{d,n}^{\text{hd}}+\psi_{n,i}'(\beta_{d,n}^{\text{hd}}-\hat{\beta}_{d,n}^{\text{hd}}))^2+(\alpha_{d,n}^{\text{hd}}-\hat{\alpha}_{d,n}^{\text{hd}})^2\right]||\hat{\beta}_n^{\text{hd}-\text{f}}||_2^2+o_P(1)\\
&=o_P(1)\ .
\end{aligned}
$$

Assumption 3.1 can be verified in the same manner as we did in the proof of Theorem 5.2.

Last, we compare $\sigma_n^{\text{na},2}$, $\sigma_n^{\text{hd},2}$, $\sigma_n^{\text{hd}-\text{f},2}$. Recall $\sigma_2^2(Q)$ and $\sigma_3^2(Q)$ defined in Theorem 3.1. As we have already verified (6) for $\hat{m}_{d,n}(X_i,W_{n,i})=\hat{\Gamma}_{n,i}\hat{\beta}_n^{\text{hd}-\text{f}}$ and $m_{d,n}(X_i,W_{n,i})=\Gamma_{n,i}\beta_n^{\text{hd}-\text{f}}$, we have, for $b\in\{\text{na},\text{hd},(\text{pfe},\text{hd})\}$, that

$$
\sigma_n^{\text{b},2}-\sigma_2^2(Q)-\sigma_3^2(Q)=\frac{1}{2}E\left[\text{Var}[E[Y_i(1)+Y_i(0)|X_i,W_{n,i}]-\Gamma_{n,i}'\gamma^b|X_i]\right]
$$

with

$$
\gamma^{\text{unadj}}=(0,0)'\ ,\quad \gamma^{\text{hd}}=(1,1)'\ ,\quad \text{and}\quad \gamma^{\text{hd}-\text{f}}=\beta_n^{\text{hd}-\text{f}}\ .
$$

In addition, we note that

$$
\frac{1}{2}E\left[\text{Var}[E[Y_i(1)+Y_i(0)|X_i,W_{n,i}]-\Gamma_{n,i}'\gamma|X_i]\right]
$$

is minimized at $\gamma=\beta_n^{\text{hd}-\text{f}}$, which leads to the desired result. ■

B Auxiliary Lemmas

Lemma B.1. *Suppose $\phi_n, n\ge 1$ is a sequence of random variables satisfying*

$$
\lim_{\lambda\to\infty}\limsup_{n\to\infty}E[|\phi_n|I\{|\phi_n|>\lambda\}]=0\ . \tag{61}
$$

Suppose X is another random variable defined on the same probability space with $\phi_n, n\ge 1$. Then,

$$
\lim_{\gamma\to\infty}\limsup_{n\to\infty}E[E[|\phi_n||X]I\{E[|\phi_n||X]>\gamma\}]=0\ . \tag{62}
$$

PROOF. Fix $\epsilon > 0$. We will show there exists $\gamma > 0$ so that

$$\limsup_{n\to\infty} E[E[|\phi_n||X]I\{E[|\phi_n||X] > \gamma\}] < \epsilon \ . \tag{63}$$

First note the event $\{E[|\phi_n||X] > \gamma\}$ is measurable with respect to the σ-algebra generated by X, and therefore

$$E[E[|\phi_n||X]I\{E[|\phi_n||X] > \gamma\}] = E[|\phi_n|I\{E[|\phi_n||X] > \gamma\}] \ . \tag{64}$$

Next, by Theorem 10.3.5 of Dudley (1989), (61) implies that there exists a $\delta > 0$ such that for any sequence of events A_n such that $\limsup_{n\to\infty} P\{A_n\} < \delta$, we have

$$\limsup_{n\to\infty} E[|\phi_n|I\{A_n\}] < \epsilon \ . \tag{65}$$

In light of the previous result, note

$$P\{E[|\phi_n||X] > \gamma\} \leq \frac{E[E[|\phi_n||X]]}{\gamma} = \frac{E[|\phi_n|]}{\gamma}$$

By Theorem 10.3.5 of Dudley (1989) again, (61) implies $\limsup_{n\to\infty} E[|\phi_n|] < \infty$, so by choosing γ large enough, we can make sure

$$\limsup_{n\to\infty} P\{E[|\phi_n||X] > \gamma\} < \delta \text{ for all } n \ .$$

(63) then follows from (64)–(65). ■

Lemma B.2. *Suppose Assumptions 2.1–2.3 and 3.1 hold. Then,*

$$\frac{s_n^2}{nE[\mathrm{Var}[\phi_{1,n,i}|X_i]]} \xrightarrow{P} 1 \ .$$

PROOF. To begin, note it follows from Assumption 2.2 and $Q_n = Q^{2n}$ that

$$\frac{1}{n}\sum_{1\leq i\leq 2n} D_i \mathrm{Var}[\phi_{1,n,i}|X_i] = \frac{1}{2n}\sum_{1\leq i\leq 2n} \mathrm{Var}[\phi_{1,n,i}|X_i] + \frac{1}{2n}\sum_{1\leq i\leq 2n: D_i=1} \mathrm{Var}[\phi_{1,n,i}|X_i] - \frac{1}{2n}\sum_{1\leq i\leq 2n: D_i=0} \mathrm{Var}[\phi_{1,n,i}|X_i] \ . \tag{66}$$

Next,

$$\left|\frac{1}{2n}\sum_{1\leq i\leq 2n: D_i=1} \mathrm{Var}[\phi_{1,n,i}|X_i] - \frac{1}{2n}\sum_{1\leq i\leq 2n: D_i=0} \mathrm{Var}[\phi_{1,n,i}|X_i]\right| \leq \frac{1}{2n}\sum_{1\leq j\leq n} |\mathrm{Var}[\phi_{1,n,\pi(2j-1)}|X_{\pi(2j-1)}] - \mathrm{Var}[\phi_{1,n,\pi(2j)}|X_{\pi(2j)}]| \ . \tag{67}$$

In what follows, we will show

$$\frac{1}{n}\sum_{1\leq j\leq n} |\mathrm{Cov}[Y_{\pi(2j-1)}(1), m_{1,n}(X_{\pi(2j-1)}, W_{\pi(2j-1)})|X_{\pi(2j-1)}]]$$

$$- \text{Cov}[Y_{\pi(2j)}(1), m_{1,n}(X_{\pi(2j)}, W_{\pi(2j)})|X_{\pi(2j)}]]| \xrightarrow{P} 0 \ .$$

To that end, first note from Assumptions 2.3 and 3.1(c) that

$$\begin{aligned}
&\frac{1}{n}\sum_{1\le j\le n}|E[Y_{\pi(2j-1)}(1)m_{1,n}(X_{\pi(2j-1)}, W_{\pi(2j-1)})|X_{\pi(2j-1)}] - E[Y_{\pi(2j)}(1)m_{1,n}(X_{\pi(2j)}, W_{\pi(2j)})|X_{\pi(2j)}]| \\
&\lesssim \frac{1}{n}\sum_{1\le j\le n}|X_{\pi(2j-1)} - X_{\pi(2j)}| \xrightarrow{P} 0 \ .
\end{aligned}$$

Next, note

$$\begin{aligned}
&\frac{1}{n}\sum_{1\le j\le n}|E[Y_{\pi(2j-1)}(1)|X_{\pi(2j-1)}]E[m_{1,n}(X_{\pi(2j-1)}, W_{\pi(2j-1)})|X_{\pi(2j-1)}] \\
&\qquad - E[Y_{\pi(2j)}(1)|X_{\pi(2j)}]E[m_{1,n}(X_{\pi(2j)}, W_{\pi(2j)})|X_{\pi(2j)}]| \\
&\le \frac{1}{n}\sum_{1\le j\le n}|E[Y_{\pi(2j-1)}(1)|X_{\pi(2j-1)}]||E[m_{1,n}(X_{\pi(2j-1)}, W_{\pi(2j-1)})|X_{\pi(2j-1)}] - E[m_{1,n}(X_{\pi(2j)}, W_{\pi(2j)})|X_{\pi(2j)}]| \\
&\qquad + \frac{1}{n}\sum_{1\le j\le n}|E[Y_{\pi(2j-1)}(1)|X_{\pi(2j-1)}] - E[Y_{\pi(2j)}(1)|X_{\pi(2j)}]||E[m_{1,n}(X_{\pi(2j)}, W_{\pi(2j)})|X_{\pi(2j)}]| \\
&\le \left(\frac{1}{n}\sum_{1\le j\le n}|E[Y_{\pi(2j-1)}(1)|X_{\pi(2j-1)}]|^2\right)^{1/2} \\
&\qquad\qquad \times \left(\frac{1}{n}\sum_{1\le j\le n}|E[m_{1,n}(X_{\pi(2j-1)}, W_{\pi(2j-1)})|X_{\pi(2j-1)}] - E[m_{1,n}(X_{\pi(2j)}, W_{\pi(2j)})|X_{\pi(2j)}]|^2\right)^{1/2} \\
&\qquad + \left(\frac{1}{n}\sum_{1\le j\le n}|E[m_{1,n}(X_{\pi(2j)}, W_{\pi(2j)})|X_{\pi(2j)}]|^2\right)^{1/2} \\
&\qquad\qquad \times \left(\frac{1}{n}\sum_{1\le j\le n}|E[Y_{\pi(2j-1)}(1)|X_{\pi(2j-1)}] - E[Y_{\pi(2j)}(1)|X_{\pi(2j)}]|^2\right)^{1/2} \\
&\lesssim \left(\frac{1}{n}\sum_{1\le i\le 2n}|E[Y_i(1)|X_i]|^2\right)^{1/2}\left(\frac{1}{n}\sum_{1\le j\le n}|X_{\pi(2j-1)} - X_{\pi(2j)}|^2\right)^{1/2} \\
&\qquad + \left(\frac{1}{n}\sum_{1\le i\le 2n}|E[m_{1,n}(X_i, W_i)|X_i]|^2\right)^{1/2}\left(\frac{1}{n}\sum_{1\le j\le n}|X_{\pi(2j-1)} - X_{\pi(2j)}|^2\right)^{1/2} \xrightarrow{P} 0 \ ,
\end{aligned}$$

where the first inequality follows from the triangle inequality, the second follows from the Cauchy-Schwarz inequality, the last follows from Assumptions 2.1(c) and 3.1(c). To see the convergence holds, first note because

$$E[|E[Y_i(1)|X_i]|^2] \le E[E[Y_i^2(1)|X_i]] = E[Y_i^2(1)] < \infty \ ,$$

the weak law of large numbers implies

$$\frac{1}{n}\sum_{1\le i\le 2n}|E[Y_i(1)|X_i]|^2 \xrightarrow{P} 2E[|E[Y_i(1)|X_i]|^2] < \infty \ .$$

On the other hand,

$$\frac{1}{2n}\sum_{1\le i\le 2n}|E[m_{1,n}(X_i,W_i)|X_i]|^2\le\frac{1}{2n}\sum_{1\le i\le 2n}E[m_{1,n}^2(X_i,W_i)|X_i]\ .$$

Assumption 3.1(b) and Lemma B.1 imply

$$\lim_{\lambda\to\infty}\limsup_{n\to\infty}E[E[m_{1,n}^2(X_i,W_i)|X_i]I\{E[m_{1,n}^2(X_i,W_i)|X_i]>\lambda\}]=0\ .$$

Therefore, Lemma 11.4.2 of Lehmann and Romano (2005) implies

$$\frac{1}{2n}\sum_{1\le i\le 2n}E[m_{1,n}^2(X_i,W_i)|X_i]-E[E[m_{1,n}^2(X_i,W_i)|X_i]]\xrightarrow{P}0\ .$$

Finally, note $E[E[m_{1,n}^2(X_i,W_i)|X_i]]=E[m_{1,n}^2(X_i,W_i)]$ is bounded for $n\ge 1$ by Assumption 3.1(b), so

$$\frac{1}{n}\sum_{1\le i\le 2n}|E[m_{1,n}(X_i,W_i)|X_i]|^2=O_P(1)\ .$$

The desired convergence therefore follows.

Similar arguments applied termwise imply the right-hand side of (67) is $o_P(1)$. (66)–(67) then imply

$$\frac{s_n^2}{n}-\frac{1}{2n}\sum_{1\le i\le 2n}\operatorname{Var}[\phi_{1,n,i}|X_i]\to 0\ . \tag{68}$$

Next, we argue

$$\frac{1}{2n}\sum_{1\le i\le 2n}\operatorname{Var}[\phi_{1,n,i}|X_i]-E[\operatorname{Var}[\phi_{1,n,i}|X_i]]\to 0\ . \tag{69}$$

To establish (69), we verify the uniform integrability condition in Lemma 11.4.2 of Lehmann and Romano (2005). To that end, we will repeatedly use the inequality

$$\left|\sum_{1\le j\le k}a_j\right|I\left\{\left|\sum_{1\le j\le k}a_j\right|>\lambda\right\}\le\sum_{1\le j\le k}k|a_j|I\left\{|a_j|>\frac{\lambda}{k}\right\} \tag{70}$$

$$|ab|I\{|ab|>\lambda\}\le|a|^2I\{|a|>\sqrt{\lambda}\}+|b|^2I\{|b|>\sqrt{\lambda}\}\ . \tag{71}$$

Note

$$\begin{aligned}
&E[|\operatorname{Var}[\phi_{1,n,i}|X_i]-E[\operatorname{Var}[\phi_{1,n,i}|X_i]]|I\{|\operatorname{Var}[\phi_{1,n,i}|X_i]-E[\operatorname{Var}[\phi_{1,n,i}|X_i]]|>\lambda\}]\\
&\lesssim E\left[|\operatorname{Var}[\phi_{1,n,i}|X_i]|I\left\{|\operatorname{Var}[\phi_{1,n,i}|X_i]|>\frac{\lambda}{2}\right\}\right]+E[\operatorname{Var}[\phi_{1,n,i}|X_i]]I\left\{E[\operatorname{Var}[\phi_{1,n,i}|X_i]]>\frac{\lambda}{2}\right\}\\
&\le E\left[E[\phi_{1,n,i}^2|X_i]I\left\{E[\phi_{1,n,i}^2|X_i]>\frac{\lambda}{2}\right\}\right]+E[\phi_{1,n,i}^2]I\left\{E[\phi_{1,n,i}^2]>\frac{\lambda}{2}\right\}\ ,
\end{aligned}$$

where in the second inequality we use the fact that the variance of a random variable is bounded by its second moment. Note Assumption 3.1 implies $E[\phi^2_{1,n,i}]$ is bounded for $n \geq 1$, and therefore

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} E[\phi^2_{1,n,i}] I\left\{E[\phi^2_{1,n,i}] > \frac{\lambda}{2}\right\} = 0 \ .$$

