

# BOOTSTRAPPING WITH AI/ML-GENERATED LABELS

By

Timothy Christensen, Sílvia Gonçalves and Benoit Perron

April 2026

COWLES FOUNDATION DISCUSSION PAPER NO. 2523



COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY  
Box 208281  
New Haven, Connecticut 06520-8281

<http://cowles.yale.edu/>

# Bootstrapping with AI/ML-generated labels\*

Timothy Christensen<sup>†</sup>      Sílvia Gonçalves<sup>‡</sup>      Benoit Perron<sup>§</sup>

April 26, 2026

## Abstract

AI/ML methods are increasingly used in economics to generate binary variables (or labels) via classification algorithms. When these generated variables are included as covariates in regressions, even small misclassification errors can induce large biases in OLS estimators and invalidate standard inference. We study whether the bootstrap can correct this bias and deliver valid inference. We first show that a seemingly natural *fixed-label bootstrap*, which generates data using estimated labels but relies on a corrupted version in estimation, is generally invalid unless a strong independence condition between the latent true labels and other covariates holds. We then propose a *coupled-label bootstrap* that jointly resamples the true and imputed labels, and show it is valid without this condition. Two finite-sample adjustments further improve coverage: a variance correction for uncertainty in estimated misclassification rates and a Hessian rotation for near-singular designs. We illustrate the methods in simulations and apply them to investigate the relationship between wages and remote work status.

*Keywords:* bootstrap, bias correction, binary labels, misclassification, AI/ML-generated data, inference.

---

\*We thank Jaap Abbring, Stephen Hansen, Christoph Rothe, and participants of the 2025 Royal Economic Society meetings in Birmingham. This material is based upon work supported by the National Science Foundation under Award No. 2521471 (Christensen). Gonçalves acknowledges support from the Social Sciences and Humanities Research Council of Canada grant RGPIN-435-2023-0352.

<sup>†</sup>Department of Economics, Yale University.

<sup>‡</sup>Department of Economics, McGill University.

<sup>§</sup>Department of Economics, Université de Montréal, CIREQ, CIRANO.

# 1 Introduction

It is now standard practice in economics and across the social sciences to generate new variables and measurements using artificial intelligence (AI) and machine learning (ML) methods. A leading use case is the generation of binary variables, or *labels*, which are imputed using classification algorithms. To give just a few examples within economics, [Goldsmith-Pinkham and Shue \(2023\)](#) impute the gender of borrowers from borrower names, [Bursztyn, Chaney, Hassan, and Rao \(2024\)](#) impute the race of charitable donors from donor names, [Adams-Prassl, Waters, Balgova, and Qian \(2023\)](#) and [Hansen, Lambert, Bloom, Davis, Sadun, and Taska \(2026\)](#) classify the flexibility of work arrangements from job postings metadata, and [Dupas, Handlan, Modestino, et al. \(2026\)](#) impute speaker gender and tone-of-voice from audio recordings of economics seminars. These imputed variables are rarely of interest themselves. Rather, they are used as inputs to econometric models for downstream inference tasks.

While the easy generation of labels has opened the door for new research, it is not without significant methodological challenges. [Battaglia, Christensen, Hansen, and Sacher \(2025, BCHS hereafter\)](#) show theoretically and empirically that even high performance classifiers with error rates below 1% can generate substantial biases in parameter estimates due to misclassification error in the generated labels. Moreover, failure to correct for these biases leads to invalid inference on regression parameters. There is therefore a need for methods that correct bias and restore valid inference.

A central challenge in developing such methods is data constraints. A widely discussed approach for correcting bias requires large validation samples, in which both the imputed variables and ground-truth labels are observed alongside other variables in the model.<sup>1</sup> In many economic applications, however, such data are unavailable. Instead, researchers may only observe an external sample from which the accuracy of the classifier can be measured. For instance, [Bursztyn et al. \(2024\)](#) estimate the accuracy of their classifier using an external sample of North Carolina

---

<sup>1</sup>In economics, this approach goes back to [Bound and Krueger \(1991\)](#), [Carroll and Wand \(1991\)](#), [Sepanski and Carroll \(1993\)](#), [Bound, Brown, Duncan, and Rodgers \(1994\)](#), and [Lee and Sepanski \(1995\)](#), among others, and has connections with the literature on auxiliary data ([Chen, Hong, and Tamer, 2005](#); [Chen, Hong, and Tarozzi, 2008](#)). Recent papers proposing these methods within the context of AI/ML-generated data in economics include [Ludwig, Mullainathan, and Rambachan \(2025\)](#) and [Carlson and Dell \(2026\)](#). Similar approaches are taken in the recent statistics literature on “prediction-powered inference” ([Angelopoulos et al., 2023a,b](#)). See also [Fong and Tyler \(2021\)](#) and [Egami, Hinck, Stewart, and Wei \(2024\)](#), among others, for related work in political science.

voter registration data which contains self-reported ethnicity, but is missing data on charitable donations. Moreover, because collecting ground-truth labels is typically much more costly than generating labels with a classifier, the size  $m$  of this external sample is usually much smaller than the sample size  $n$  used for model estimation ( $m \ll n$ ). [BCHS](#) propose analytic bias corrections that account for both these data constraints.

In this paper, we study whether the bootstrap can correct bias and deliver valid inference. There is a long tradition of using the bootstrap to reduce bias and improve the coverage of confidence intervals ([Efron, 1981](#); [Hall, 1992](#)). In our setting, however, the problem is deceptively difficult. With binary labels, measurement error is necessarily “nonclassical” ([Aigner, 1973](#)): a misclassified 1 must be a 0, and vice versa. Consequently, common intuitions and methods developed in settings with “classical” measurement error do not easily carry over.

More formally, consider the regression model

$$\begin{aligned} Y_i &= X_i' \beta + u_i, \\ X_i &= g(\theta_i, Z_i), \end{aligned}$$

where  $\theta_i$  is a latent binary variable,  $Y_i$  and  $Z_i$  are observed, and the functional form of  $g$  is known. For example,  $g(\theta_i, Z_i) = (\theta_i, Z_i)'$  when  $\theta_i$  enters the regression additively, or  $g(\theta_i, Z_i) = (\theta_i Z_i', Z_i)'$  when it enters through interactions. In practice, researchers replace  $\theta_i$  with an estimate  $\hat{\theta}_i$ , regress  $Y_i$  on  $\hat{X}_i \equiv g(\hat{\theta}_i, Z_i)$ , and report confidence intervals and standard errors for  $\beta$  treating  $\hat{X}_i$  as though it is the truth. Misclassification ( $\hat{\theta}_i \neq \theta_i$  for some observations) induces measurement error in the  $\hat{X}_i$ , so regression estimators can be biased and standard inference can be invalid.

As in [BCHS](#), we assume the researcher observes a large sample  $(Y_i, Z_i, \hat{\theta}_i)_{i=1}^n$  and a much smaller external sample  $(\theta_i, \hat{\theta}_i)_{i=1}^m$  used to estimate classifier accuracy.<sup>2</sup> The question we address is: how should one bootstrap in this scenario?