On the other hand

$$\begin{aligned} & E\left[E[\phi^2_{1,n,i}|X_i] I\left\{E[\phi^2_{1,n,i}|X_i] > \frac{\lambda}{2}\right\}\right] \qquad (72) \\ & \lesssim E\left[E[Y_i^2(1)|X_i] I\left\{E[Y_i^2(1)|X_i] > \frac{\lambda}{12}\right\}\right] + E\left[E[m^2_{1,n}(X_i,W_i)|X_i] I\left\{E[m^2_{1,n}(X_i,W_i)|X_i] > \frac{\lambda}{3}\right\}\right] \\ & \quad + E\left[E[m^2_{0,n}(X_i,W_i)|X_i] I\left\{E[m^2_{0,n}(X_i,W_i)|X_i] > \frac{\lambda}{3}\right\}\right] \\ & \quad + E\left[|E[Y_i(1) m_{1,n}(X_i,W_i)|X_i]| I\left\{|E[Y_i(1) m_{1,n}(X_i,W_i)|X_i]| > \frac{\lambda}{12}\right\}\right] \\ & \quad + E\left[|E[Y_i(1) m_{0,n}(X_i,W_i)|X_i]| I\left\{|E[Y_i(1) m_{0,n}(X_i,W_i)|X_i]| > \frac{\lambda}{12}\right\}\right] \\ & \quad + E\left[|E[m_{1,n}(X_i,W_i) m_{0,n}(X_i,W_i)|X_i]| I\left\{|E[m_{1,n}(X_i,W_i) m_{0,n}(X_i,W_i)|X_i]| > \frac{\lambda}{6}\right\}\right] \ . \end{aligned}$$

It follows from Assumptions 2.1(b) and 3.1(b) together with Lemma B.1 that

$$\begin{aligned} \lim_{\lambda\to\infty} \limsup_{n\to\infty} E\left[E[Y_i^2(1)|X_i] I\left\{E[Y_i^2(1)|X_i] > \frac{\lambda}{12}\right\}\right] &= 0 \\ \lim_{\lambda\to\infty} \limsup_{n\to\infty} E\left[E[m^2_{1,n}(X_i,W_i)|X_i] I\left\{E[m^2_{1,n}(X_i,W_i)|X_i] > \frac{\lambda}{3}\right\}\right] &= 0 \\ \lim_{\lambda\to\infty} \limsup_{n\to\infty} E\left[E[m^2_{0,n}(X_i,W_i)|X_i] I\left\{E[m^2_{0,n}(X_i,W_i)|X_i] > \frac{\lambda}{3}\right\}\right] &= 0 \ . \end{aligned}$$

For the last term in (72), note

$$\begin{aligned} & E\left[|E[m_{1,n}(X_i,W_i) m_{0,n}(X_i,W_i)|X_i]| I\left\{|E[m_{1,n}(X_i,W_i) m_{0,n}(X_i,W_i)|X_i]| > \frac{\lambda}{6}\right\}\right] \\ & \leq E\left[E[|m_{1,n}(X_i,W_i) m_{0,n}(X_i,W_i)||X_i] I\left\{E[|m_{1,n}(X_i,W_i) m_{0,n}(X_i,W_i)||X_i] > \frac{\lambda}{6}\right\}\right] \ . \end{aligned}$$

Meanwhile,

$$\begin{aligned} & E\left[E[|m_{1,n}(X_i,W_i) m_{0,n}(X_i,W_i)| I\left\{|m_{1,n}(X_i,W_i) m_{0,n}(X_i,W_i)| > \lambda\right\}\right] \\ & \leq E[m^2_{1,n}(X_i,W_i) I\{|m_{1,n}(X_i,W_i)| > \sqrt{\lambda}\}] + E[m^2_{0,n}(X_i,W_i) I\{|m_{0,n}(X_i,W_i)| > \sqrt{\lambda}\}] \ . \end{aligned}$$

It then follows from the previous two inqualities, Assumption 3.1(b), and Lemma B.1 that

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} E\left[|E[m_{1,n}(X_i,W_i) m_{0,n}(X_i,W_i)|X_i]| I\left\{|E[m_{1,n}(X_i,W_i) m_{0,n}(X_i,W_i)|X_i]| > \frac{\lambda}{6}\right\}\right] = 0 \ .$$

Similar arguments establish

$$\lim_{\lambda\to\infty}\limsup_{n\to\infty} E\left[|E[Y_i(1)m_{1,n}(X_i,W_i)|X_i]|I\left\{|E[Y_i(1)m_{1,n}(X_i,W_i)|X_i]|>\frac{\lambda}{12}\right\}\right]=0$$

$$\lim_{\lambda\to\infty}\limsup_{n\to\infty} E\left[|E[Y_i(1)m_{0,n}(X_i,W_i)|X_i]|I\left\{|E[Y_i(1)m_{0,n}(X_i,W_i)|X_i]|>\frac{\lambda}{12}\right\}\right]=0\ .$$

Therefore, (69) follows. The conclusion then follows from (68)–(69) and Assumption 3.1(a). ■

Lemma B.3. *Suppose Assumptions 2.1–2.3 and 3.1 hold. Then,*

$$\lim_{\gamma\to\infty}\limsup_{n\to\infty} E[|\phi_{1,n,i}-E[\phi_{1,n,i}|X_i]|^2 I\{|\phi_{1,n,i}-E[\phi_{1,n,i}|X_i]|^2>\gamma\}]=0\ .$$

Proof. Note

$$\begin{aligned}
&E[|\phi_{1,n,i}-E[\phi_{1,n,i}|X_i]|^2 I\{|\phi_{1,n,i}-E[\phi_{1,n,i}|X_i]|^2>\gamma\}]\\
&\lesssim E\left[(\phi_{1,n,i}^2+E[\phi_{1,n,i}|X_i]^2)I\left\{\phi_{1,n,i}^2+E[\phi_{1,n,i}|X_i]^2>\frac{\gamma}{2}\right\}\right]\\
&\lesssim E\left[\phi_{1,n,i}^2 I\left\{\phi_{1,n,i}^2>\frac{\gamma}{4}\right\}\right]+E\left[E[\phi_{1,n,i}|X_i]^2 I\left\{E[\phi_{1,n,i}|X_i]^2>\frac{\gamma}{4}\right\}\right]\ .
\end{aligned}$$

where the first inequality follows from $(a+b)^2\le 2(a^2+b^2)$ and the second inequality follows from (70). Next, note

$$\begin{aligned}
&E\left[E[\phi_{1,n,i}|X_i]^2 I\left\{E[\phi_{1,n,i}|X_i]^2>\frac{\gamma}{4}\right\}\right]\\
&\lesssim E\left[E[Y_i(1)|X_i]^2 I\left\{E[Y_i(1)|X_i]^2>\frac{\gamma}{24}\right\}\right]+E\left[E[m_{1,n}(X_i,W_i)|X_i]^2 I\left\{E[m_{1,n}(X_i,W_i)|X_i]^2>\frac{\gamma}{6}\right\}\right]\\
&\quad+E\left[E[m_{0,n}(X_i,W_i)|X_i]^2 I\left\{E[m_{0,n}(X_i,W_i)|X_i]^2>\frac{\gamma}{6}\right\}\right]\\
&\quad+E\left[|E[Y_i(1)|X_i]E[m_{1,n}(X_i,W_i)|X_i]|I\left\{|E[Y_i(1)|X_i]E[m_{1,n}(X_i,W_i)|X_i]|>\frac{\gamma}{24}\right\}\right]\\
&\quad+E\left[|E[Y_i(1)|X_i]E[m_{0,n}(X_i,W_i)|X_i]|I\left\{|E[Y_i(1)|X_i]E[m_{0,n}(X_i,W_i)|X_i]|>\frac{\gamma}{24}\right\}\right]\\
&\quad+E\left[|E[m_{1,n}(X_i,W_i)|X_i]E[m_{0,n}(X_i,W_i)|X_i]|I\left\{|E[m_{1,n}(X_i,W_i)|X_i]E[m_{0,n}(X_i,W_i)|X_i]|>\frac{\gamma}{12}\right\}\right]\\
&\lesssim E\left[E[Y_i^2(1)|X_i]I\left\{E[Y_i^2(1)|X_i]>\frac{\gamma}{24}\right\}\right]+E\left[E[m_{1,n}^2(X_i,W_i)|X_i]I\left\{E[m_{1,n}^2(X_i,W_i)|X_i]>\frac{\gamma}{6}\right\}\right]\\
&\quad+E\left[E[m_{0,n}^2(X_i,W_i)|X_i]I\left\{E[m_{0,n}^2(X_i,W_i)|X_i]>\frac{\gamma}{6}\right\}\right]\\
&\quad+E\left[|E[Y_i(1)|X_i]|I\left\{|E[Y_i(1)|X_i]|>\sqrt{\frac{\gamma}{24}}\right\}\right]\\
&\quad+E\left[|E[m_{1,n}(X_i,W_i)|X_i]|I\left\{|E[m_{1,n}(X_i,W_i)|X_i]|>\sqrt{\frac{\gamma}{24}}\right\}\right]\\
&\quad+E\left[|E[m_{0,n}(X_i,W_i)|X_i]|I\left\{|E[m_{0,n}(X_i,W_i)|X_i]|>\sqrt{\frac{\gamma}{24}}\right\}\right]\\
&\quad+E\left[|E[m_{1,n}(X_i,W_i)|X_i]|I\left\{|E[m_{1,n}(X_i,W_i)|X_i]|>\sqrt{\frac{\gamma}{12}}\right\}\right]\\
&\quad+E\left[|E[m_{0,n}(X_i,W_i)|X_i]|I\left\{|E[m_{0,n}(X_i,W_i)|X_i]|>\sqrt{\frac{\gamma}{12}}\right\}\right]\\
&\le E\left[E[Y_i^2(1)|X_i]I\left\{E[Y_i^2(1)|X_i]>\frac{\gamma}{24}\right\}\right]+E\left[E[m_{1,n}^2(X_i,W_i)|X_i]I\left\{E[m_{1,n}^2(X_i,W_i)|X_i]>\frac{\gamma}{6}\right\}\right]
\end{aligned}$$

$$+E\left[E[m_{0,n}^2(X_i,W_i)|X_i]I\left\{E[m_{0,n}^2(X_i,W_i)|X_i]>\frac{\gamma}{6}\right\}\right]$$
$$+E\left[E[Y_i^2(1)|X_i]I\left\{E[Y_i^2(1)|X_i]>\frac{\gamma}{24}\right\}\right]$$
$$+E\left[E[m_{1,n}^2(X_i,W_i)|X_i]I\left\{E[m_{1,n}^2(X_i,W_i)|X_i]>\frac{\gamma}{24}\right\}\right]$$
$$+E\left[E[m_{0,n}^2(X_i,W_i)|X_i]I\left\{E[m_{0,n}^2(X_i,W_i)|X_i]>\sqrt{\frac{\gamma}{24}}\right\}\right]$$
$$+E\left[E[m_{1,n}^2(X_i,W_i)|X_i]I\left\{E[m_{1,n}^2(X_i,W_i)|X_i]>\sqrt{\frac{\gamma}{12}}\right\}\right]$$
$$+E\left[E[m_{0,n}^2(X_i,W_i)|X_i]I\left\{E[m_{0,n}^2(X_i,W_i)|X_i]>\sqrt{\frac{\gamma}{12}}\right\}\right] ,$$

where the first inequality follows from (70), the second one follows from the conditional Jensen's inequality and (71), and the third one follows again from the conditional Jensen's inequality. It then follows from Lemma B.1 together with Assumptions 2.1(b) and 3.1(b) that

$$\lim_{\gamma\to\infty}\limsup_{n\to\infty}E\left[E[\phi_{1,n,i}|X_i]^2 I\left\{E[\phi_{1,n,i}|X_i]^2>\frac{\gamma}{4}\right\}\right]=0 .$$

Similar arguments lead to

$$\lim_{\gamma\to\infty}\limsup_{n\to\infty}E\left[\phi_{1,n,i}^2 I\left\{\phi_{1,n,i}^2>\frac{\gamma}{4}\right\}\right]=0 .$$

The conclusion then follows. ■

Lemma B.4. *Suppose Assumptions in Theorem 5.1 hold. Then,*

$$\left\|\frac{1}{n}\sum_{1\le j\le n}\delta_{\epsilon,j}\right\|_\infty=O_P\left(\sqrt{\frac{\log(2np_n)}{n}}+\zeta_n\right)\quad \text{and}\quad \left\|\Omega_n^{-1}\frac{1}{n}\sum_{1\le j\le n}\delta_{\epsilon,j}\delta_{\psi,j}\right\|_\infty=O_P\left(\sqrt{\frac{\log(2np_n)}{n}}+\zeta_n\right).$$

PROOF. Note $\omega_{n,l}\ge\underline{\sigma}>0$. It suffices to bound $\frac{1}{n}\sum_{1\le j\le n}\delta_{\epsilon,j}\delta_{\psi,j}$. We have

$$\begin{aligned}
\delta_{\epsilon,j}&=\delta_{Y,j}-\delta_{\psi,j}'\beta_n^{\text{hd-pd}}-\alpha_n^{\text{hd-pd}}\\
&=(D_{\pi(2j-1)}-D_{\pi(2j)})((Y_{\pi(2j-1)}-Y_{\pi(2j)})-(\psi_{n,\pi(2j-1)}-\psi_{n,\pi(2j)})'\beta_n^{\text{hd-pd}})-\alpha_n^{\text{hd-pd}}\\
&=(D_{\pi(2j-1)}-D_{\pi(2j)})((Y_{\pi(2j-1)}-Y_{\pi(2j)})-(\tilde\psi_{n,\pi(2j-1)}-\tilde\psi_{n,\pi(2j)})'\beta_n^{\text{hd-pd}})-\alpha_n^{\text{hd-pd}}\\
&-(D_{\pi(2j-1)}-D_{\pi(2j)})(\Psi(X_{\pi(2j-1)})-\Psi(X_{\pi(2j)}))'\beta_n^{\text{hd-pd}}-\alpha_n^{\text{hd-pd}}\\
&=(D_{\pi(2j-1)}-D_{\pi(2j)})(\epsilon_{n,\pi(2j)}(D_{\pi(2j-1)})-\epsilon_{n,\pi(2j-1)}(D_{\pi(2j)}))\\
&-(D_{\pi(2j-1)}-D_{\pi(2j)})(\Psi(X_{\pi(2j-1)})-\Psi(X_{\pi(2j)}))'\beta_n^{\text{hd-pd}}\\
&+(D_{\pi(2j-1)}-D_{\pi(2j)})(\tilde\psi_{n,\pi(2j-1)}+\tilde\psi_{n,\pi(2j)})'\left(\frac{\beta^{\text{hd-pd}}_{D_{\pi(2j-1)},n}-\beta^{\text{hd-pd}}_{D_{\pi(2j)},n}}{2}\right)\\
&+(D_{\pi(2j-1)}-D_{\pi(2j)})(\alpha^{\text{hd-pd}}_{D_{\pi(2j-1)},n}-\alpha^{\text{hd-pd}}_{D_{\pi(2j)},n})-\alpha_n^{\text{hd-pd}}\\
&=(D_{\pi(2j-1)}-D_{\pi(2j)})(\epsilon_{n,\pi(2j-1)}(D_{\pi(2j-1)})-\epsilon_{n,\pi(2j)}(D_{\pi(2j)}))\\
&-(D_{\pi(2j-1)}-D_{\pi(2j)})(\Psi(X_{\pi(2j-1)})-\Psi(X_{\pi(2j)}))'\beta_n^{\text{hd-pd}}\\
&+(D_{\pi(2j-1)}-D_{\pi(2j)})(\tilde\psi_{n,\pi(2j-1)}+\tilde\psi_{n,\pi(2j)})'\left(\frac{\beta^{\text{hd-pd}}_{D_{\pi(2j-1)},n}-\beta^{\text{hd-pd}}_{D_{\pi(2j)},n}}{2}\right), \qquad (73)
\end{aligned}$$

where we use the fact that

$$(D_{\pi(2j-1)} - D_{\pi(2j)})(\alpha^{\text{hd-pd}}_{D_{\pi(2j-1)},n} - \alpha^{\text{hd-pd}}_{D_{\pi(2j)},n}) = \alpha^{\text{hd-pd}}_{1,n} - \alpha^{\text{hd-pd}}_{0,n} = \alpha^{\text{hd-pd}}_{n}.$$

To see the first result in Lemma B.4, we note that that

$$\begin{aligned}
&\frac{1}{n}\sum_{j\in[n]}(D_{\pi(2j-1)} - D_{\pi(2j)})(\epsilon_{n,\pi(2j-1)}(D_{\pi(2j-1)}) - \epsilon_{n,\pi(2j)}(D_{\pi(2j)})) \\
&= \frac{1}{2n}\sum_{i\in[2n]} I(D_i = 1)\epsilon_{n,i}(1) - \frac{1}{2n}\sum_{i\in[2n]} I(D_i = 0)\epsilon_{n,i}(0) \\
&= \frac{1}{2n}\sum_{i\in[2n]} I(D_i = 1)(\epsilon_{n,i}(1) - E(\epsilon_{n,i}(1)|X_i)) - \frac{1}{2n}\sum_{i\in[2n]} I(D_i = 0)(\epsilon_{n,i}(0) - E(\epsilon_{n,i}(0)|X_i)) \\
&+ \frac{1}{2n}\sum_{i\in[2n]} I(D_i = 1)E(\epsilon_{n,i}(1)|X_i) - E(\epsilon_{n,i}(1)) + \frac{1}{2n}\sum_{i\in[2n]} I(D_i = 0)E(\epsilon_{n,i}(0)|X_i) - E(\epsilon_{n,i}(0)) \\
&+ (E(\epsilon_{n,i}(1)) - E(\epsilon_{n,i}(0))).
\end{aligned}$$

Following the same arguments used in the proof of $\mathcal{E}_{n,1}(d)$ and $\mathcal{E}_{n,2}(d)$ below, we can show that

$$\begin{aligned}
&\frac{1}{2n}\sum_{i\in[2n]} I(D_i = 1)(\epsilon_{n,i}(1) - E(\epsilon_{n,i}(1)|X_i)) - \frac{1}{2n}\sum_{i\in[2n]} I(D_i = 0)(\epsilon_{n,i}(0) - E(\epsilon_{n,i}(0)|X_i)) = O_P\left(\frac{1}{\sqrt{n}}\right) \\
&\frac{1}{2n}\sum_{i\in[2n]} I(D_i = 1)E(\epsilon_{n,i}(1)|X_i) - E(\epsilon_{n,i}(1)) + \frac{1}{2n}\sum_{i\in[2n]} I(D_i = 0)E(\epsilon_{n,i}(0)|X_i) - E(\epsilon_{n,i}(0)) = O_P\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

Then, Assumption 5.1 implies

$$\frac{1}{n}\sum_{j\in[n]}(D_{\pi(2j-1)} - D_{\pi(2j)})(\epsilon_{n,\pi(2j-1)}(D_{\pi(2j-1)}) - \epsilon_{n,\pi(2j)}(D_{\pi(2j)})) = O_P\left(\sqrt{\frac{\log(2np_n)}{n}}\right).$$

In addition, by Lemma B.5, we have

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{1\le j\le n}(D_{\pi(2j)} - D_{\pi(2j-1)})(\Psi(X_{\pi(2j)}) - \Psi(X_{\pi(2j-1)}))'\beta^{\text{hd-pd}}_{n}\right| = O_P(\zeta_n) \ , \\
&\left|\frac{1}{n}\sum_{1\le j\le n}(D_{\pi(2j)} - D_{\pi(2j-1)})(\tilde{\psi}_{n,\pi(2j)} + \tilde{\psi}_{n,\pi(2j-1)})'\left(\frac{\beta^{\text{hd-pd}}_{D_{\pi(2j)},n} - \beta^{\text{hd-pd}}_{D_{\pi(2j-1)},n}}{2}\right)\right| = O_P\left(\frac{1}{\sqrt{n}}\right) \ ,
\end{aligned}$$

which leads to the first result in Lemma B.4.

In addition, (73) implies

$$\begin{aligned}
&\frac{1}{n}\sum_{1\le j\le n}\delta_{\epsilon,j}\delta_{\psi,j} \\
&= \frac{1}{n}\sum_{1\le j\le n}(\psi_{n,\pi(2j)} - \psi_{n,\pi(2j-1)})(\epsilon_{n,\pi(2j)}(D_{\pi(2j)}) - \epsilon_{n,\pi(2j-1)}(D_{\pi(2j-1)}))
\end{aligned}$$

$$
\begin{aligned}
&- \frac{1}{n} \sum_{1 \le j \le n} (\psi_{n,\pi(2j)} - \psi_{n,\pi(2j-1)})(\Psi(X_{\pi(2j)}) - \Psi(X_{\pi(2j-1)}))' \beta_n^{\mathrm{hd-pd}} \\
&+ \frac{1}{n} \sum_{1 \le j \le n} (\psi_{n,\pi(2j)} - \psi_{n,\pi(2j-1)})(\tilde{\psi}_{n,\pi(2j)} + \tilde{\psi}_{n,\pi(2j-1)})' \left(\frac{\beta^{\mathrm{hd-pd}}_{D_{\pi(2j)},n} - \beta^{\mathrm{hd-pd}}_{D_{\pi(2j-1)},n}}{2} \right).
\end{aligned}
$$

Lemma B.5 below shows

$$
\left\| \frac{1}{n} \sum_{1 \le j \le n} (\psi_{n,\pi(2j)} - \psi_{n,\pi(2j-1)})(\Psi(X_{\pi(2j)}) - \Psi(X_{\pi(2j-1)}))' \beta_n^{\mathrm{hd-pd}} \right\|_\infty = o_P \left(\sqrt{\frac{\log(2np_n)}{n}} + \zeta_n \right)
$$

and

$$
\left\| \frac{1}{n} \sum_{1 \le j \le n} (\psi_{n,\pi(2j)} - \psi_{n,\pi(2j-1)})(\tilde{\psi}_{n,\pi(2j)} + \tilde{\psi}_{n,\pi(2j-1)})' \left(\frac{\beta^{\mathrm{hd-pd}}_{D_{\pi(2j)},n} - \beta^{\mathrm{hd-pd}}_{D_{\pi(2j-1)},n}}{2} \right) \right\|_\infty = O_P \left(\sqrt{\frac{\log(2np_n)}{n}} + \zeta_n \right) .
$$

It remains to bound $\frac{1}{n} \sum_{1 \le j \le n} (\psi_{n,\pi(2j)} - \psi_{n,\pi(2j-1)})(\epsilon_{n,\pi(2j)}(D_{\pi(2j)}) - \epsilon_{n,\pi(2j-1)}(D_{\pi(2j-1)}))$. We have

$$
\begin{aligned}
&\left\| \frac{1}{n} \sum_{1 \le j \le n} (\psi_{n,\pi(2j)} - \psi_{n,\pi(2j-1)})(\epsilon_{n,\pi(2j)}(D_{\pi(2j)}) - \epsilon_{n,\pi(2j-1)}(D_{\pi(2j-1)})) \right\|_\infty \\
&\le \left\| \frac{1}{n} \sum_{1 \le j \le n} (\tilde{\psi}_{n,\pi(2j)} - \tilde{\psi}_{n,\pi(2j-1)})(\epsilon_{n,\pi(2j)}(D_{\pi(2j)}) - \epsilon_{n,\pi(2j-1)}(D_{\pi(2j-1)})) \right\|_\infty \\
&+ \left\| \frac{1}{n} \sum_{1 \le j \le n} (\Psi(X_{\pi(2j)}) - \Psi(X_{\pi(2j-1)}))(\epsilon_{n,\pi(2j)}(D_{\pi(2j)}) - \epsilon_{n,\pi(2j-1)}(D_{\pi(2j-1)})) \right\|_\infty \\
&\equiv I + II.
\end{aligned}
$$

Let $p = q/(q-1) < 2$. For II, we have

$$
\begin{aligned}
II &\lesssim \frac{1}{n} \sum_{1 \le j \le n} ||X_{\pi(2j)} - X_{\pi(2j-1)}||_2 |\epsilon_{n,\pi(2j)}(D_{\pi(2j)}) - \epsilon_{n,\pi(2j-1)}(D_{\pi(2j-1)})| \\
&\le \left(\frac{1}{n} \sum_{1 \le j \le n} ||X_{\pi(2j)} - X_{\pi(2j-1)}||_2^p \right)^{1/p} \left(\frac{1}{n} \sum_{1 \le j \le n} |\epsilon_{n,\pi(2j)}(D_{\pi(2j)}) - \epsilon_{n,\pi(2j-1)}(D_{\pi(2j-1)})|^q \right)^{1/q} \\
&\lesssim \left(\frac{1}{n} \sum_{1 \le j \le n} ||X_{\pi(2j)} - X_{\pi(2j-1)}||_2^p \right)^{1/p} \left(\frac{1}{n} \sum_{1 \le i \le 2n} (|\epsilon_{n,i}(1)|^q + |\epsilon_{n,i}(0)|^q) \right)^{1/q} \\
&= O_P(\zeta_n).
\end{aligned}
$$

For I, we have

$$
\left\| \frac{1}{n} \sum_{1 \le j \le n} (\tilde{\psi}_{n,\pi(2j)} - \tilde{\psi}_{n,\pi(2j-1)})(\epsilon_{n,\pi(2j)}(D_{\pi(2j)}) - \epsilon_{n,\pi(2j-1)}(D_{\pi(2j-1)})) \right\|_\infty
$$

$$
\begin{aligned}
&\leq \sum_{d=0,1} \left\| \frac{1}{n} \sum_{1\leq i\leq 2n} I(D_i = d)\tilde{\psi}_{n,i}\epsilon_{n,i}(d) \right\|_\infty \\
&+ \left\| \frac{1}{n} \sum_{1\leq j\leq n} \left[\tilde{\psi}_{n,\pi(2j)}\epsilon_{n,\pi(2j-1)}(D_{\pi(2j-1)}) + \tilde{\psi}_{n,\pi(2j-1)}\epsilon_{n,\pi(2j)}(D_{\pi(2j)}) \right] \right\|_\infty \\
&\leq \sum_{d=0,1} \left\| \frac{1}{n} \sum_{1\leq i\leq 2n} I(D_i = d)\tilde{\psi}_{n,i}\epsilon_{n,i}(d) \right\|_\infty + O_P\left(\sqrt{\frac{\log(2np_n)}{n}} \right) ,
\end{aligned}
$$

where the last inequality is by Lemma B.5. To bound the first term on the RHS of the above display, we further define

$$
\mathcal{E}_{n,0}(d) = \begin{pmatrix} \frac{1}{2n}\sum_{1\leq i\leq 2n} E[\epsilon_{n,i}^4(d)|X_i] \leq c_0 < \infty \ , \\ \min_{1\leq l\leq p_n} \frac{1}{n}\sum_{1\leq i\leq 2n} I\{D_i = d\}\operatorname{Var}[\tilde{\psi}_{n,i,l}\epsilon_{n,i}(d)|X_i] \geq \underline{\sigma}^2 > 0 \ , \end{pmatrix}
$$

$$
\mathcal{E}_{n,1}(d) = \left\{ \left\| \frac{1}{n} \sum_{1\leq i\leq 2n} I\{D_i = d\}(\tilde{\psi}_{n,i}\epsilon_{n,i}(d) - E[\tilde{\psi}_{n,i}\epsilon_{n,i}(d)|X_i]) \right\|_\infty \leq C\sqrt{\log(2np_n)/n} \right\} ,
$$

$$
\mathcal{E}_{n,2}(d) = \left\{ \left\| \frac{1}{n} \sum_{1\leq i\leq 2n} I\{D_i = d\}(E[\tilde{\psi}_{n,i}\epsilon_{n,i}(d)|X_i] - E[\tilde{\psi}_{n,i}\epsilon_{n,i}(d)]) \right\|_\infty \leq C\sqrt{\log(2np_n)/n} \right\} ,
$$

$$
\mathcal{E}_{n,3}(d) = \left\{ \max_{1\leq l\leq p_n} \left| \frac{1}{2n} \sum_{1\leq i\leq 2n} (E[\tilde{\psi}_{n,i,l}^2\epsilon_{n,i}^2(d)|X_i] - E[\tilde{\psi}_{n,i,l}^2\epsilon_{n,i}^2(d)]) \right|^{1/2} \leq C \right\} ,
$$

and

$$
\mathcal{E}_{n,4}(d) = \left\{ \max_{1\leq l\leq p_n} \left| \frac{1}{2n} \sum_{1\leq i\leq 2n} (2I\{D_i = d\} - 1)(E[\epsilon_{n,i}^2(d)\tilde{\psi}_{n,i,l}^2|X_i] - E[\epsilon_{n,i}^2(d)\tilde{\psi}_{n,i,l}^2]) \right|^{1/2} \leq C \right\} .
$$

We aim to show that $P\{\mathcal{E}_{n,1}(d)\} \to 1$ and $P\{\mathcal{E}_{n,2}(d)\} \to 1$ for some sufficiently large constant C, which implies

$$
\begin{aligned}
&P\left\{ \left\| \frac{1}{n} \sum_{1\leq i\leq 2n} I(D_i = d)\tilde{\psi}_{n,i}\epsilon_{n,i}(d) \right\|_\infty \geq C\left(\sqrt{\frac{\log(2np_n)}{n}} + \zeta_n \right) \right\} \\
&\leq P\left\{ \left\| \frac{1}{n} \sum_{1\leq i\leq 2n} I(D_i = d)(\tilde{\psi}_{n,i}\epsilon_{n,i}(d) - E\tilde{\psi}_{n,i}\epsilon_{n,i}(d)) \right\|_\infty \geq 0.5C\left(\sqrt{\frac{\log(2np_n)}{n}} + \zeta_n \right) \right\} \\
&\leq P\{\mathcal{E}_{n,1}^c(d)\} + P\{\mathcal{E}_{n,2}^c(d)\} \to 0 \ ,
\end{aligned}
$$

where the first inequality is by (23).