Naively resampling  $(Y_i, \hat{X}_i)$ , or using equivalent residual-based approaches such as the wild bootstrap, does not work. These methods do not introduce any measurement error in the covariates in the bootstrap world, because the same covariates are used in data generation and parameter estimation. As a result, they fail to reproduce the bias of the OLS estimator and yield invalid inference.

---

<sup>2</sup>In fact, neither we nor [BCHS](#) need the full external sample, just estimates  $\hat{F}_+$  and  $\hat{F}_-$  of the false positive and negative rates of the classifier, and the sample size  $m$  used to estimate those rates.

A seemingly natural approach is a *fixed-label bootstrap*, a variant of the wild bootstrap that builds in measurement error. Outcomes are generated as  $Y_i^* = \hat{X}_i' \hat{\beta} + u_i^*$ , where  $\hat{\beta}$  is the OLS estimator from regression of  $Y_i$  on  $\hat{X}_i$ , and  $u_i^*$  is a bootstrap version of the OLS residual. Then,  $Y_i^*$  is regressed on  $\hat{X}_i^* \equiv g(\hat{\theta}_i^*, Z_i)$ , where  $\hat{\theta}_i^*$  is a corrupted version of  $\hat{\theta}_i$  constructed so that the false positive and negative rates for predicting  $\hat{\theta}_i$  with  $\hat{\theta}_i^*$  match those observed in the external sample. Analogues of this bootstrap have been used to correct bias in other settings with generated covariates, including factor-augmented regressions (Gonçalves and Perron, 2014; Gonçalves, Koh, and Perron, 2025) and panel data models (Gonçalves and Kaffo, 2015; Higgins and Jochmans, 2024; Li, Shen, and Zhou, 2024). However, we show that this bootstrap is generally inconsistent: it has the correct asymptotic variance but fails to replicate the asymptotic bias of the OLS estimator unless certain independence conditions hold between  $\theta_i$  and  $Z_i$ .

To address this problem, we introduce a novel *coupled-label bootstrap*. Outcomes are generated as  $Y_i^* = X_i^{*'} \hat{\beta} + u_i^*$ , where  $X_i^* = g(\theta_i^*, Z_i)$ ,  $\theta_i^*$  is a bootstrap version of the true latent  $\theta_i$ , and  $\hat{\beta}$  and  $u_i^*$  are defined above. Estimation then proceeds by regressing  $Y_i^*$  on  $\hat{X}_i^*$  as before. The key innovation is to generate the pairs  $(\theta_i^*, \hat{\theta}_i^*)$  jointly so that the misclassification rates for predicting  $\theta_i^*$  with  $\hat{\theta}_i^*$  match those in the external sample. We show that this bootstrap reproduces the correct asymptotic bias and variance of the OLS estimator, which justifies the use of confidence sets and hypothesis tests constructed via this bootstrap.

While our results establish asymptotic validity of the coupled-label bootstrap, two further modifications improve its finite-sample performance. The first is a variance correction that accounts for additional sampling error from estimating the misclassification rates. While BCHS develop analytic corrections, we show that a similar adjustment can be implemented via a simple modification of the bootstrap. The second is a rotation of the inverse Hessian, which is useful in near-singular designs where only a small fraction of  $\theta_i$  are 0 (or 1). In these settings, misclassification can introduce discrepancies between the bootstrap and estimated inverse Hessian that are asymptotically negligible but material in finite samples. Simulations calibrated to our empirical application show that these modifications substantially improve finite-sample performance. We therefore recommend the coupled-label bootstrap with both modifications.

The remainder of the paper is organized as follows. Section 2 provides high-level

conditions for bootstrap validity with generated covariates. Section 3 develops theory for the binary labels case, analyzing a general wild bootstrap and providing results for the fixed-label and coupled-label bootstraps. Section 4 introduces two modifications that improve finite-sample performance. Section 5 presents simulation evidence in a design calibrated to the application, and Section 6 revisits Hansen et al. (2026), who study the relationship between wages and remote work status. All proofs are presented in the appendix.

**Notation** Let  $\mathbb{P}^*$  denote the bootstrap probability measure conditional on the data, and  $\mathbb{E}^*$  the corresponding expectation. For bootstrap random vectors  $Y_n^*$ , we write  $Y_n^* \rightarrow_{p^*} 0$  or  $Y_n^* = o_{p^*}(1)$  if  $\mathbb{P}^*(\|Y_n^*\| > \epsilon) \rightarrow_p 0$  for all  $\epsilon > 0$ . Similarly,  $Y_n^* = O_{p^*}(1)$  if  $\limsup_{n \rightarrow \infty} \mathbb{P}^*(\|Y_n^*\| > M) \rightarrow_p 0$  as  $M \rightarrow \infty$ . Finally, we write  $Z_n^* \rightarrow_{d^*} Z$  to denote that  $\mathbb{P}^*(Z_n^* \leq z) - F(z) \rightarrow_p 0$  for all continuity points  $z$  of the CDF  $F$  of  $Z$ .

## 2 General Case

In this section, we consider the general case in which  $Y$  is regressed on a vector of generated covariates  $\hat{X}$ . We first recap results from BCHS, who show that regressing  $Y$  on  $\hat{X}$  induces a bias in the asymptotic distribution of the OLS estimator that invalidates standard inference. We then provide high-level conditions under which the bootstrap can match the asymptotic bias and variance of the OLS estimator, justifying the use of bootstrap percentile intervals to perform valid inference. The central question is then whether one can design a bootstrap satisfying these conditions in settings of interest. Section 3 addresses this question for the binary labels case.

### 2.1 Asymptotic Distribution of the OLS Estimator

We wish to perform inference on the parameter vector  $\beta$  in the regression model

$$Y_i = \beta' X_i + u_i,$$

where the true covariates  $X_i$  are latent and generated covariates  $\hat{X}_i$  are used in their place. We are interested in the large-sample properties of the OLS estimator

$$\hat{\beta} = \left( n^{-1} \sum_{i=1}^n \hat{X}_i \hat{X}_i' \right)^{-1} n^{-1} \sum_{i=1}^n \hat{X}_i Y_i$$

from a regression of  $Y_i$  on a vector of generated covariates  $\hat{X}_i$ .

The following regularity conditions are similar to those in [BCHS](#):

**Assumption 1.** (i)  $\frac{1}{n} \sum_{i=1}^n \|\hat{X}_i - X_i\|^2 \rightarrow_p 0$ ;

(ii)  $\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{X}_i (\hat{X}_i - X_i)' \rightarrow_p B$ ;

(iii)  $\frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{X}_i - X_i) u_i \rightarrow_p 0$ .

**Assumption 2.** (i)  $\frac{1}{n} \sum_{i=1}^n X_i X_i' \rightarrow_p Q \equiv \mathbb{E}[X_i X_i'] > 0$ , for  $\mathbb{E}[\|X_i\|^2] < \infty$ ;

(ii)  $\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i u_i \rightarrow_d N(0, \Sigma)$ , for  $\Sigma = \mathbb{E}[X_i X_i' u_i^2]$ .

Assumption 1 imposes conditions on the generated variable  $\hat{X}_i$  while Assumption 2 is standard and only concerns the latent variable  $X_i$ . Assumption 1 allows for both classical and nonclassical error in  $\hat{X}_i$ . The key condition is Assumption 1(ii). In the classical case, this condition requires the measurement error variance to be  $O(n^{-1/2})$ , similar to the small measurement error asymptotic frameworks of [Chesher \(1991\)](#) and [Evdokimov and Zeleneev \(2023\)](#). This scaling ensures that measurement error persists but does not dominate sampling error in the asymptotic distribution of  $\hat{\beta}$ . As a result, asymptotics deliver tractable and useful approximations to the finite-sample distribution of  $\hat{\beta}$ , where both sources of error play a role. Section 3 provides sufficient conditions for Assumption 1 for the binary labels case.

The next result follows by similar arguments to [BCHS](#) and so a proof is omitted:

**Theorem 1.** *Suppose that Assumptions 1 and 2 hold. Then as  $n \rightarrow \infty$ ,*

$$\sqrt{n}(\hat{\beta} - \beta) \rightarrow_d N(b, V),$$

where  $b = -Q^{-1}B\beta$ , and  $V = Q^{-1}\Sigma Q^{-1}$ .

As noted in [BCHS](#), Theorem 1 shows that if  $b \neq 0$ , then  $\hat{\beta}$  has a first-order asymptotic bias and the usual approach to inference is invalid.

## 2.2 Consistency of the Bootstrap

Suppose that we generate the bootstrap data as

$$Y_i^* = \hat{\beta}' X_i^* + u_i^*$$

where  $u_i^*$  is a bootstrap version of  $\hat{u}_i = Y_i - \hat{\beta}' \hat{X}_i$ . For now, we do not specify a particular bootstrap method used in generating  $u_i^*$ . Later we will consider a wild bootstrap for i.i.d. data. Similarly,  $X_i^*$  is a bootstrap analogue of the true latent  $X_i$ . A natural choice for  $X_i^*$  is  $\hat{X}_i$ , as in a standard fixed-regressor bootstrap. As we will see below, allowing for  $X_i^*$  to differ from  $\hat{X}_i$  turns out to be important for the binary labels application, so we allow for this possibility here.

Let  $\hat{X}_i^*$  denote a bootstrap analogue of  $\hat{X}_i$  which will be context-specific. The bootstrap analogue of  $\hat{\beta}$  is

$$\hat{\beta}^* = \left( n^{-1} \sum_{i=1}^n \hat{X}_i^* \hat{X}_i^{*'} \right)^{-1} n^{-1} \sum_{i=1}^n \hat{X}_i^* Y_i^*.$$

The following set of high-level conditions on  $(u_i^*, X_i^*, \hat{X}_i^*)$  ensure the bootstrap distribution of  $\sqrt{n}(\hat{\beta}^* - \hat{\beta})$  is consistent for the distribution of  $\sqrt{n}(\hat{\beta} - \beta)$  given in Theorem 1 as  $n \rightarrow \infty$ .

**Assumption B.1.** (i)  $\frac{1}{n} \sum_{i=1}^n \|\hat{X}_i^* - X_i^*\|^2 \rightarrow_{p^*} 0$  and  $\frac{1}{n} \sum_{i=1}^n \|X_i^* - \hat{X}_i\|^2 \rightarrow_{p^*} 0$ .

(ii)  $\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{X}_i^* (\hat{X}_i^* - X_i^*)' \rightarrow_{p^*} B$ .

(iii)  $\frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{X}_i^* - X_i^*) u_i^* \rightarrow_{p^*} 0$  and  $\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i^* - \hat{X}_i) u_i^* \rightarrow_{p^*} 0$ .

**Assumption B.2.**  $\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{X}_i u_i^* \rightarrow_{d^*} N(0, \Sigma)$ .

Assumption B.1 is the bootstrap analogue of Assumption 1. Part (i) is used to show that  $n^{-1} \sum_{i=1}^n \hat{X}_i^* \hat{X}_i^{*'} \rightarrow_{p^*} Q$ , given  $n^{-1} \sum_{i=1}^n \hat{X}_i \hat{X}_i' \rightarrow_p Q$  under Assumptions 1 and 2. Part (iii) and Assumption B.2 imply that  $n^{-1/2} \sum_{i=1}^n \hat{X}_i^* u_i^* \rightarrow_{d^*} N(0, \Sigma)$ . In particular, Assumption B.2 is the bootstrap analogue of Assumption 2(ii) and follows by the application of a bootstrap CLT. Assumption B.1 (ii) is the crucial assumption that ensures that the bootstrap mimics the bias term  $B$  in Assumption 1(ii).

**Theorem 2.** *Suppose that Assumptions 1 and 2 hold. If  $(u_i^*, X_i^*, \hat{X}_i^*)$  satisfy Assumptions B.1 and B.2, then as  $n \rightarrow \infty$ ,*

$$\sqrt{n}(\hat{\beta}^* - \hat{\beta}) \rightarrow_{d^*} N(b, V),$$

where  $b = -Q^{-1}B\beta$ , and  $V = Q^{-1}\Sigma Q^{-1}$ .

Theorem 2 shows that the bootstrap correctly replicates the limiting distribution of  $\hat{\beta}$ , including its bias and variance, thereby justifying the construction of bootstrap percentile confidence intervals for the elements of  $\beta$ . Specifically, a  $100(1 - \alpha)\%$  percentile interval for the  $j$ -th element  $\beta_j$  of  $\beta$  is given by

$$[\hat{\beta}_j - \hat{q}_{1-\alpha/2,j}, \hat{\beta}_j - \hat{q}_{\alpha/2,j}],$$

where  $\hat{q}_{\alpha,j}$  denotes the  $\alpha$ -quantile of the bootstrap distribution of  $\hat{\beta}_j^* - \hat{\beta}_j$ . Unlike asymptotic theory-based intervals, this interval requires neither an explicit bias correction nor a variance estimator.

### 3 Binary Labels Case

We now turn to the binary-label case. We begin by describing the setup and specializing Theorem 1 to this setting. We then consider two bootstrap methods, both of which are instances of the wild bootstrap. The first is a seemingly natural *fixed-label bootstrap*, analogous to the standard fixed-regressor bootstrap, which is generally inconsistent unless strong independence conditions hold between  $\theta_i$  and  $Z_i$ . We next present a *coupled-label bootstrap* and show that it is consistent in settings where the fixed-label bootstrap fails. We therefore recommend the coupled-label bootstrap, together with the finite-sample modifications described in the next section.

#### 3.1 Setup

In this setting,

$$X_i = g(\theta_i, Z_i),$$

where  $g : \{0, 1\} \times \mathbb{R}^{d_z} \rightarrow \mathbb{R}^k$  is known,  $\theta_i$  is a latent binary random variable, and  $Z_i \in \mathbb{R}^{d_z}$  is an observed random vector. Since  $\theta_i$  is latent, it is common practice

to generate an estimate  $\hat{\theta}_i$  of  $\theta_i$  using a classification algorithm, then regress  $Y_i$  on  $\hat{X}_i = g(\hat{\theta}_i, Z_i)$ . For ease of exposition we will focus on the case of scalar  $\theta_i$ , though it is straightforward to extend our analysis to the vector case.