First, we show $P\{\mathcal{E}_{n,3}(d)\} \to 1$. Let

$$t_n = C\sqrt{\frac{\log(np_n)\Xi_n^2}{n}}$$

for some sufficiently large constant C and $\{e_i\}_{1\le i\le 2n}$ be a sequence of i.i.d. Rademacher random variables independent of everything else. Then, we have

$$\begin{aligned}
&\left(1 - \frac{4\max_{1\le l\le p_n}\mathrm{Var}[E[\tilde{\psi}^2_{n,i,l}\epsilon^2_{n,i}(d)|X_i]]}{2nt_n^2}\right) P\left\{\max_{1\le l\le p_n}\left|\frac{1}{2n}\sum_{1\le i\le 2n}\left[E[\tilde{\psi}^2_{n,i,l}\epsilon^2_{n,i}(d)|X_i] - E[\tilde{\psi}^2_{n,i,l}\epsilon^2_{n,i}(d)]\right]\right| \ge t_n\right\}\\
&\le 2P\left\{\max_{1\le l\le p_n}\left|\frac{1}{2n}\sum_{1\le i\le 2n}4e_iE[\tilde{\psi}^2_{n,i,l}\epsilon^2_{n,i}(d)|X_i]\right| \ge t_n\right\}\\
&= o(1) + 2E\left[P\left\{\max_{1\le l\le p_n}\left|\frac{1}{2n}\sum_{1\le i\le 2n}4e_iE[\tilde{\psi}^2_{n,i,l}\epsilon^2_{n,i}(d)|X_i]\right| \ge t_n\middle|X^{(n)}\right\}I\{\mathcal{E}_{n,0}(d)\}\right]\\
&\lesssim o(1) + p_n\exp\left(-\frac{nt_n^2}{\Xi_n^2 c}\right) = o(1),
\end{aligned}$$

where the first inequality is by van der Vaart and Wellner (1996, Lemma 2.3.7), the second inequality is by the Hoeffding's inequality conditional on $X^{(n)}$ and the fact that, on $\mathcal{E}_{n,0}(d)$,

$$\begin{aligned}
\frac{1}{2n}\sum_{1\le i\le 2n}(E[\tilde{\psi}^2_{n,i,l}\epsilon^2_{n,i}(d)|X_i])^2 &\le \frac{1}{2n}\sum_{1\le i\le 2n}E[\tilde{\psi}^4_{n,i,l}|X_i]E[\epsilon^4_{n,i}(d)|X_i]\\
&\le \frac{2\Xi_n^2}{n}\sum_{1\le i\le 2n}E[\tilde{\psi}^2_{n,i,l}|X_i]E[\epsilon^4_{n,i}(d)|X_i] \le \Xi_n^2 Cc_0\ .
\end{aligned}$$

To see the above inequality, note that C is a fixed constant, the second last inequality is by $|\tilde{\psi}_{n,i,l}| \le 2\Xi_n$, and the last inequality is by the fact that $\log(p_n)\Xi_n^2 = o(n)$. Furthermore, we note that

$$\begin{aligned}
&\frac{4\max_{1\le l\le p_n}\mathrm{Var}[E[\tilde{\psi}^2_{n,i,l}\epsilon^2_{n,i}(d)|X_i]]}{2nt_n^2}\\
&\lesssim \frac{\max_{1\le l\le p_n}E\left[E[\tilde{\psi}^4_{n,i,l}|X_i]E[\epsilon^4_{n,i}(d)|X_i]\right]}{nt_n^2}\\
&\lesssim \frac{\Xi_n^2\max_{1\le l\le p_n}E\left[E[\tilde{\psi}^2_{n,i,l}|X_i]E[\epsilon^4_{n,i}(d)|X_i]\right]}{n}\\
&\lesssim \frac{\Xi_n^2E[\epsilon^4_{n,i}(d)]}{nt_n^2}\\
&= o(1).
\end{aligned}$$

Therefore, we have $P\{\mathcal{E}_{n,3}(d)\} \to 1$.

Next, we show $P\{\mathcal{E}_{n,4}(d)\} \to 1$. Define $a_{n,i,l} = E[\epsilon^2_{n,i}(d)\tilde{\psi}^2_{n,i,l}|X_i] - E[\epsilon^2_{n,i}(d)\tilde{\psi}^2_{n,i,l}]$. Then, we have

$$
\begin{aligned}
& P\left\{\max_{1\le l\le p_n}\left|\frac{1}{2n}\sum_{1\le i\le 2n}(2I\{D_i=d\}-1)(E[\epsilon^2_{n,i}(d)\tilde{\psi}^2_{n,i,l}|X_i]-E[\epsilon^2_{n,i}(d)\tilde{\psi}^2_{n,i,l}])\right| > t_n \middle| X^{(n)}\right\} I\{\mathcal{E}_{n,0}(d)\} \\
& \le \sum_{1\le l\le p_n} P\left\{\left|\frac{1}{2n}\sum_{1\le j\le n}(I\{D_{\pi(2j-1)}=d\}-I\{D_{\pi(2j)}=d\})(a_{n,\pi(2j-1),l}-a_{n,\pi(2j),l})\right| > t_n \middle| X^{(n)}\right\} I\{\mathcal{E}_{n,0}(d)\} \\
& \le \sum_{1\le l\le p_n} \exp\left(-\frac{2nt_n^2}{\frac{1}{n}\sum_{1\le j\le n}(a_{n,\pi(2j-1),l}-a_{n,\pi(2j),l})^2}\right) I\{\mathcal{E}_{n,0}(d)\} \\
& \le \exp\left(\log(p_n) - \frac{2nt_n^2}{\Xi_n^2 c^2}\right) ,
\end{aligned}
$$

where, conditional on $X^{(n)}$, $\{I\{D_{\pi(2j-1)}=d\}-I\{D_{\pi(2j)}=d\}\}_{1\le j\le n}$ is a sequence of i.i.d. Rademacher random variables, the second last inequality is by Hoeffding's inequality, and the last inequality is by that, on $\mathcal{E}_{n,0}(d)$,

$$
\begin{aligned}
& \left(\frac{1}{n}\sum_{1\le j\le n}(a_{n,\pi(2j-1),l}-a_{n,\pi(2j),l})^2\right)^{1/2} \\
& \le \left(\frac{1}{n}\sum_{1\le j\le n}(E[\tilde{\psi}^2_{n,\pi(2j-1),l}\epsilon^2_{n,i}(d)|X^{(n)}])^2\right)^{1/2} + \left(\frac{1}{n}\sum_{1\le j\le n}(E[\tilde{\psi}^2_{n,\pi(2j),l}\epsilon^2_{n,i}(d)|X^{(n)}])^2\right)^{1/2} \\
& \le \left(\frac{2}{n}\sum_{1\le i\le 2n}(E[\tilde{\psi}^2_{n,i,l}\epsilon^2_{n,i}(d)|X_i])^2\right)^{1/2} \\
& \le \Xi_n c .
\end{aligned}
$$

Recall $t_n = C\sqrt{\frac{\log(np_n)\Xi_n^2}{n}}$ for some sufficiently large C and note that $P\{\mathcal{E}_{n,0}(d)\} \to 1$. We have

$$
\max_{1\le l\le p_n}\left|\sum_{1\le i\le 2n}\frac{(2I\{D_i=d\}-1)(E[\epsilon^2_{n,i}(d)\tilde{\psi}^2_{n,i,l}|X_i]-E[\epsilon^2_{n,i}(d)\tilde{\psi}^2_{n,i,l}])}{2n}\right| = O_P\left(\sqrt{\frac{\log(np_n)\Xi_n^2}{n}}\right) ,
$$

and thus, $P\{\mathcal{E}_{n,4}(d)\} \to 1$.

Next, we show $P\{\mathcal{E}_{n,1}(d)\} \to 1$. We note that, for $d \in \{0,1\}$, conditional on $(D^{(n)}, X^{(n)})$, $\{\tilde{\psi}_{n,i}\epsilon_{n,i}(d)\}_{1\le i\le 2n}$ are independent. In what follows, we couple

$$
\mathbb{U}_n = \frac{1}{n}\sum_{1\le i\le 2n} I\{D_i=d\}(\tilde{\psi}_{n,i}\epsilon_{n,i}(d) - E[\tilde{\psi}_{n,i}\epsilon_{n,i}(d)|X_i])
$$

with a centered Gaussian random vector as in Theorem 2.1 in Chernozhukov et al. (2017). Let $Z = (Z_1, \ldots, Z_{p_n})$ be a Gaussian random vector with $E[Z_l] = 0$ for $1 \le l \le p_n$ and $\text{Var}[Z] = \text{Var}[\mathbb{U}_n|X^{(n)}, D^{(n)}]$ that additionally satisfies the conditions of that theorem. Specifically, $Z = (Z_1, \cdots, Z_{p_n})$ is a centered

Gaussian random vector in R^{p_n} such that on $\mathcal{E}_{n,0}(d) \cap \mathcal{E}_{n,3}(d) \cap \mathcal{E}_{n,4}(d)$,

$$E[ZZ'] = \frac{1}{n^2} \sum_{1 \le i \le 2n} I\{D_i = d\} E[\epsilon_{n,i}^2(d) \tilde{\psi}_{n,i} \tilde{\psi}_{n,i}' | X_i]$$
$$- \frac{1}{n} \left(\frac{1}{n} \sum_{1 \le i \le 2n} I\{D_i = d\} E[\epsilon_{n,i}(d) \tilde{\psi}_{n,i} | X_i] \right) \left(\frac{1}{n} \sum_{1 \le i \le 2n} I\{D_i = d\} E[\epsilon_{n,i}(d) \tilde{\psi}_{n,i} | X_i] \right)'$$

and by the definitions of $E_{n,3}(d)$ and $E_{n,4}(d)$,

$$\begin{aligned}
\max_{1 \le l \le p_n} E[Z_l^2] &\le \frac{\max_{1 \le l \le p_n} \sum_{1 \le i \le 2n} I\{D_i = d\} E[\epsilon_{n,i}^2(d) \tilde{\psi}_{n,i,l}^2 | X_i]}{n^2} \\
&\le \frac{C\overline{\sigma}^2}{n} + \frac{\max_{1 \le l \le p_n} \sum_{1 \le i \le 2n} I\{D_i = d\} (E[\epsilon_{n,i}^2(d) \tilde{\psi}_{n,i,l}^2 | X_i] - E[\epsilon_{n,i}^2(d) \tilde{\psi}_{n,i,l}^2])}{n^2} \\
&\le \frac{C\overline{\sigma}^2}{n} + \frac{\max_{1 \le l \le p_n} \sum_{1 \le i \le 2n} (2I\{D_i = d\} - 1)(E[\epsilon_{n,i}^2(d) \tilde{\psi}_{n,i,l}^2 | X_i] - E[\epsilon_{n,i}^2(d) \tilde{\psi}_{n,i,l}^2])}{2n^2} \\
&+ \frac{\max_{1 \le l \le p_n} \sum_{1 \le i \le 2n} (E[\epsilon_{n,i}^2(d) \tilde{\psi}_{n,i,l}^2 | X_i] - E[\epsilon_{n,i}^2(d) \tilde{\psi}_{n,i,l}^2])}{2n^2} \\
&\le \frac{C\overline{\sigma}^2}{n} ~.
\end{aligned}$$

Further define $q(1-\alpha)$ as the $(1-\alpha)$ quantile of $||Z||_\infty$. Then, we have

$$q(1 - 1/n) \le \frac{C\overline{\sigma}(\sqrt{2\log(2p_n)} + \sqrt{2\log(n)})}{\sqrt{n}} \le 2C\overline{\sigma}\sqrt{\log(2np_n)/n} ~,$$

where the first inequality is by the last display in the proof of Lemma E.2 in Chetverikov and Sørensen (2022) and the second inequality is by the fact that $\sqrt{a} + \sqrt{b} \le \sqrt{2(a+b)}$ for $a, b > 0$. Therefore, we have

$$\begin{aligned}
P\{\mathcal{E}_{n,1}^c(d)) &\le P\{\mathcal{E}_{n,1}^c(d), \mathcal{E}_{n,0}(d), \mathcal{E}_{n,3}(d), \mathcal{E}_{n,4}(d)\} + o(1) \\
&= EP\{\mathcal{E}_{n,1}^c(d) | D^{(n)}, X^{(n)}\} I\{\mathcal{E}_{n,0}(d), \mathcal{E}_{n,3}(d), \mathcal{E}_{n,4}(d)\} + o(1) \\
&\le E[P\{||Z||_\infty \ge 2C\overline{\sigma}\sqrt{\log(2np_n)/n} | D^{(n)}, X^{(n)}\} I\{\mathcal{E}_{n,0}(d), \mathcal{E}_{n,3}(d), \mathcal{E}_{n,4}(d)\}] + o(1) \\
&\le E[P\{||Z||_\infty \ge q(1-1/n) | D^{(n)}, X^{(n)}\} I\{\mathcal{E}_{n,0}(d), \mathcal{E}_{n,3}(d), \mathcal{E}_{n,4}(d)\}] + o(1) = o(1) ~,
\end{aligned}$$

where the second inequality is by Theorem 2.1 in Chernozhukov et al. (2017).