**Example 1.** *This framework includes models where  $\theta_i$  is interacted with a subset of the  $Z_i$  variables, such as*

$$Y_i = \beta_1' \theta_i Z_{1i} + \beta_2' Z_{2i} + u_i, \quad (1)$$

*to capture, e.g., effects that vary by membership of a race or gender group. In this case,  $g(\theta_i, Z_i) = (\theta_i Z_{1i}', Z_{2i}')' \in \mathbb{R}^k$ .*

**Example 2.** *Example 1 includes as a special case the model studied in [BCHS](#), in which*

$$Y_i = \beta_1 \theta_i + \beta_2' Z_{2i} + u_i, \quad (2)$$

*where  $Z_{1i} = 1$  and  $Z_{2i} = Z_i$ . For this model,  $g(\theta_i, Z_i) = (\theta_i, Z_i')' \in \mathbb{R}^k$ .*

To maintain flexibility, we will not take a stand on the algorithm with which  $\hat{\theta}_i$  is generated. Instead, we shall assume that  $(Y_i, Z_i, \hat{\theta}_i, \theta_i)_{i=1}^n$  are drawn i.i.d. from a distribution  $P_n$ . Here we use the notation  $P_n$  because our asymptotic framework introduced below allows the distribution of the data to vary with  $n$ , so that both misclassification and sampling error feature in the asymptotic distribution of the OLS estimator. The econometrician observes  $(Y_i, Z_i, \hat{\theta}_i)_{i=1}^n$  and, for the purposes of bias correction, has access to an external sample  $(\theta_i, \hat{\theta}_i)_{i=1}^m$ , which can be used to assess the accuracy of the classifier. As in [BCHS](#), we allow the external sample size  $m$  to be of much smaller magnitude than  $n$ . This reflects common empirical settings with large amounts of generated data and small amounts of ground-truth observations.

### 3.2 Asymptotic Bias of the OLS Estimator

We now provide primitive conditions for Assumption 1 for the binary labels case. To this end we impose some structure on  $\hat{\theta}_i$  and  $\theta_i$ . We follow [BCHS](#) and adopt an asymptotic framework where the (unconditional) false-positive rate  $\mathbb{E}[\hat{\theta}_i(1 - \theta_i)]$  and (unconditional) false-negative rate  $\mathbb{E}[\theta_i(1 - \hat{\theta}_i)]$  both tend to zero as  $n \rightarrow \infty$ , so that both sampling error in the regression and measurement error in the  $\hat{\theta}_i$  remain relevant in large samples. This ensures the asymptotic distribution mimics the finite-sample setting where both play roles.

To introduce the assumptions, note that we can write

$$\hat{X}_i - X_i = \hat{\theta}_i(1 - \theta_i)(g(1, Z_i) - g(0, Z_i)) + \theta_i(1 - \hat{\theta}_i)(g(0, Z_i) - g(1, Z_i)),$$

from which it follows that

$$\hat{X}_i(\hat{X}_i - X_i)' = \hat{\theta}_i(1 - \theta_i)D_{+i} + \theta_i(1 - \hat{\theta}_i)D_{-i},$$

where

$$\begin{aligned} D_{+i} &= g(1, Z_i)(g(1, Z_i) - g(0, Z_i))', \\ D_{-i} &= g(0, Z_i)(g(0, Z_i) - g(1, Z_i))'. \end{aligned}$$

Let  $G_i = \max\{\|g(0, Z_i)\|, \|g(1, Z_i)\|\}$  and let  $\delta > 0$ .

- Assumption 3.** (i)  $\mathbb{E}[X_i u_i] = 0$  and  $\mathbb{E}[|u_i|^{4+\delta}] < \infty$ ;
- (ii)  $\mathbb{E}[X_i X_i'] > 0$  and  $\mathbb{E}[G_i^{4+\delta}] < \infty$ ;
- (iii)  $\mathbb{E}[(\hat{X}_i - X_i)u_i] = o(n^{-1/2})$  as  $n \rightarrow \infty$ ;
- (iv)  $\sqrt{n}\mathbb{E}[\hat{\theta}_i(1 - \theta_i)] \rightarrow \kappa_+$  and  $\sqrt{n}\mathbb{E}[\theta_i(1 - \hat{\theta}_i)] \rightarrow \kappa_-$  as  $n \rightarrow \infty$ , for  $0 \leq \kappa_+, \kappa_- < \infty$ ;
- (v)  $\sqrt{n}(\mathbb{E}[\hat{\theta}_i(1 - \theta_i)D_{+i} + \theta_i(1 - \hat{\theta}_i)D_{-i}] - \mathbb{E}[\hat{\theta}_i(1 - \theta_i)]\mathbb{E}[D_{+i}] - \mathbb{E}[\theta_i(1 - \hat{\theta}_i)]\mathbb{E}[D_{-i}]) \rightarrow 0$  as  $n \rightarrow \infty$ .

Assumption 3(i)–(ii) are standard regularity conditions. Assumption 3(iii) allows for some correlation between the measurement error in  $\hat{X}_i$  and  $u_i$ , provided it vanishes at a sufficiently fast rate. For instance, suppose  $\theta_i$  and  $\hat{\theta}_i$  are functions of  $Z_i$  and some auxiliary information  $\mathcal{U}_i$ . Then by iterated expectations,

$$\begin{aligned} \mathbb{E}[(\hat{X}_i - X_i)u_i] &= \mathbb{E}[\hat{\theta}_i(1 - \theta_i)(g(1, Z_i) - g(0, Z_i))\mathbb{E}[u_i|Z_i, \mathcal{U}_i]] \\ &\quad + \mathbb{E}[\theta_i(1 - \hat{\theta}_i)(g(0, Z_i) - g(1, Z_i))\mathbb{E}[u_i|Z_i, \mathcal{U}_i]], \end{aligned}$$

so it suffices that  $\mathbb{E}[u_i|Z_i, \mathcal{U}_i] = 0$  for Assumption 3(iii) to be satisfied.

Assumption 3(iv) formalizes the asymptotic framework described above. This condition requires the false positive and false negative rates to vanish at the same order as sampling uncertainty in the regression (i.e.,  $O(n^{-1/2})$ ). It provides a more robust starting point for empirical work than the standard approach of simply regressing  $Y_i$  on  $\hat{X}_i$  and performing OLS inference, which implicitly assumes  $\kappa_+$  and  $\kappa_-$  are zero.

Finally, Assumption 3(v) limits the dependence between  $g(0, Z_i)$  and  $g(1, Z_i)$  and the classification errors in  $\hat{\theta}_i$  (but not the dependence between  $Z_i$  and  $\theta_i$ ). This condition relates both to the functional form of  $g$  and the statistical nature of the classification errors. This condition allows the conditional probabilities of misclassification  $\mathbb{E}[\hat{\theta}_i(1 - \theta_i)|Z_i]$  and  $\mathbb{E}[\theta_i(1 - \hat{\theta}_i)|Z_i]$  to depend on  $Z_i$ , provided the misclassification errors are asymptotically uncorrelated with  $D_{+i}$  and  $D_{-i}$ . As this assumption is important, before proceeding we discuss some sufficient conditions for it.

**Example 2 (continued).** Consider (2) as in BCHS. Then

$$D_{+i} = \begin{pmatrix} 1 & 0 \\ Z_i & 0 \end{pmatrix}, \quad D_{-i} = \begin{pmatrix} 0 & 0 \\ -Z_i & 0 \end{pmatrix},$$

so a sufficient condition for Assumption 3(v) is

$$\begin{aligned} \sqrt{n}\mathbb{E}[\theta_i - \hat{\theta}_i] &\rightarrow 0, \\ \sqrt{n}\mathbb{E}[(\theta_i - \hat{\theta}_i)Z_i] &\rightarrow 0. \end{aligned} \tag{3}$$

Suppose that  $\hat{\theta}_i$  is generated by a logistic classifier with (potentially many) variables  $W_{i,n}$ , including  $Z_i$  and a constant. That is,

$$\mathbb{P}(\hat{\theta}_i = 1|W_{i,n}) = \frac{e^{W'_{i,n}\gamma_n}}{1 + e^{W'_{i,n}\gamma_n}}. \tag{4}$$

It is important to note that we are not assuming the logistic model is the DGP for  $\theta_i$ , just that it generates  $\hat{\theta}_i$ . Suppose the classifier has been pre-trained on a large external data set of  $(\theta_i, W_{i,n})$ . The parameters  $\gamma_n$  of the classifier maximize the population likelihood

$$L(\gamma) = \mathbb{E}[\theta_i W'_{i,n} \gamma - \log(1 + e^{W'_{i,n} \gamma})],$$

and therefore satisfy the first-order condition

$$0 = \mathbb{E} \left[ \left( \theta_i - \frac{e^{W'_{i,n} \gamma_n}}{1 + e^{W'_{i,n} \gamma_n}} \right) W_{i,n} \right].$$

Hence, by (4) and iterated expectations, we have

$$\mathbb{E}[(\theta_i - \hat{\theta}_i)W_{i,n}] = \mathbb{E}[(\theta_i - \mathbb{E}[\hat{\theta}_i|W_{i,n}])W_{i,n}] = 0,$$

from which it follows that (3) holds. Note that the first condition in (3) together with  $\sqrt{n}\mathbb{E}[\hat{\theta}_i(1-\theta_i)] \rightarrow \kappa_+$  implies that  $\sqrt{n}\mathbb{E}[\theta_i(1-\hat{\theta}_i)] \rightarrow \kappa_+$  also holds, and thus  $\kappa_+ = \kappa_-$ .

**Lemma 1.** *Suppose that Assumption 3 holds. Then Assumptions 1 and 2 hold with*

$$B = \kappa_+ \mathbb{E}[D_{+i}] + \kappa_- \mathbb{E}[D_{-i}].$$

Before proceeding, we characterize the bias term  $B$  in the context of the two running examples.

**Example 1 (continued).** *Consider the interactions model (1). Then*

$$B = \begin{pmatrix} \kappa_+ \mathbb{E}[Z_{1i} Z'_{1i}] & 0 \\ (\kappa_+ - \kappa_-) \mathbb{E}[Z_{2i} Z'_{1i}] & 0 \end{pmatrix}.$$

**Example 2 (continued).** *Consider the additive model (2). Then*

$$B = \begin{pmatrix} \kappa_+ & 0 \\ (\kappa_+ - \kappa_-) \mathbb{E}[Z_i] & 0 \end{pmatrix}.$$

If  $\kappa_+ = \kappa_-$ , then

$$B = \begin{pmatrix} \kappa_+ & 0 \\ 0 & 0 \end{pmatrix},$$

as in [BCHS](#).

### 3.3 Wild Bootstrap with Binary Labels

Both the fixed-label and coupled-label bootstraps are versions of the following wild bootstrap with binary labels, but differ in how  $(\theta_i^*, \hat{\theta}_i^*)$  are drawn:

$$\begin{aligned} Y_i^* &= \hat{\beta}' X_i^* + u_i^*, \\ X_i^* &= g(\theta_i^*, Z_i), \\ \hat{X}_i^* &= g(\hat{\theta}_i^*, Z_i), \\ u_i^* &= \hat{u}_i \eta_i, \end{aligned} \tag{5}$$

where  $\eta_i$  are i.i.d.  $(0, 1)$  and generated independently of the bootstrap labels  $(\theta_i^*, \hat{\theta}_i^*)$ , conditional on the original sample. To understand when these methods are valid, we

use the following set of conditions on  $(\theta_i^*, \hat{\theta}_i^*)$ .

**Assumption L.** (i)  $\frac{1}{n} \sum_{i=1}^n \mathbb{E}^*[\mathbb{I}[\theta_i^* \neq \hat{\theta}_i^*]] \rightarrow_p 0$ .

(ii)  $\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{E}^*[\hat{\theta}_i^*(1 - \theta_i^*)] \rightarrow_p \kappa_+$ ,  $\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{E}^*[\theta_i^*(1 - \hat{\theta}_i^*)] \rightarrow_p \kappa_-$ .

(iii)  $\text{Var}^*\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\theta}_i^*(1 - \theta_i^*)\right) \rightarrow_p 0$  and  $\text{Var}^*\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \theta_i^*(1 - \hat{\theta}_i^*)\right) \rightarrow_p 0$ .

(iv)  $\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\theta}_i^*(1 - \theta_i^*)(D_{+i} - \mathbb{E}[D_{+i}]) + \theta_i^*(1 - \hat{\theta}_i^*)(D_{-i} - \mathbb{E}[D_{-i}]) \rightarrow_{p^*} 0$ .

Assumption **L(i)** requires the bootstrap “true” labels  $\theta_i^*$  to match the estimated labels  $\hat{\theta}_i$  on average with probability converging to one. This condition is automatically satisfied if  $\theta_i^* = \hat{\theta}_i$ . Assumption **L(ii)** is the bootstrap analogue of Assumption **3(iv)**. It requires the average bootstrap false positive and negative rates, when scaled by  $\sqrt{n}$ , to converge in probability to  $\kappa_+$  and  $\kappa_-$ . Together with Assumptions **L(iii)** and **L(iv)**, this condition ensures that the bootstrap reproduces the asymptotic bias in the OLS estimator. Assumption **L(iv)** can be viewed as the bootstrap analogue of Assumption **3(v)**.

The following lemma shows that the high-level Assumptions **B.1** and **B.2** hold for the wild bootstrap with binary labels provided the method for generating  $(\theta_i^*, \hat{\theta}_i^*)$  satisfies Assumption **L**.

**Lemma 2.** *Suppose Assumption **3** holds. Let  $u_i^* = \hat{u}_i \eta_i$  where  $\eta_i$  are i.i.d.  $(0, 1)$  independently of  $(\theta_i^*, \hat{\theta}_i^*)$  such that  $\mathbb{E}^*[|\eta_i|^{2+\delta}] < \infty$  for  $\delta > 0$ . If, in addition,  $(\theta_i^*, \hat{\theta}_i^*)$  satisfy Assumption **L**, then Assumptions **B.1** and **B.2** hold.*

## 3.4 Specific Bootstrap Methods

Here we study two methods for generating  $(\theta_i^*, \hat{\theta}_i^*)$ . The first is a standard fixed-regressor bootstrap which sets  $\theta_i^* = \hat{\theta}_i$ . Its validity requires  $\mathbb{E}[D_{+i} | \theta_i = 0] = \mathbb{E}[D_{+i}]$  and  $\mathbb{E}[D_{-i} | \theta_i = 1] = \mathbb{E}[D_{-i}]$ , which holds if  $D_{+i}$  and  $D_{-i}$  are mean independent of  $\theta_i$ , but not necessarily otherwise. We therefore introduce a coupled-label bootstrap that jointly resamples  $(\theta_i^*, \hat{\theta}_i^*)$  independently of the original sample to mimic the bias  $B$  without this strong independence assumption.