Finally, we turn to $\mathcal{E}_{n,2}(d)$ with $d = 1$. We have

$$\begin{aligned}
&\frac{1}{n} \sum_{1 \le i \le 2n} I\{D_i = 1\} (E[\tilde{\psi}_{n,i} \epsilon_{n,i}(1) | X_i] - E[\tilde{\psi}_{n,i} \epsilon_{n,i}(1)]) \\
&= \frac{1}{2n} \sum_{1 \le i \le 2n} (E[\tilde{\psi}_{n,i} \epsilon_{n,i}(1) | X_i] - E[\tilde{\psi}_{n,i} \epsilon_{n,i}(1)]) + \frac{1}{2n} \sum_{1 \le i \le 2n} (2D_i - 1)(E[\tilde{\psi}_{n,i} \epsilon_{n,i}(1) | X_i] - E[\tilde{\psi}_{n,i} \epsilon_{n,i}(1)]).
\end{aligned} \tag{74}$$

Note $\{E[\tilde{\psi}_{n,i}\epsilon_{n,i}(1)|X_i] - E[\psi_{n,i}\epsilon_{n,i}(1)]\}_{1\leq i\leq 2n}$ is a sequence of independent centered random variables and

$$\max_{1\leq l\leq p_n} E[(E[\tilde{\psi}_{n,i,l}\epsilon_{n,i}(1)|X_i] - E[\tilde{\psi}_{n,i,l}\epsilon_{n,i}(1)])^2] \leq C\overline{\sigma}^2.$$

Following Theorem 2.1 in Chernozhukov et al. (2017), Lemma E.2 in Chetverikov and Sørensen (2022), and similar arguments to the coupling argument above, we have

$$P\left\{\left\|\frac{1}{2n}\sum_{1\leq i\leq 2n}(E[\tilde{\psi}_{n,i}\epsilon_{n,i}(1)|X_i] - E[\tilde{\psi}_{n,i}\epsilon_{n,i}(1)])\right\|_\infty \leq C\overline{\sigma}\sqrt{\log(2np_n)/n}\right) \to 1. \tag{75}$$

For the second term on the RHS of (74), we define $g_{n,i,l} = E[\tilde{\psi}_{n,i,l}\epsilon_{n,i}(1)|X_i] - E[\tilde{\psi}_{n,i,l}\epsilon_{n,i}(1)]$. We have

$$\begin{aligned}
&P\left\{\left\|\frac{1}{2n}\sum_{1\leq i\leq 2n}(2D_i-1)E[\tilde{\psi}_{n,i}\epsilon_{n,i}(1)|X_i] - E[\tilde{\psi}_{n,i}\epsilon_{n,i}(1)]\right\|_\infty > t\middle| X^{(n)}\right\} \\
&\leq \sum_{1\leq l\leq p_n} P\left\{\left|\frac{1}{2n}\sum_{1\leq j\leq n}(D_{\pi(2j-1)} - D_{\pi(2j)})(g_{n,\pi(2j-1),l} - g_{n,\pi(2j),l})\right| > t\middle| X^{(n)}\right\} \\
&\leq \sum_{1\leq l\leq p_n} \exp\left(-\frac{2nt^2}{\frac{1}{n}\sum_{1\leq j\leq n}(g_{n,\pi(2j-1),l} - g_{n,\pi(2j),l})^2}\right),
\end{aligned}$$

where, conditional on $X^{(n)}$, $\{(D_{\pi(2j-1)} - D_{\pi(2j)})\}_{1\leq j\leq n}$ is a sequence of i.i.d. Rademacher random variables and the last inequality is by Hoeffding's inequality. In addition, on $\mathcal{E}_{n,3}(1)$, we have

$$\begin{aligned}
&\left(\frac{1}{n}\sum_{1\leq j\leq n}(g_{n,\pi(2j-1),l} - g_{n,\pi(2j),l})^2\right)^{1/2} \\
&\leq \left(\frac{1}{n}\sum_{1\leq j\leq n}(E[\tilde{\psi}_{n,\pi(2j-1),l}\epsilon_{n,\pi(2j-1)}(1)|X_{\pi(2j-1)}])^2\right)^{1/2} + \left(\frac{1}{n}\sum_{1\leq j\leq n}(E[\tilde{\psi}_{n,\pi(2j),l}\epsilon_{n,\pi(2j)}(1)|X_{\pi(2j)}])^2\right)^{1/2} \\
&\leq \left(\frac{2}{n}\sum_{1\leq i\leq 2n}(E[\tilde{\psi}_{n,i,l}\epsilon_{n,i}(1)|X_i])^2\right)^{1/2} \\
&\leq \left(\frac{2}{n}\sum_{1\leq i\leq 2n}E[\tilde{\psi}^2_{n,i,l}\epsilon^2_{n,i}(1)|X_i]\right)^{1/2} \\
&\leq \left(\frac{2}{n}\sum_{1\leq i\leq 2n}\left[E[\tilde{\psi}^2_{n,i,l}\epsilon^2_{n,i}(1)|X_i] - E[\tilde{\psi}^2_{n,i,l}\epsilon^2_{n,i}(1)]\right]\right)^{1/2} + 2\overline{\sigma} \\
&\leq C\overline{\sigma}.
\end{aligned}$$

Therefore, we have

$$P\left\{\left\|\frac{1}{2n}\sum_{1\leq i\leq 2n}(2D_i-1)E[\tilde{\psi}_{n,i}\epsilon_{n,i}(1)|X_i] - E[\tilde{\psi}_{n,i}\epsilon_{n,i}(1)]\right\|_\infty > C\sqrt{\frac{\log(np_n)\overline{\sigma}^2}{n}}\right\}$$

$$\leq P\left\{\left\|\frac{1}{2n}\sum_{1\leq i\leq 2n}(2D_i-1)E[\tilde{\psi}_{n,i}\epsilon_{n,i}(1)|X_i]-E[\tilde{\psi}_{n,i}\epsilon_{n,i}(1)]\right\|_\infty > C\sqrt{\frac{\log(np_n)\overline{\sigma}^2}{n}}, \mathcal{E}_{n,3}(1)\right\}+o(1)$$

$$\leq E\left[P\left\{\left\|\frac{1}{2n}\sum_{1\leq i\leq 2n}(2D_i-1)E[\tilde{\psi}_{n,i}\epsilon_{n,i}(1)|X_i]-E[\tilde{\psi}_{n,i}\epsilon_{n,i}(1)]\right\|_\infty > C\sqrt{\frac{\log(np_n)\overline{\sigma}^2}{n}}\middle|X^{(n)}\right\}I\{\mathcal{E}_{n,3}(1)\}\right]+o(1)$$

$$= o(1)\ . \tag{76}$$

Combining (74), (75), and (76), we have $P\{\mathcal{E}_{n,2}(1)\} \to 1$. The same result holds for $\mathcal{E}_{n,2}(0)$. ■

Lemma B.5. *Supposes Assumptions in Theorem 5.1 hold. Then, we have*

$$\left\|\frac{1}{n}\sum_{1\leq j\leq n}(\psi_{n,\pi(2j)}-\psi_{n,\pi(2j-1)})(\Psi(X_{\pi(2j)})-\Psi(X_{\pi(2j-1)}))'\beta_n^{\text{hd}-\text{pd}}\right\|_\infty = o_P\left(\zeta_n\right)\ ,$$

$$\left|\frac{1}{n}\sum_{1\leq j\leq n}(D_{\pi(2j)}-D_{\pi(2j-1)})(\Psi(X_{\pi(2j)})-\Psi(X_{\pi(2j-1)}))'\beta_n^{\text{hd}-\text{pd}}\right| = O_P\left(\zeta_n\right)\ ,$$

$$\left\|\frac{1}{n}\sum_{1\leq j\leq n}(\psi_{n,\pi(2j)}-\psi_{n,\pi(2j-1)})(\tilde{\psi}_{n,\pi(2j)}+\tilde{\psi}_{n,\pi(2j-1)})'\left(\frac{\beta^{\text{hd}-\text{pd}}_{D_{\pi(2j)},n}-\beta^{\text{hd}-\text{pd}}_{D_{\pi(2j-1)},n}}{2}\right)\right\|_\infty = O_P\left(\sqrt{\frac{\log(p_n)}{n}}+\zeta_n\right)\ ,$$

$$\left|\frac{1}{n}\sum_{1\leq j\leq n}(D_{\pi(2j)}-D_{\pi(2j-1)})(\tilde{\psi}_{n,\pi(2j)}+\tilde{\psi}_{n,\pi(2j-1)})'\left(\frac{\beta^{\text{hd}-\text{pd}}_{D_{\pi(2j)},n}-\beta^{\text{hd}-\text{pd}}_{D_{\pi(2j-1)},n}}{2}\right)\right| = O_P\left(\frac{1}{\sqrt{n}}\right)\ ,$$

$$\left\|\frac{1}{n}\sum_{1\leq j\leq n}\tilde{\psi}_{n,\pi(2j)}\epsilon_{n,\pi(2j-1)}(D_{\pi(2j-1)})\right\|_\infty = O_P\left(\sqrt{\frac{\log(p_n)}{n}}\right)\ , \quad \textit{and}$$

$$\left\|\frac{1}{n}\sum_{1\leq j\leq n}\tilde{\psi}_{n,\pi(2j-1)}\epsilon_{n,\pi(2j)}(D_{\pi(2j)})\right\|_\infty = O_P\left(\sqrt{\frac{\log(p_n)}{n}}\right)\ .$$

Proof. Recall $\tilde{\psi}_{n,i} = \psi_{n,i} - \Psi(X_i)$. For the first result, we note that

$$\begin{aligned}
&\left\|\frac{1}{n}\sum_{1\leq j\leq n}(\psi_{n,\pi(2j)}-\psi_{n,\pi(2j-1)})(\Psi(X_{\pi(2j)})-\Psi(X_{\pi(2j-1)}))'\beta_n^{\text{hd}-\text{pd}}\right\|_\infty \\
&\leq \left\|\frac{1}{n}\sum_{1\leq j\leq n}(\tilde{\psi}_{n,\pi(2j)}-\tilde{\psi}_{n,\pi(2j-1)})(\Psi(X_{\pi(2j)})-\Psi(X_{\pi(2j-1)}))'\beta_n^{\text{hd}-\text{pd}}\right\|_\infty \\
&+\left\|\frac{1}{n}\sum_{1\leq j\leq n}(\Psi(X_{\pi(2j)})-\Psi(X_{\pi(2j-1)}))(\Psi(X_{\pi(2j)})-\Psi(X_{\pi(2j-1)}))'\beta_n^{\text{hd}-\text{pd}}\right\|_\infty \\
&= \left\|\frac{1}{n}\sum_{1\leq j\leq n}(\tilde{\psi}_{n,\pi(2j)}-\tilde{\psi}_{n,\pi(2j-1)})(\Psi(X_{\pi(2j)})-\Psi(X_{\pi(2j-1)}))'\beta_n^{\text{hd}-\text{pd}}\right\|_\infty \\
&+\frac{C}{n}\sum_{1\leq j\leq n}(X_{\pi(2j)}-X_{\pi(2j-1)})^2
\end{aligned}$$

$$= \left\| \frac{1}{n} \sum_{1\le j\le n} (\tilde{\psi}_{n,\pi(2j)} - \tilde{\psi}_{n,\pi(2j-1)})(\Psi(X_{\pi(2j)}) - \Psi(X_{\pi(2j-1)}))' \beta_n^{\text{hd}-\text{pd}} \right\|_\infty + O_P(\zeta_n^2). \quad (77)$$

For the First term on the RHS of the above display, we note that, conditional on $(D^{(n)}, X^{(n)})$, $\{U_{n,j} = (\tilde{\psi}_{n,\pi(2j)} - \tilde{\psi}_{n,\pi(2j-1)})(\Psi(X_{\pi(2j)}) - \Psi(X_{\pi(2j-1)}))' \beta_n^{\text{hd}-\text{pd}}\}_{1\le j\le n}$ are independent and mean zero. We have

$$\max_{1\le l\le p_n} \frac{1}{n} \sum_{1\le j\le n} U_{n,j,l}^2 \le \frac{C\Xi_n^2}{n} \sum_{1\le j\le n} (X_{\pi(2j)} - X_{\pi(2j-1)})^2 = O_P(\Xi_n^2 \zeta_n^2).$$

For an arbitrary $\epsilon > 0$, let $\mathcal{A}_{n,1} = \{\max_{1\le l\le p_n} \frac{1}{n} \sum_{1\le j\le n} U_{n,j,l}^2 \le C_1 \Xi_n^2 \zeta_n^2\}$ for some large C_1 so that $P\{\mathcal{A}_{n,1}\} \ge 1 - \epsilon$, where $U_{n,j,l}$ is the lth element of $U_{n,j}$. Further denote $\{e_j\}_{1\le j\le n}$ as a sequence of independent Rademacher random variables. Then, for $t_n = M \log^{1/2}(p_n) \zeta_n \Xi_n n^{-1/2}$ with some sufficiently large M, we have

$$\begin{aligned}
&\left(1 - \frac{4\max_{1\le l\le p_n} \sum_{1\le j\le n} \text{Var}[U_{n,j,l} | D^{(n)}, X^{(n)}]}{n^2 t_n^2}\right) P\left\{ \max_{1\le l\le p_n} \left| \frac{1}{n} \sum_{1\le j\le n} U_{n,j,l} \right| \ge t_n \middle| D^{(n)}, X^{(n)} \right\} \\
&\le 2P\left\{ \max_{1\le l\le p_n} \left| \frac{1}{n} \sum_{1\le j\le n} 4e_i U_{n,j,l} \right| \ge t_n \middle| D^{(n)}, X^{(n)} \right\} \\
&\le \epsilon + 2E\left[P\left\{ \max_{1\le l\le p_n} \left| \frac{1}{n} \sum_{1\le j\le n} 4e_i U_{n,j,l} \right| \ge t \middle| D^{(n)}, X^{(n)}, W_n^{(n)} \right\} I\{\mathcal{A}_{n,1}\} \middle| D^{(n)}, X^{(n)} \right] \\
&\le \epsilon + Cp_n \exp\left(-\frac{nt_n^2}{\Xi_n^2 C_1 \Xi_n^2} \right) \le C\epsilon,
\end{aligned}$$

where the first inequality is by van der Vaart and Wellner (1996, Lemma 2.3.7) and the second last inequality is by the Hoeffding's inequality. In addition, for a sufficiently large M and by Assumption 5.2, we have

$$\frac{4\max_{1\le l\le p_n} \sum_{1\le j\le n} \text{Var}[U_{n,j,l} | D^{(n)}, X^{(n)}]}{n^2 t_n^2} \le \frac{C\zeta_n^2}{\Xi_n^2 \zeta_n^2 M^2 \log(p_n)} < 1,$$

where the first inequality is because, by Assumption 5.1,

$$\max_{1\le l\le p_n} E\left[(\tilde{\psi}_{n,\pi(2j),l} - \tilde{\psi}_{n,\pi(2j-1),l})^2 | D^{(n)}, X^{(n)} \right] \le \max_{1\le i\le n} \max_{1\le l\le p_n} CE[\psi_{n,i,l}^2 | X_i] \le C$$

for some constant C. This implies

$$\left\| \frac{1}{n} \sum_{1\le j\le n} (\tilde{\psi}_{n,\pi(2j)} - \tilde{\psi}_{n,\pi(2j-1)})(\Psi(X_{\pi(2j)}) - \Psi(X_{\pi(2j-1)}))' \beta_n^{\text{hd}} \right\|_\infty = O_P(\log^{1/2}(p_n) \zeta_n \Xi_n n^{-1/2}) = o_P(\zeta_n).$$

The second result in Lemma B.5 is a direct consequence of the Cauchy Schwartz inequality and Assumption 5.2(a).