### 3.4.1 Fixed-Label Bootstrap

This method sets  $\theta_i^* = \hat{\theta}_i$  in the bootstrap DGP for  $Y_i^*$ , treating the bootstrap latent labels as fixed. The bootstrap estimated labels  $\hat{\theta}_i^*$  are generated as i.i.d. Bernoulli

random variables, conditional on  $\hat{\theta}_i$ :

$$\mathbb{P}^*(\hat{\theta}_i^* = \theta | \hat{\theta}_i = 1) = \begin{cases} \hat{F}_- / \hat{\pi} & \text{if } \theta = 0, \\ 1 - \hat{F}_- / \hat{\pi} & \text{if } \theta = 1, \end{cases}$$

and

$$\mathbb{P}^*(\hat{\theta}_i^* = \theta | \hat{\theta}_i = 0) = \begin{cases} \hat{F}_+ / (1 - \hat{\pi}) & \text{if } \theta = 1, \\ 1 - \hat{F}_+ / (1 - \hat{\pi}) & \text{if } \theta = 0. \end{cases}$$

Here,  $\hat{F}_+$  and  $\hat{F}_-$  are such that  $\sqrt{n}\hat{F}_+ \rightarrow_p \kappa_+$  and  $\sqrt{n}\hat{F}_- \rightarrow_p \kappa_-$ , respectively, and  $\hat{\pi} = n^{-1} \sum_{i=1}^n \hat{\theta}_i$ . To this end, we follow [BCHS](#) and rely on an external sample  $(\theta_i, \hat{\theta}_i)_{i=1}^m$  used to assess the accuracy of the classifier. [BCHS](#) show that

$$\hat{\kappa}_+ = \sqrt{n}\hat{F}_+, \quad \text{with} \quad \hat{F}_+ = \frac{1}{m} \sum_{i=1}^m \hat{\theta}_i(1 - \theta_i)$$

is consistent as  $n, m \rightarrow \infty$  with  $n/m^2 \rightarrow 0$  under Assumption [3\(iv\)](#). An analogous result holds for

$$\hat{\kappa}_- = \sqrt{n}\hat{F}_-, \quad \text{with} \quad \hat{F}_- = \frac{1}{m} \sum_{i=1}^m \theta_i(1 - \hat{\theta}_i).$$

Importantly, these results allow for the external sample size  $m$  to be much smaller than the original sample size  $n$  used in the regression.

For this method, Assumption [L\(i\)](#) is automatically satisfied since  $\theta_i^* = \hat{\theta}_i$ . One can also show that Assumptions [L\(ii\)](#)-[L\(iii\)](#) are satisfied provided  $\sqrt{n}\hat{F}_+ \rightarrow_p \kappa_+$  and  $\sqrt{n}\hat{F}_- \rightarrow_p \kappa_-$ . However, Assumption [L\(iv\)](#) fails whenever  $(D_{+i}, D_{-i})$  are not mean independent of  $\hat{\theta}_i$ . This implies that the fixed-label bootstrap only replicates  $B$  if we assume that this strong independence condition holds. To see this, note that the bootstrap bias is determined by (the probability limit of) the bootstrap expectation of

$$\hat{B}^* \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{\theta}_i^*(1 - \theta_i^*)D_{+i} + \theta_i^*(1 - \hat{\theta}_i^*)D_{-i}), \quad \text{where } \theta_i^* = \hat{\theta}_i.$$

Fixing the ‘‘latent’’ bootstrap labels  $\theta_i^*$  at  $\hat{\theta}_i$  implies that the bootstrap expectation

of  $\hat{B}^*$  is equal to

$$\begin{aligned}\hat{B} \equiv \mathbb{E}^*[\hat{B}^*] &= \hat{\kappa}_+ \frac{1}{1-\hat{\pi}} \frac{1}{n} \sum_{i=1}^n (1-\hat{\theta}_i) D_{+i} + \hat{\kappa}_- \frac{1}{\hat{\pi}} \frac{1}{n} \sum_{i=1}^n \hat{\theta}_i D_{-i} \\ &\rightarrow_p \kappa_+ \mathbb{E}[D_{+i} | \theta_i = 0] + \kappa_- \mathbb{E}[D_{-i} | \theta_i = 1],\end{aligned}$$

whereas

$$B = \kappa_+ \mathbb{E}[D_{+i}] + \kappa_- \mathbb{E}[D_{-i}].$$

Hence,  $\hat{B}$  does not consistently estimate the bias  $B$  in general, unless  $(D_{+i}, D_{-i})$  are mean independent of  $\theta_i$ . We summarize this result next.

**Corollary 1.** *Under Assumption 3, the fixed-label bootstrap satisfies the conditions of Theorem 2 if  $\sqrt{n}\hat{F}_+ \rightarrow_p \kappa_+$  and  $\sqrt{n}\hat{F}_- \rightarrow_p \kappa_-$ , and  $(D_{+i}, D_{-i})$  are mean independent of  $\theta_i$ .*

**Example 2 (continued).** *One instance in which this bootstrap is valid is the special case of model (2) where  $Z_i$  contains just a constant for the intercept. Then,  $D_{+i}$  and  $D_{-i}$  are non-stochastic, and so  $\hat{B} \rightarrow_p B$ .*

### 3.4.2 Coupled-Label Bootstrap

The coupled-label bootstrap generates the pair  $(\theta_i^*, \hat{\theta}_i^*)$  jointly for each observation  $i = 1, \dots, n$ , independently across  $i$  and independently of  $u_i^* = \hat{u}_i \eta_i$ , according to the following categorical distributions:

$$\mathbb{P}^*(\theta_i^* = \theta_1, \hat{\theta}_i^* = \theta_2 \mid \hat{\theta}_i = 1) = \begin{cases} 1 - \hat{F}_+ - \frac{\hat{F}_-}{\hat{\pi}} & \text{if } \theta_1 = 1, \theta_2 = 1, \\ \hat{F}_- & \text{if } \theta_1 = 1, \theta_2 = 0, \\ \hat{F}_+ & \text{if } \theta_1 = 0, \theta_2 = 1, \\ \hat{F}_- \frac{1-\hat{\pi}}{\hat{\pi}} & \text{if } \theta_1 = 0, \theta_2 = 0, \end{cases}$$

and

$$\mathbb{P}^*(\theta_i^* = \theta_1, \hat{\theta}_i^* = \theta_2 \mid \hat{\theta}_i = 0) = \begin{cases} \hat{F}_+ \frac{\hat{\pi}}{1-\hat{\pi}} & \text{if } \theta_1 = 1, \theta_2 = 1, \\ \hat{F}_- & \text{if } \theta_1 = 1, \theta_2 = 0, \\ \hat{F}_+ & \text{if } \theta_1 = 0, \theta_2 = 1, \\ 1 - \hat{F}_+ \frac{1}{1-\hat{\pi}} - \hat{F}_- & \text{if } \theta_1 = 0, \theta_2 = 0. \end{cases}$$

The key feature of this method is that the probability of discordant pairs  $(\theta_i^*, \hat{\theta}_i^*) = (1, 0)$  and  $(0, 1)$  equals  $\hat{F}_-$  and  $\hat{F}_+$ , respectively, for every observation  $i$ , regardless of  $\hat{\theta}_i$ . As a result, the bootstrap false positive and negative rates  $\mathbb{E}^*[\hat{\theta}_i^*(1 - \theta_i^*)] = \hat{F}_+$  and  $\mathbb{E}^*[\theta_i^*(1 - \hat{\theta}_i^*)] = \hat{F}_-$  are constant across  $i$ . This ensures that the bootstrap correctly replicates the asymptotic bias  $B$  without requiring independence between  $\theta_i$  and  $Z_i$ , since the bootstrap misclassification indicators are decoupled from  $(D_{+i}, D_{-i})$ .