For the third result in Lemma B.5, we have

$$\begin{aligned}
&\left\|\frac{1}{n}\sum_{1\le j\le n}(\psi_{n,\pi(2j)}-\psi_{n,\pi(2j-1)})(\tilde\psi_{n,\pi(2j)}+\tilde\psi_{n,\pi(2j-1)})'\left(\frac{\beta^{\text{hd-pd}}_{D_{\pi(2j)},n}-\beta^{\text{hd-pd}}_{D_{\pi(2j-1)},n}}{2}\right)\right\|_\infty\\
&\le\left\|\frac{1}{n}\sum_{1\le j\le n}(\tilde\psi_{n,\pi(2j)}-\tilde\psi_{n,\pi(2j-1)})(\tilde\psi_{n,\pi(2j)}+\tilde\psi_{n,\pi(2j-1)})'\left(\frac{\beta^{\text{hd-pd}}_{D_{\pi(2j)},n}-\beta^{\text{hd-pd}}_{D_{\pi(2j-1)},n}}{2}\right)\right\|_\infty\\
&+\frac{1}{n}\sum_{1\le j\le n}||X_{\pi(2j)}-X_{\pi(2j-1)}||_2\left|(\tilde\psi_{n,\pi(2j)}+\tilde\psi_{n,\pi(2j-1)})'\left(\frac{\beta^{\text{hd-pd}}_{D_{\pi(2j)},n}-\beta^{\text{hd-pd}}_{D_{\pi(2j-1)},n}}{2}\right)\right|\\
&\le\left\|\frac{1}{n}\sum_{1\le j\le n}(\tilde\psi_{n,\pi(2j)}-\tilde\psi_{n,\pi(2j-1)})(\tilde\psi_{n,\pi(2j)}+\tilde\psi_{n,\pi(2j-1)})'\left(\frac{\beta^{\text{hd-pd}}_{D_{\pi(2j)},n}-\beta^{\text{hd-pd}}_{D_{\pi(2j-1)},n}}{2}\right)\right\|_\infty\\
&+\left[\frac{1}{n}\sum_{1\le j\le n}||X_{\pi(2j)}-X_{\pi(2j-1)}||_2^2\right]^{1/2}\left[\frac{1}{n}\sum_{1\le j\le n}\left|(\tilde\psi_{n,\pi(2j)}+\tilde\psi_{n,\pi(2j-1)})'\left(\frac{\beta^{\text{hd-pd}}_{D_{\pi(2j)},n}-\beta^{\text{hd-pd}}_{D_{\pi(2j-1)},n}}{2}\right)\right|^2\right]\\
&\le\left\|\frac{1}{n}\sum_{1\le j\le n}(\tilde\psi_{n,\pi(2j)}\tilde\psi'_{n,\pi(2j)}-\tilde\psi_{n,\pi(2j-1)}\tilde\psi'_{n,\pi(2j-1)})\left(\frac{\beta^{\text{hd-pd}}_{D_{\pi(2j)},n}-\beta^{\text{hd-pd}}_{D_{\pi(2j-1)},n}}{2}\right)\right\|_\infty\\
&+\left\|\frac{1}{n}\sum_{1\le j\le n}\tilde\psi_{n,\pi(2j)}\tilde\psi'_{n,\pi(2j-1)}\left(\frac{\beta^{\text{hd-pd}}_{D_{\pi(2j)},n}-\beta^{\text{hd-pd}}_{D_{\pi(2j-1)},n}}{2}\right)\right\|_\infty\\
&+\left\|\frac{1}{n}\sum_{1\le j\le n}\tilde\psi_{n,\pi(2j-1)}\tilde\psi'_{n,\pi(2j)}\left(\frac{\beta^{\text{hd-pd}}_{D_{\pi(2j)},n}-\beta^{\text{hd-pd}}_{D_{\pi(2j-1)},n}}{2}\right)\right\|_\infty+O_P(\zeta_n)\\
&\equiv R_{n,1}+R_{n,2}+R_{n,3}+O_P(\zeta_n).
\end{aligned}$$

For $R_{n,1}$, we note that, conditional on $(D^{(n)},X^{(n)})$,

$$\left\{(\tilde\psi_{n,\pi(2j)}\tilde\psi'_{n,\pi(2j)}-\tilde\psi_{n,\pi(2j-1)}\tilde\psi'_{n,\pi(2j-1)})\left(\frac{\beta^{\text{hd-pd}}_{D_{\pi(2j)},n}-\beta^{\text{hd-pd}}_{D_{\pi(2j-1)},n}}{2}\right)\right\}_{1\le j\le n}$$

are independent and mean zero. In addition, we have

$$\begin{aligned}
&\max_{1\le l\le p_n}\frac{1}{n}\sum_{1\le j\le n}\left((\tilde\psi_{n,\pi(2j),l}\tilde\psi'_{n,\pi(2j)}-\tilde\psi_{n,\pi(2j-1),l}\tilde\psi'_{n,\pi(2j-1)})\left(\frac{\beta^{\text{hd-pd}}_{D_{\pi(2j)},n}-\beta^{\text{hd-pd}}_{D_{\pi(2j-1)},n}}{2}\right)\right)^2\\
&\lesssim\max_{1\le l\le p_n}\left[\frac{1}{2n}\sum_{1\le i\le 2n}\tilde\psi^q_{n,i,l}\right]^{2/q}\left[\frac{1}{2n}\sum_{1\le i\le 2n}|\tilde\psi'_{n,i,l}\beta^{\text{hd-pd}}_{1,n}|^{2q/(q-2)}+|\tilde\psi'_{n,i,l}\beta^{\text{hd-pd}}_{0,n}|^{2q/(q-2)}\right]^{q/(q-2)}=O_P(1)
\end{aligned}$$

and

$$\max_{1\le l\le p_n}E\left\{\frac{1}{n}\sum_{1\le j\le n}\left((\tilde\psi_{n,\pi(2j),l}\tilde\psi'_{n,\pi(2j)}-\tilde\psi_{n,\pi(2j-1),l}\tilde\psi'_{n,\pi(2j-1)})\left(\frac{\beta^{\text{hd-pd}}_{D_{\pi(2j)},n}-\beta^{\text{hd-pd}}_{D_{\pi(2j-1)},n}}{2}\right)\right)^2\Big|D^{(n)},X^{(n)}\right\}$$

$$\lesssim \max_{1\leq l\leq p_n}\left[E\left\{\frac{1}{n}\sum_{1\leq i\leq 2n}\tilde{\psi}_{n,i,l}^{q}\middle|X_i\right\}\right]^{2/q}\left[\frac{1}{2n}\sum_{1\leq i\leq 2n}E\left\{|\tilde{\psi}'_{n,i,l}\beta_{1,n}^{\mathrm{hd-pd}}|^{2q/(q-2)}+|\tilde{\psi}'_{n,i,l}\beta_{0,n}^{\mathrm{hd-pd}}|^{2q/(q-2)}\middle|X_i\right\}\right]^{q/(q-2)}$$
$$= O(1).$$

Therefore, by the same argument as that used to bound the first term on the RHS of (77), we have

$$\max_{1\leq l\leq p_n}\frac{1}{n}\sum_{1\leq j\leq n}\left[(\tilde{\psi}_{n,\pi(2j),l}\tilde{\psi}'_{n,\pi(2j)}-\tilde{\psi}_{n,\pi(2j-1),l}\tilde{\psi}'_{n,\pi(2j-1)})\left(\frac{\beta^{\mathrm{hd-pd}}_{D_{\pi(2j)},n}-\beta^{\mathrm{hd-pd}}_{D_{\pi(2j-1)},n}}{2}\right)\right]=O_P\left(\sqrt{\frac{\log(p_n)}{n}}\right).$$

Also note that, conditional on $(D^{(n)},X^{(n)})$,

$$\left\{\tilde{\psi}_{n,\pi(2j)}\tilde{\psi}'_{n,\pi(2j-1)}\left(\frac{\beta^{\mathrm{hd-pd}}_{D_{\pi(2j)},n}-\beta^{\mathrm{hd-pd}}_{D_{\pi(2j-1)},n}}{2}\right)\right\}_{1\leq j\leq n}\quad\text{and}\quad\left\{\tilde{\psi}_{n,\pi(2j-1)}\tilde{\psi}'_{n,\pi(2j)}\left(\frac{\beta^{\mathrm{hd-pd}}_{D_{\pi(2j)},n}-\beta^{\mathrm{hd-pd}}_{D_{\pi(2j-1)},n}}{2}\right)\right\}_{1\leq j\leq n}$$

are independent cross j and mean zero. By the same argument as that used to bound the first term on the RHS of (77), we can show

$$\left\|\frac{1}{n}\sum_{1\leq j\leq n}\tilde{\psi}_{n,\pi(2j)}\tilde{\psi}'_{n,\pi(2j-1)}\left(\frac{\beta^{\mathrm{hd-pd}}_{D_{\pi(2j)},n}-\beta^{\mathrm{hd-pd}}_{D_{\pi(2j-1)},n}}{2}\right)\right\|_\infty=O_P\left(\sqrt{\frac{\log(p_n)}{n}}\right)$$

and

$$\left\|\frac{1}{n}\sum_{1\leq j\leq n}\tilde{\psi}_{n,\pi(2j-1)}\tilde{\psi}'_{n,\pi(2j)}\left(\frac{\beta^{\mathrm{hd-pd}}_{D_{\pi(2j)},n}-\beta^{\mathrm{hd-pd}}_{D_{\pi(2j-1)},n}}{2}\right)\right\|_\infty=O_P\left(\sqrt{\frac{\log(p_n)}{n}}\right).$$

This leads to the desired result.

The fourth result holds because conditionally on (D^n,X^n), $\{\tilde{\psi}_{n,\pi(2j-1)}+\tilde{\psi}_{n,\pi(2j)}\}_{j\in[n]}$ are mean zero and independent.

For the fifth result in Lemma B.5, we note that, conditional on $(D^{(n)},X^{(n)})$,

$$\{\psi_{n,\pi(2j)}-\Psi(X_{\pi(2j)})\epsilon_{n,\pi(2j-1)}(D_{\pi(2j-1)})\}_{1\leq j\leq n}$$

are independent and mean zero. In addition, there exist constants (b,C) such that

$$\begin{aligned}
0<b&\leq\min_{1\leq l\leq p_n}\frac{1}{n}\sum_{1\leq j\leq n}E\left\{(\psi_{n,\pi(2j),l}-\Psi(X_{\pi(2j),l})\epsilon_{n,\pi(2j-1)}(D_{\pi(2j-1)}))^2\middle|D^{(n)},X^{(n)}\right\}\\
&\leq\max_{1\leq l\leq p_n}\frac{1}{n}\sum_{1\leq j\leq n}E\left\{(\psi_{n,\pi(2j),l}-\Psi(X_{\pi(2j),l})\epsilon_{n,\pi(2j-1)}(D_{\pi(2j-1)}))^2\middle|D^{(n)},X^{(n)}\right\}\\
&\leq\max_{1\leq l\leq p_n}\frac{C}{n}\sum_{1\leq i\leq 2n}\left(E\left\{\epsilon^2_{n,i}(1)+\epsilon^2_{n,i}(0)\middle|X_i\right\}\right)\\
&\leq C<\infty.
\end{aligned}$$

Therefore, following Theorem 2.1 in Chernozhukov et al. (2017), Lemma E.2 in Chetverikov and Sørensen (2022), and the coupling argument used to bound $\mathcal{E}_{n,1}(d)$ in the proof of Lemma B.4, we have

$$\left\| \frac{1}{n} \sum_{1 \le j \le n} (\psi_{n,\pi(2j)} - \Psi(X_{\pi(2j)}))\epsilon_{n,\pi(2j-1)}(D_{\pi(2j-1)}) \right\|_\infty = O_P\left(\sqrt{\frac{\log(p_n)}{n}} \right) . \tag{78}$$

The last result of Lemma B.5 can be established in the same manner.

■

Lemma B.6. *Suppose Assumptions in Theorem 5.2 hold. Then,*

$$P\left\{ \left| \frac{1}{n} \sum_{i \in [2n]} I\{D_i = d\}\epsilon_{n,i}(d) - E[\epsilon_{n,i}(d)] \right| \le \sqrt{\frac{\log(2np_n)}{n}} \right\} \to 1$$

and

$$P\left\{ \left\| \Omega_n^{-1}(d) \frac{1}{n} \sum_{1 \le i \le 2n} I\{D_i = d\}(\psi_{n,i}\epsilon_{n,i}(d) - E[\psi_{n,i}\epsilon_{n,i}(d)]) \right\|_\infty \le \frac{6\bar{\sigma}}{\underline{\sigma}} \sqrt{\frac{\log(2np_n)}{n}} \right\} \to 1 \ .$$

Proof. For the first result, we note that

$$\begin{aligned} \left| \frac{1}{n} \sum_{i \in [2n]} I\{D_i = d\}\epsilon_{n,i}(d) - E[\epsilon_{n,i}(d)] \right| \le & \left| \frac{1}{n} \sum_{i \in [2n]} I\{D_i = d\}(\epsilon_{n,i}(d) - E[\epsilon_{n,i}(d)|X_i]) \right| \\ & + \left| \frac{1}{n} \sum_{i \in [2n]} (I\{D_i = d\} - 1/2)(E[\epsilon_{n,i}(d)|X_i] - E[\epsilon_{n,i}(d)]) \right| \\ & + \left| \frac{1}{2n} \sum_{i \in [2n]} (E[\epsilon_{n,i}(d)|X_i] - E[\epsilon_{n,i}(d)]) \right| . \end{aligned}$$

The first two terms on the RHS of the above display are $O_P(1/\sqrt{n})$ following the proof of $\mathcal{E}_{n,1}(d)$ and $\mathcal{E}_{n,2}(d)$ in Lemma B.4. The last term on the RHS is also $O_P(1/\sqrt{n})$ by Chebyshev's inequality. This implies the desired result.

For the second result, define

$$\mathcal{E}_{n,0}(d) = \begin{pmatrix} \max_{d \in \{0,1\}} \frac{1}{2n} \sum_{1 \le i \le 2n} E[\epsilon_{n,i}^4(d)|X_i] \le c_0 < \infty \ , \\ \min_{1 \le l \le p_n} \frac{1}{n} \sum_{1 \le i \le 2n} I\{D_i = d\} \operatorname{Var}[\psi_{n,i,l}\epsilon_{n,i}(d)|X_i] \ge \underline{\sigma}^2 > 0 \ , \end{pmatrix}$$

$$\mathcal{E}_{n,1}(d) = \left\{ \left\| \frac{1}{n} \sum_{1 \le i \le 2n} I\{D_i = d\}(\psi_{n,i}\epsilon_{n,i}(d) - E[\psi_{n,i}\epsilon_{n,i}(d)|X_i]) \right\|_\infty \le 2.04\bar{\sigma}\sqrt{\log(2np_n)/n} \right\} \ ,$$

$$\mathcal{E}_{n,2}(d) = \left\{ \left\| \frac{1}{n} \sum_{1 \le i \le 2n} I\{D_i = d\}(E[\psi_{n,i}\epsilon_{n,i}(d)|X_i] - E[\psi_{n,i}\epsilon_{n,i}(d)]) \right\|_\infty \le 3.96\bar{\sigma}\sqrt{\log(2np_n)/n} \right\} ,$$

$$\mathcal{E}_{n,3}(d) = \left\{ \max_{1 \le l \le p_n} \left| \frac{1}{2n} \sum_{1 \le i \le 2n} (E[\psi^2_{n,i,l}\epsilon^2_{n,i}(d)|X_i] - E[\psi^2_{n,i,l}\epsilon^2_{n,i}(d)]) \right|^{1/2} \le 0.01\bar{\sigma} \right\} ,$$

and

$$\mathcal{E}_{n,4}(d) = \left\{ \max_{1 \le l \le p_n} \left| \frac{1}{2n} \sum_{1 \le i \le 2n} (2I\{D_i = d\} - 1)(E[\epsilon^2_{n,i}(d)\psi^2_{n,i,l}|X_i] - E[\epsilon^2_{n,i}(d)\psi^2_{n,i,l}]) \right|^{1/2} \le 0.01\bar{\sigma} \right\} .$$

We aim to show that $P\{\mathcal{E}_{n,1}(d)\} \to 1$ and $P\{\mathcal{E}_{n,2}(d)\} \to 1$. Then, by letting $C = 6\bar{\sigma}/\underline{\sigma}$ which implies

$$\begin{aligned}
P\{\mathcal{E}_n(d)\} &= 1 - P\{\mathcal{E}^c_n(d)\} \\
&\ge 1 - P\left\{ \left\| \frac{1}{n} \sum_{1 \le i \le 2n} I\{D_i = d\}(\psi_{n,i}\epsilon_{n,i}(d) - E[\psi_{n,i}\epsilon_{n,i}(d)]) \right\|_\infty \ge C\underline{\sigma}\sqrt{\frac{\log(2np_n)}{n}} \right\} \\
&= 1 - P\left\{ \left\| \frac{1}{n} \sum_{1 \le i \le 2n} I\{D_i = d\}(\psi_{n,i}\epsilon_{n,i}(d) - E[\psi_{n,i}\epsilon_{n,i}(d)]) \right\|_\infty \ge 6\bar{\sigma}\sqrt{\frac{\log(2np_n)}{n}} \right\} \\
&\ge 1 - P\{\mathcal{E}^c_{n,1}(d)\} - P\{\mathcal{E}^c_{n,2}(d)\} \to 1 .
\end{aligned}$$

First, we show $P\{\mathcal{E}_{n,3}(d)\} \to 1$. Let

$$t_n = C\sqrt{\frac{\log(np_n)\Xi^2_n}{n}} \to 0$$

for some sufficiently large constant $C > 0$ and $\{e_i\}_{1 \le i \le 2n}$ be a sequence of i.i.d. Rademacher random variables independent of everything else. Then, for any fixed $t > 0$, we have

$$\begin{aligned}
&\left(1 - \frac{4\max_{1 \le l \le p_n} \operatorname{Var}[E[\psi^2_{n,i,l}\epsilon^2_{n,i}(d)|X_i]]}{2nt^2}\right) P\left\{ \max_{1 \le l \le p_n} \left| \frac{1}{2n} \sum_{1 \le i \le 2n} \left[E[\psi^2_{n,i,l}\epsilon^2_{n,i}(d)|X_i] - E[\psi^2_{n,i,l}\epsilon^2_{n,i}(d)]\right] \right| \ge t \right\} \\
&\le 2P\left\{ \max_{1 \le l \le p_n} \left| \frac{1}{2n} \sum_{1 \le i \le 2n} 4e_i E[\psi^2_{n,i,l}\epsilon^2_{n,i}(d)|X_i] \right| \ge t \right\} \\
&= o(1) + 2E\left[P\left\{ \max_{1 \le l \le p_n} \left| \frac{1}{2n} \sum_{1 \le i \le 2n} 4e_i E[\psi^2_{n,i,l}\epsilon^2_{n,i}(d)|X_i] \right| \ge t \middle| X^{(n)} \right\} I\{\mathcal{E}_{n,0}(d)\} \right] \\
&\lesssim o(1) + p_n \exp\left(-\frac{nt^2}{\Xi^2_n c}\right) = o(1),
\end{aligned}$$

where the first inequality is by van der Vaart and Wellner (1996, Lemma 2.3.7), the second inequality is by the Hoeffding's inequality conditional on $X^{(n)}$ and the fact that, on $\mathcal{E}_{n,0}(d)$,

$$\begin{aligned}\frac{1}{2n}\sum_{1\le i\le 2n}(E[\psi^2_{n,i,l}\epsilon^2_{n,i}(d)|X_i])^2 &\le \frac{1}{2n}\sum_{1\le i\le 2n}E[\psi^4_{n,i,l}|X_i]E[\epsilon^4_{n,i}(d)|X_i]\\ &\le \frac{\Xi_n^2}{2n}\sum_{1\le i\le 2n}E[\psi^2_{n,i,l}|X_i]E[\epsilon^4_{n,i}(d)|X_i]\le \Xi_n^2 Cc_0,\end{aligned}$$

where C is a fixed constant, and the last equality is by the fact that $\log(p_n)\Xi_n^2 = o(n)$. Furthermore, we note that

$$\begin{aligned}&\frac{4\max_{1\le l\le p_n}\mathrm{Var}[E[\psi^2_{n,i,l}\epsilon^2_{n,i}(d)|X_i]]}{2n}\\ &\lesssim \frac{\max_{1\le l\le p_n}E\left[E[\psi^4_{n,i,l}|X_i]E[\epsilon^4_{n,i}(d)|X_i]\right]}{n}\\ &\lesssim \frac{\Xi_n^2\max_{1\le l\le p_n}E\left[E[\psi^2_{n,i,l}|X_i]E[\epsilon^4_{n,i}(d)|X_i]\right]}{n}\\ &\lesssim \frac{\Xi_n^2 E[\epsilon^4_{n,i}(d)]}{n}\\ &= o(1).\end{aligned}$$

Therefore, we have

$$P\left\{\max_{1\le l\le p_n}\left|\frac{1}{2n}\sum_{1\le i\le 2n}\left[E[\psi^2_{n,i,l}\epsilon^2_{n,i}(d)|X_i]-E[\psi^2_{n,i,l}\epsilon^2_{n,i}(d)]\right]\right|\ge t\right\}=o(1)$$

for any fixed $t>0$, which is the desired result.

Next, we show $P\{\mathcal{E}_{n,4}(d)\}\to 1$. Define $a_{n,i,l}=E[\epsilon^2_{n,i}(d)\psi^2_{n,i,l}|X_i]-E[\epsilon^2_{n,i}(d)\psi^2_{n,i,l}]$. Then, we have

$$\begin{aligned}&P\left\{\max_{1\le l\le p_n}\left|\frac{1}{2n}\sum_{1\le i\le 2n}(2I\{D_i=d\}-1)(E[\epsilon^2_{n,i}(d)\psi^2_{n,i,l}|X_i]-E[\epsilon^2_{n,i}(d)\psi^2_{n,i,l}])\right|>t\middle|X^{(n)}\right\}I\{\mathcal{E}_{n,0}(d)\}\\ &\le \sum_{1\le l\le p_n}P\left\{\left|\frac{1}{2n}\sum_{1\le j\le n}(I\{D_{\pi(2j-1)}=d\}-I\{D_{\pi(2j)}=d\})(a_{n,\pi(2j-1),l}-a_{n,\pi(2j),l})\right|>t\middle|X^{(n)}\right\}I\{\mathcal{E}_{n,0}(d)\}\\ &\le \sum_{1\le l\le p_n}\exp\left(-\frac{2nt^2}{\frac{1}{n}\sum_{1\le j\le n}(a_{n,\pi(2j-1),l}-a_{n,\pi(2j),l})^2}\right)I\{\mathcal{E}_{n,0}(d)\}\\ &\le \exp\left(\log(p_n)-\frac{2nt^2}{\Xi_n^2c^2}\right)\ ,\end{aligned}$$

where, conditional on $X^{(n)}$, $\{I\{D_{\pi(2j-1)}=d\}-I\{D_{\pi(2j)}=d\}\}_{1\le j\le n}$ is a sequence of i.i.d. Rademacher random variables, the second last inequality is by Hoeffding's inequality, and the last inequality is by that, on $\mathcal{E}_{n,0}(d)$,

$$\begin{aligned}
&\left(\frac{1}{n}\sum_{1\leq j\leq n}(a_{n,\pi(2j-1),l}-a_{n,\pi(2j),l})^2\right)^{1/2}\\
&\leq\left(\frac{1}{n}\sum_{1\leq j\leq n}(E[\psi^2_{n,\pi(2j-1),l}\epsilon^2_{n,i}(d)|X_{\pi(2j-1)}])^2\right)^{1/2}+\left(\frac{1}{n}\sum_{1\leq j\leq n}(E[\psi^2_{n,\pi(2j),l}\epsilon^2_{n,i}(d)|X_{\pi(2j)}])^2\right)^{1/2}\\
&\leq\left(\frac{2}{n}\sum_{1\leq i\leq 2n}(E[\psi^2_{n,i,l}\epsilon^2_{n,i}(d)|X_i])^2\right)^{1/2}\\
&\leq \Xi_n c\ .
\end{aligned}$$

By letting $t=C\sqrt{\frac{\log(p_n)\Xi_n^2}{n}}$ for some sufficiently large C and noting that $P\{\mathcal{E}_{n,0}(d)\}\to 1$, we have

$$\max_{1\leq l\leq p_n}\left|\sum_{1\leq i\leq 2n}\frac{(2I\{D_i=d\}-1)(E[\epsilon^2_{n,i}(d)\psi^2_{n,i,l}|X_i]-E[\epsilon^2_{n,i}(d)\psi^2_{n,i,l}])}{2n}\right|=O_p\left(\sqrt{\frac{\log(p_n)\Xi_n^2}{n}}\right)\ ,$$

and thus, $P\{\mathcal{E}_{n,4}(d)\}\to 1$.

Next, we show $P\{\mathcal{E}_{n,1}(d)\}\to 1$. We note that, for $d\in\{0,1\}$, conditional on $(D^{(n)},X^{(n)})$, $\{\psi_{n,i}\epsilon_{n,i}(d)\}_{1\leq i\leq 2n}$ are independent. In what follows, we couple

$$\mathbb{U}_n=\frac{1}{n}\sum_{1\leq i\leq 2n}I\{D_i=d\}(\psi_{n,i}\epsilon_{n,i}(d)-E[\psi_{n,i}\epsilon_{n,i}(d)|X_i])$$

with a centered Gaussian random vector as in Theorem 2.1 in Chernozhukov et al. (2017). Let $Z=(Z_1,\ldots,Z_{p_n})$ be a Gaussian random vector with $E[Z_l]=0$ for $1\leq l\leq p_n$ and $\text{Var}[Z]=\text{Var}[\mathbb{U}_n|X^{(n)},D^{(n)}]$ that additionally satisfies the conditions of that theorem. Specifically, $Z=(Z_1,\cdots,Z_{p_n})$ is a centered Gaussian random vector in R^{p_n} such that on $\mathcal{E}_{n,0}(d)\cap\mathcal{E}_{n,3}(d)\cap\mathcal{E}_{n,4}(d)$,

$$\begin{aligned}
E[ZZ']=&\frac{1}{n^2}\sum_{1\leq i\leq 2n}I\{D_i=d\}E[\epsilon^2_{n,i}(d)\psi_{n,i}\psi'_{n,i}|X_i]\\
&-\frac{1}{n}\left(\frac{1}{n}\sum_{1\leq i\leq 2n}I\{D_i=d\}E[\epsilon_{n,i}(d)\psi_{n,i}|X_i]\right)\left(\frac{1}{n}\sum_{1\leq i\leq 2n}I\{D_i=d\}E[\epsilon_{n,i}(d)\psi_{n,i}|X_i]\right)'
\end{aligned}$$

and

$$\begin{aligned}
\max_{1\leq l\leq p_n}E[Z_l^2]\leq&\frac{\max_{1\leq l\leq p_n}\sum_{1\leq i\leq 2n}I\{D_i=d\}E[\epsilon^2_{n,i}(d)\psi^2_{n,i,l}|X_i]}{n^2}\\
\leq&\frac{\bar{\sigma}^2}{n}+\frac{\max_{1\leq l\leq p_n}\sum_{1\leq i\leq 2n}I\{D_i=d\}(E[\epsilon^2_{n,i}(d)\psi^2_{n,i,l}|X_i]-E[\epsilon^2_{n,i}(d)\psi^2_{n,i,l}])}{n^2}\\
\leq&\frac{\bar{\sigma}^2}{n}+\frac{\max_{1\leq l\leq p_n}\sum_{1\leq i\leq 2n}(2I\{D_i=d\}-1)(E[\epsilon^2_{n,i}(d)\psi^2_{n,i,l}|X_i]-E[\epsilon^2_{n,i}(d)\psi^2_{n,i,l}])}{2n^2}\\
&+\frac{\max_{1\leq l\leq p_n}\sum_{1\leq i\leq 2n}(E[\epsilon^2_{n,i}(d)\psi^2_{n,i,l}|X_i]-E[\epsilon^2_{n,i}(d)\psi^2_{n,i,l}])}{2n^2}
\end{aligned}$$

$$\leq \frac{1.02\overline{\sigma}^2}{n} \ .$$

Further define $q(1-\alpha)$ as the $(1-\alpha)$ quantile of $||Z||_\infty$. Then, we have

$$q(1-1/n) \leq \frac{1.02\overline{\sigma}(\sqrt{2\log(2p_n)} + \sqrt{2\log(n)})}{\sqrt{n}} \leq 2.04\overline{\sigma}\sqrt{\log(2np_n)/n},$$

where the first inequality is by the last display in the proof of Lemma E.2 in Chetverikov and Sørensen (2022) and the second inequality is by the fact that $\sqrt{a}+\sqrt{b} \leq \sqrt{2(a+b)}$ for $a, b > 0$. Therefore, we have

$$\begin{aligned}
P\{\mathcal{E}_{n,1}^c(d)) &\leq P\{\mathcal{E}_{n,1}^c(d), \mathcal{E}_{n,0}(d), \mathcal{E}_{n,3}(d), \mathcal{E}_{n,4}(d)\} + o(1) \\
&= EP\{\mathcal{E}_{n,1}^c(d)|D^{(n)}, X^{(n)}\}I\{\mathcal{E}_{n,0}(d), \mathcal{E}_{n,3}(d), \mathcal{E}_{n,4}(d)\} + o(1) \\
&\leq E[P\{||Z||_\infty \geq 2.04\overline{\sigma}\sqrt{\log(2np_n)/n}|D^{(n)}, X^{(n)})\}I\{\mathcal{E}_{n,0}(d), \mathcal{E}_{n,3}(d), \mathcal{E}_{n,4}(d)\}] + o(1) \\
&\leq E[P\{||Z||_\infty \geq q(1-1/n)|D^{(n)}, X^{(n)}\}] = o(1),
\end{aligned}$$

where the second inequality is by Theorem 2.1 in Chernozhukov et al. (2017).