The conditional probabilities assigned to the concordant pairs  $(\theta_i^*, \hat{\theta}_i^*) = (1, 1)$  and  $(0, 0)$  ensure that the bootstrap Hessian  $\hat{Q}^*$  converges to the empirical Hessian  $\hat{Q}$  conditional on the data. Since  $\hat{Q}^*$  depends only on  $\hat{\theta}_i^*$  interacted with  $g(1, Z_i)g(1, Z_i)'$  and  $g(0, Z_i)g(0, Z_i)'$ , its bootstrap expectation is determined by  $\mathbb{E}^*[\hat{\theta}_i^*]$ . As the discordant pair probabilities are set equal to  $\hat{F}_+$  and  $\hat{F}_-$ , the key requirement is that  $\mathbb{P}^*(\theta_i^* = 1, \hat{\theta}_i^* = 1 | \hat{\theta}_i = 1) \rightarrow_p 1$  and  $\mathbb{P}^*(\theta_i^* = 1, \hat{\theta}_i^* = 1 | \hat{\theta}_i = 0) \rightarrow_p 0$ , both of which are satisfied by the proposed construction. Under this distribution,  $\mathbb{E}^*[\hat{\theta}_i^*] = \hat{\theta}_i(1 - \hat{F}_-/\hat{\pi}) + (1 - \hat{\theta}_i)\hat{F}_+/(1 - \hat{\pi})$ , which ensures that the coupled-label bootstrap Hessian matches the fixed-label bootstrap Hessian in expectation. Other choices are nevertheless possible.

The following result formally establishes the validity of this bootstrap:

**Theorem 3.** *Suppose that Assumption 3 holds. Let  $u_i^* = \hat{u}_i\eta_i$  where  $\eta_i$  are i.i.d.  $(0, 1)$  independently of  $(\theta_i^*, \hat{\theta}_i^*)$  such that  $\mathbb{E}^*[|\eta_i|^{2+\delta}] < \infty$  for  $\delta > 0$ . If, in addition,  $\sqrt{n}\hat{F}_+ \rightarrow_p \kappa_+$  and  $\sqrt{n}\hat{F}_- \rightarrow_p \kappa_-$ , then the coupled-label bootstrap satisfies the conditions of Theorem 2 and, consequently,*

$$\sqrt{n}(\hat{\beta}^* - \hat{\beta}) \rightarrow_{d^*} N(b, V),$$

as  $n \rightarrow \infty$ , with  $b = -Q^{-1}B\beta$  and  $V = Q^{-1}\Sigma Q^{-1}$ .

Theorem 3 shows that the coupled-label bootstrap correctly replicates the bias and variance of the asymptotic distribution of the OLS estimator in Theorem 1, thus justifying the use of percentile intervals for valid inference. However, as our simulations show, finite-sample coverage can deviate from the nominal level. We therefore recommend using the coupled-label bootstrap with the two finite-sample modifications described in the next section.

## 4 Finite-Sample Modifications

The asymptotic validity of the bootstrap percentile interval requires the bootstrap to correctly mimic the asymptotic distribution of  $\sqrt{n}(\hat{\beta} - \beta)$ , including its mean and variance. From the proofs of Lemma A.3 and Theorem 2,

$$\sqrt{n}(\hat{\beta}^* - \hat{\beta}) = Z^* + \hat{b}^*, \quad (6)$$

where  $Z^* \rightarrow_{d^*} N(0, V)$  and

$$\hat{b}^* = -\hat{Q}^{*-1} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\theta}_i^* (1 - \theta_i^*) D_{+i} + \theta_i^* (1 - \hat{\theta}_i^*) D_{-i} \right) \hat{\beta} \equiv -\hat{Q}^{*-1} \hat{B}^* \hat{\beta},$$

where  $\hat{Q}^* = n^{-1} \sum_{i=1}^n \hat{X}_i^* \hat{X}_i^{*'}$  is the bootstrap Hessian matrix. To show that  $\hat{b}^* \rightarrow_{p^*} b$ , we add and subtract appropriately to obtain

$$\hat{b}^* = -\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\theta}_i^* (1 - \theta_i^*) \Gamma_+ \beta + \theta_i^* (1 - \hat{\theta}_i^*) \Gamma_- \beta + o_{p^*}(1) \equiv b^* + o_{p^*}(1), \quad (7)$$

where  $\Gamma_+ = Q^{-1} \mathbb{E}[D_{+i}]$  and  $\Gamma_- = Q^{-1} \mathbb{E}[D_{-i}]$ . While the coupled-label bootstrap bias  $b^*$  is consistent for  $b$ , finite-sample distortions may arise. In particular, as noted in BCHS, estimating  $F_+$  and  $F_-$  introduces additional noise that can inflate the variance of  $b^*$ , requiring a variance correction to improve coverage of bias-corrected confidence intervals when  $m/n$  is small. The same issue arises here, but a similar correction can easily be incorporated into the coupled-label bootstrap. A second source of distortion arises from the  $o_{p^*}(1)$  term representing the difference between  $\hat{b}^*$  and  $b^*$ . This term depends on the difference between  $\hat{Q}^{*-1}$  and  $\hat{Q}^{-1}$ , which can be non-negligible in finite samples, especially when the proportion of  $\hat{\theta}_i$  equal to 1 or 0 is small. In such cases, the bootstrap mean of  $\hat{b}^*$  is different from that of  $b^*$ , introducing a location bias in the bootstrap. This motivates rotating  $\hat{\beta}^* - \hat{\beta}$  by  $\hat{R}^* = \hat{Q}^{-1} \hat{Q}^*$ .

We discuss these two modifications next. We first describe the variance correction for the coupled-label bootstrap, establish its validity, and show it is able to replicate the two leading higher-order variance terms. We then describe the rotation that resolves the mismatch between the bootstrap Hessian and its sample analogue.