Finally, we turn to $\mathcal{E}_{n,2}(d)$ with $d = 1$. We have

$$\begin{aligned}
&\frac{1}{n}\sum_{1\leq i\leq 2n} I\{D_i = 1\}(E[\psi_{n,i}\epsilon_{n,i}(1)|X_i] - E[\psi_{n,i}\epsilon_{n,i}(1)]) \\
&= \frac{1}{2n}\sum_{1\leq i\leq 2n}(E[\psi_{n,i}\epsilon_{n,i}(1)|X_i] - E[\psi_{n,i}\epsilon_{n,i}(1)]) + \frac{1}{2n}\sum_{1\leq i\leq 2n}(2D_i - 1)(E[\psi_{n,i}\epsilon_{n,i}(1)|X_i] - E[\psi_{n,i}\epsilon_{n,i}(1)]).
\end{aligned} \tag{79}$$

Note $\{E[\psi_{n,i}\epsilon_{n,i}(1)|X_i] - E[\psi_{n,i}\epsilon_{n,i}(1)]\}_{1\leq i\leq 2n}$ is a sequence of independent centered random variables and

$$\max_{1\leq l\leq p_n} E[(E[\psi_{n,i,l}\epsilon_{n,i}(1)|X_i] - E[\psi_{n,i,l}\epsilon_{n,i}(1)])^2] \leq \overline{\sigma}^2.$$

Following Theorem 2.1 in Chernozhukov et al. (2017), Lemma E.2 in Chetverikov and Sørensen (2022), and similar arguments to the ones above, we have

$$P\left\{\left\|\frac{1}{2n}\sum_{1\leq i\leq 2n}(E[\psi_{n,i}\epsilon_{n,i}(1)|X_i] - E[\psi_{n,i}\epsilon_{n,i}(1)])\right\|_\infty \leq \overline{\sigma}\sqrt{2\log(2np_n)/n}\right) \to 1. \tag{80}$$

For the second term on the RHS of (79), we define $g_{n,i,l} = E[\psi_{n,i,l}\epsilon_{n,i}(1)|X_i] - E[\psi_{n,i,l}\epsilon_{n,i}(1)]$. We have

$$\begin{aligned}
&P\left\{\left\|\frac{1}{2n}\sum_{1\leq i\leq 2n}(2D_i - 1)E[\psi_{n,i}\epsilon_{n,i}(1)|X_i] - E[\psi_{n,i}\epsilon_{n,i}(1)]\right\|_\infty > t\middle|X^{(n)}\right\} \\
&\leq \sum_{1\leq l\leq p_n} P\left\{\left|\frac{1}{2n}\sum_{1\leq j\leq n}(D_{\pi(2j-1)} - D_{\pi(2j)})(g_{n,\pi(2j-1),l} - g_{n,\pi(2j),l})\right| > t\middle|X^{(n)}\right\} \\
&\leq \sum_{1\leq l\leq p_n} \exp\left(-\frac{2nt^2}{\frac{1}{n}\sum_{1\leq j\leq n}(g_{n,\pi(2j-1),l} - g_{n,\pi(2j),l})^2}\right) \ ,
\end{aligned}$$

where, conditional on $X^{(n)}$, $\{(D_{\pi(2j-1)} - D_{\pi(2j)})\}_{1\leq j\leq n}$ is a sequence of i.i.d. Rademacher random variables and the last inequality is by Hoeffding's inequality. In addition, on $\mathcal{E}_{n,3}(1)$, we have

$$\begin{aligned}
&\left(\frac{1}{n}\sum_{1\leq j\leq n}(g_{n,\pi(2j-1),l} - g_{n,\pi(2j),l})^2\right)^{1/2} \\
&\leq \left(\frac{1}{n}\sum_{1\leq j\leq n}(E[\psi_{n,\pi(2j-1),l}\epsilon_{n,i}(1)|X_{\pi(2j-1)}])^2\right)^{1/2} + \left(\frac{1}{n}\sum_{1\leq j\leq n}(E[\psi_{n,\pi(2j),l}\epsilon_{n,i}(1)|X_{\pi(2j)}])^2\right)^{1/2} \\
&\leq \left(\frac{2}{n}\sum_{1\leq i\leq 2n}(E[\psi_{n,i,l}\epsilon_{n,i}(1)|X_i])^2\right)^{1/2} \\
&\leq \left(\frac{2}{n}\sum_{1\leq i\leq 2n}E[\psi_{n,i,l}^2\epsilon_{n,i}^2(1)|X_i]\right)^{1/2} \\
&\leq \left(\frac{2}{n}\sum_{1\leq i\leq 2n}\left[E[\psi_{n,i,l}^2\epsilon_{n,i}^2(1)|X_i] - E[\psi_{n,i,l}^2\epsilon_{n,i}^2(1)]\right]\right)^{1/2} + 2\overline{\sigma} \\
&\leq 2.02\overline{\sigma}.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&P\left\{\left\|\frac{1}{2n}\sum_{1\leq i\leq 2n}(2D_i - 1)E[\psi_{n,i}\epsilon_{n,i}(1)|X_i] - E[\psi_{n,i}\epsilon_{n,i}(1)]\right\|_\infty > 2.02\sqrt{\frac{\log(np_n)\overline{\sigma}^2}{n}}\right\} \\
&\leq P\left\{\left\|\frac{1}{2n}\sum_{1\leq i\leq 2n}(2D_i - 1)E[\psi_{n,i}\epsilon_{n,i}(1)|X_i] - E[\psi_{n,i}\epsilon_{n,i}(1)]\right\|_\infty > 2.02\sqrt{\frac{\log(np_n)\overline{\sigma}^2}{n}}, \mathcal{E}_{n,3}(1)\right\} + o(1) \\
&\leq E\left[P\left\{\left\|\frac{1}{2n}\sum_{1\leq i\leq 2n}(2D_i - 1)E[\psi_{n,i}\epsilon_{n,i}(1)|X_i] - E[\psi_{n,i}\epsilon_{n,i}(1)]\right\|_\infty > 2.02\sqrt{\frac{\log(np_n)\overline{\sigma}^2}{n}}\middle|X^{(n)}\right\} I\{\mathcal{E}_{n,3}(1)\}\right] + o(1) \\
&= o(1)\ . \qquad (81)
\end{aligned}$$

Combining (79), (80), (81), and the fact that $\sqrt{2} + 2.02 \leq 3.98$, we have $P\{\mathcal{E}_{n,2}(1)\} \to 1$. The same result holds for $\mathcal{E}_{n,2}(0)$. ■

C Details for Simulations

The regressors in both the first and second LASSO-based adjustments are as follows.

(i) For Models 1-6, we use $\{1, X_i, W_i, X_i^2, W_i^2, X_iW_i, (X_i - \tilde{X})I\{X_i > \tilde{X}\}, (W_i - \tilde{W})I\{W_i > \tilde{W}\}, (X_i - \tilde{X})^2 I\{X_i > \tilde{X}\}, (W_i - \tilde{W})^2 I\{W_i > \tilde{W}\}\}$ where $\tilde{X}$ and $\tilde{W}$ are the sample medians of $\{X_i\}_{i\in[2n]}$ and $\{W_i\}_{i\in[2n]}$, respectively.

(ii) For Models 7-9, we use $\{1, X_i, W_i, X_i^2, W_i^2, X_iW_i, (X_{ij} - \tilde{X}_j)I\{X_{ij} > \tilde{X}_j\}, (X_{ij} - \tilde{X}_j)^2 I\{X_{ij} > \tilde{X}_j\}, (W_{ij} -$

$\tilde{W}_1)I\{W_{ij} > \tilde{W}_j\}, (W_{ij} - \tilde{W}_j)^2 I\{W_{ij} > \tilde{W}_j\}\}$ where $\tilde{X}_j$ and $\tilde{W}_j$, for $j = 1, 2$, are the sample medians of $\{X_{ij}\}_{i\in[2n]}$ and $\{W_{ij}\}_{i\in[2n]}$, respectively.

(iii) For Models 9-11, we use $\{1, X_i, W_i, X_i^2, W_i^2, X_{i1}W_{i1}, X_{i2}W_{i2}, X_{i3}W_{i1}, X_{i4}W_{i2}, (X_{ij} - \tilde{X}_j)I\{X_{ij} > \tilde{X}_j\}, (X_{ij} - \tilde{X}_j)^2 I\{X_{ij} > \tilde{X}_j\}, (W_{ij} - \tilde{W}_j)I\{W_{ij} > \tilde{W}_j\}, (W_{ij} - \tilde{W}_j)^2 I\{W_{ij} > \tilde{W}_j\}\}$ where $\tilde{X}_j$,for $j = 1, 2, 3, 4$, and $\tilde{W}_j$, for $j = 1, 2$, are the sample medians of $\{X_{ij}\}_{i\in[2n]}$ and $\{W_{ij}\}_{i\in[2n]}$, respectively.

(iv) Models 12-15 already contain high-dimensional covariates. We just use X_i and W_i as the LASSO regressors.

References

Armstrong, T. B. (2022). Asymptotic Efficiency Bounds for a Class of Experimental Designs. ArXiv:2205.02726 [stat], URL http://arxiv.org/abs/2205.02726.

Bai, Y., Liu, J. and Tabord-Meehan, M. (2022). Inference for Matched Tuples and Fully Blocked Factorial Designs. ArXiv:2206.04157 [econ, math, stat], URL http://arxiv.org/abs/2206.04157.

Bai, Y., Romano, J. P. and Shaikh, A. M. (2021). Inference in Experiments with Matched Pairs*. *Journal of the American Statistical Association*, **0** 1–37. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01621459.2021.1883437, URL https://doi.org/10.1080/01621459.2021.1883437.

Belloni, A., Chen, D., Chernozhukov, V. and Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, **80** 2369–2429.

Belloni, A., Chernozhukov, V., Fernández-Val, I. and Hansen, C. (2017). Program evaluation with high-dimensional data. *Econometrica*, **85** 233–298.

Belloni, A., Chernozhukov, V. and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, **81** 608–650.

Bickel, P. J., Ritov, Y. and Tsybakov, A. B. (2009). Simultaneous analysis of Lasso and Dantzig selector. *The Annals of Statistics*, **37** 1705–1732. Publisher: Institute of Mathematical Statistics, URL https://projecteuclid.org/journals/annals-of-statistics/volume-37/issue-4/Simultaneous-analysis-of-Lasso-and-Dantzig-selector/10.1214/08-AOS620.full.

Bruhn, M. and McKenzie, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *American Economic Journal: Applied Economics*, **1** 200–232.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2014). Gaussian approximation of suprema of empirical processes. *The Annals of Statistics*, **42** 1564–1597.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *The Annals of Probability*, **45** 2309–2352. Publisher: Institute of Mathematical Statistics, URL https://projecteuclid.org/journals/annals-of-probability/volume-45/issue-4/Central-limit-theorems-and-bootstrap-in-high-dimensions/10.1214/16-AOP1113.full.

Chetverikov, D. and Sørensen, J. R.-V. (2022). Analytic and Bootstrap-after-Cross-Validation Methods for Selecting Penalty Parameters of High-Dimensional M-Estimators. Tech. Rep. arXiv:2104.04716, arXiv. ArXiv:2104.04716 [econ, math, stat] type: article, URL http://arxiv.org/abs/2104.04716.

Cohen, P. L. and Fogarty, C. B. (2020). No-harm calibration for generalized oaxaca-blinder estimators. *arXiv preprint arXiv:2012.09246*.

Donner, A. and Klar, N. (2000). *Design and analysis of cluster randomization trials in health research*, vol. 27. Arnold London.

Dudley, R. M. (1989). *Real Analysis and Probability*. Wadsworth and Brook/Cole.

Fogarty, C. B. (2018). Regression-assisted inference for the average treatment effect in paired experiments. *Biometrika*, **105** 994–1000.

Freedman, D. A. (2008). On regression adjustments to experimental data. *Advances in Applied Mathematics*, **40** 180–193. URL http://www.sciencedirect.com/science/article/pii/S019688580700005X.

Glennerster, R. and Takavarasha, K. (2013). *Running randomized evaluations: A practical guide*. Princeton University Press.

Groh, M. and McKenzie, D. (2016). Macroinsurance for microenterprises: A randomized experiment in post-revolution egypt. *Journal of Development Economics*, **118** 13–25.

Jiang, L., Liu, X., Phillips, P. C. and Zhang, Y. (2022). Bootstrap inference for quantile treatment effects in randomized experiments with matched pairs. *Review of Economics and Statistics*. Forthcoming.

Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces: Isoperimetry and Processes*. Classics in Mathematics, Springer-Verlag, Berlin Heidelberg. URL https://www.springer.com/gp/book/9783642202117.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. 3rd ed. Springer, New York.

Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *Annals of Applied Statistics*, **7** 295–318. Publisher: Institute of Mathematical Statistics, URL https://projecteuclid.org/euclid.aoas/1365527200.

Negi, A. and Wooldridge, J. M. (2021). Revisiting regression adjustment in experiments with heterogeneous treatment effects. *Econometric Reviews*, **40** 504–534. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/07474938.2020.1824732, URL https://doi.org/10.1080/07474938.2020.1824732.

Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1995). Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data. *Journal of the American Statistical Association*, **90** 106–121. Publisher: [American Statistical Association, Taylor & Francis, Ltd.], URL https://www.jstor.org/stable/2291134.

Rosenberger, W. F. and Lachin, J. M. (2015). *Randomization in clinical trials: Theory and Practice*. John Wiley & Sons.

Tsiatis, A. A., Davidian, M., Zhang, M. and Lu, X. (2008). Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: a principled yet flexible approach. *Statistics in Medicine*, **27** 4658–4677.

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer-Verlag, New York.

Yang, L. and Tsiatis, A. A. (2001). Efficiency Study of Estimators for a Treatment Effect in a Pretest–Posttest Trial. *The American Statistician*, **55** 314–321. Publisher: Taylor & Francis _eprint: https://doi.org/10.1198/000313001753272466, URL https://doi.org/10.1198/000313001753272466.