## 4.1 Coupled-Label Bootstrap with Variance Correction

To generate each bootstrap sample, first draw  $V_+^* \sim \text{Binomial}(m, \hat{F}_+)$  and  $V_-^* \sim \text{Binomial}(m, \hat{F}_-)$  and set  $\hat{F}_+^* = V_+^*/m$  and  $\hat{F}_-^* = V_-^*/m$ . Conditional on  $\hat{F}_+^*$  and  $\hat{F}_-^*$ , the pairs  $(\theta_i^*, \hat{\theta}_i^*)$  are generated jointly for each observation  $i = 1, \dots, n$ , independently across  $i$  and independently of  $u_i^* = \hat{u}_i \eta_i$ , according to the following distributions:

$$\mathbb{P}^*(\theta_i^* = \theta_1, \hat{\theta}_i^* = \theta_2 \mid \hat{\theta}_i = 1, \hat{F}_+^*, \hat{F}_-^*) = \begin{cases} 1 - \hat{F}_+^* - \frac{\hat{F}_-^*}{\hat{\pi}} & \text{if } \theta_1 = 1, \theta_2 = 1, \\ \hat{F}_-^* & \text{if } \theta_1 = 1, \theta_2 = 0, \\ \hat{F}_+^* & \text{if } \theta_1 = 0, \theta_2 = 1, \\ \hat{F}_-^* \frac{1-\hat{\pi}}{\hat{\pi}} & \text{if } \theta_1 = 0, \theta_2 = 0, \end{cases} \quad (8)$$

and

$$\mathbb{P}^*(\theta_i^* = \theta_1, \hat{\theta}_i^* = \theta_2 \mid \hat{\theta}_i = 0, \hat{F}_+^*, \hat{F}_-^*) = \begin{cases} \hat{F}_+^* \frac{\hat{\pi}}{1-\hat{\pi}} & \text{if } \theta_1 = 1, \theta_2 = 1, \\ \hat{F}_-^* & \text{if } \theta_1 = 1, \theta_2 = 0, \\ \hat{F}_+^* & \text{if } \theta_1 = 0, \theta_2 = 1, \\ 1 - \hat{F}_+^* \frac{1}{1-\hat{\pi}} - \hat{F}_-^* & \text{if } \theta_1 = 0, \theta_2 = 0. \end{cases} \quad (9)$$

Note that the  $\hat{F}_+^*$  and  $\hat{F}_-^*$  are generated only once per bootstrap sample and are held fixed when generating the  $n$  pairs  $(\theta_i^*, \hat{\theta}_i^*)$  for  $i = 1, \dots, n$ .

The following result formally establishes validity of this bootstrap:

**Theorem 4.** *Suppose that Assumption 3 holds. Let  $u_i^* = \hat{u}_i \eta_i$  where  $\eta_i$  are i.i.d.  $(0, 1)$  independently of  $(\theta_i^*, \hat{\theta}_i^*)$  such that  $\mathbb{E}^*[|\eta_i|^{2+\delta}] < \infty$  for  $\delta > 0$ . If, in addition,  $\sqrt{n}\hat{F}_+ \rightarrow_p \kappa_+$  and  $\sqrt{n}\hat{F}_- \rightarrow_p \kappa_-$  and  $n/m^2 \rightarrow 0$ , then the coupled-label bootstrap with variance correction satisfies the conditions of Theorem 2 and, consequently,*

$$\sqrt{n}(\hat{\beta}^* - \hat{\beta}) \rightarrow_{d^*} N(b, V),$$

as  $n \rightarrow \infty$ , with  $b = -Q^{-1}B\beta$  and  $V = Q^{-1}\Sigma Q^{-1}$ .

Theorem 4 shows that this bootstrap is able to reproduce the asymptotic bias and variance of the OLS estimator. We now show it is able to match key higher-order variance terms as well.

## 4.2 Variance Analysis

### 4.2.1 Finite-Sample Correction from BCHS

We first consider the bias-corrected estimator proposed in BCHS. That estimator can be written as

$$\hat{\beta}^{bc} = (I + \hat{\Gamma}_+ \hat{F}_+ + \hat{\Gamma}_- \hat{F}_-) \hat{\beta},$$

with  $\hat{Q} = \frac{1}{n} \sum_{i=1}^n \hat{X}_i \hat{X}_i'$ ,  $\hat{\Gamma}_+ = \hat{Q}^{-1} (\frac{1}{n} \sum_{i=1}^n D_{+i})$ , and  $\hat{\Gamma}_-$  defined analogously. We may analyze the effect of  $\hat{F}_+$  and  $\hat{F}_-$  on the variance of  $\hat{\beta}^{bc}$  as follows. First, as  $\text{Var}(\sqrt{n}(\hat{\beta} - \beta)) \approx V$ , we have

$$\begin{aligned} \mathbb{E}[\text{Var}(\hat{\beta}^{bc} | \hat{F}_+, \hat{F}_-)] &\approx \frac{1}{n} \mathbb{E}[(I + \Gamma_+ \hat{F}_+ + \Gamma_- \hat{F}_-) V (I + \Gamma_+ \hat{F}_+ + \Gamma_- \hat{F}_-)] \\ &= \frac{1}{n} (V + O(n^{-1/2})), \end{aligned}$$

because  $F_+ = \mathbb{E}[\hat{F}_+]$ ,  $F_- = \mathbb{E}[\hat{F}_-]$  and  $\sqrt{n}F_+, \sqrt{n}F_- = O(1)$ . Moreover,

$$\begin{aligned} \text{Var}(\mathbb{E}[\hat{\beta}^{bc} | \hat{F}_+, \hat{F}_-]) &\approx \text{Var}((I + \Gamma_+ \hat{F}_+ + \Gamma_- \hat{F}_-) \beta) \\ &= \frac{F_+(1 - F_+)}{m} \Gamma_+ \beta \beta' \Gamma_+' + \frac{F_-(1 - F_-)}{m} \Gamma_- \beta \beta' \Gamma_-' \\ &\quad - \frac{F_+ F_-}{m} (\Gamma_+ \beta \beta' \Gamma_-' + \Gamma_- \beta \beta' \Gamma_+'), \end{aligned}$$

with  $F_+ = \mathbb{E}[\hat{F}_+]$  and  $F_- = \mathbb{E}[\hat{F}_-]$ . The law of total variance implies that the variance of  $\hat{\beta}^{bc}$  is approximately equal to the sum of these two terms. As  $\sqrt{n}F_+, \sqrt{n}F_- = O(1)$  and  $\sqrt{n}/m = o(1)$ , we have

$$\begin{aligned} \text{Var}(\sqrt{n}(\hat{\beta}^{bc} - \beta)) \\ \approx V + \frac{nF_+(1 - F_+)}{m} \Gamma_+ \beta \beta' \Gamma_+' + \frac{nF_-(1 - F_-)}{m} \Gamma_- \beta \beta' \Gamma_-' + O(n^{-1/2}). \end{aligned} \quad (10)$$

Under the framework adopted in BCHS where  $m/n \rightarrow 0$  with  $n/m^2 \rightarrow 0$ , the second and third terms are  $O(\sqrt{n}/m)$  and therefore dominate the remaining  $O(n^{-1/2})$  term. Thus, it is these two terms that are most critical to match for the purposes of finite-sample variance correction.

### 4.2.2 The Bootstrap Matches the Finite-Sample Variance Correction

We now show that the coupled-label bootstrap with variance correction is able to match the above two  $O(\sqrt{n}/m)$  terms. Recalling the bootstrap expansion (7), the bootstrap expectation of the bias term  $b^*$  is

$$\mathbb{E}^*[b^*] = -\left(\sqrt{n}\hat{F}_+\Gamma_+\beta + \sqrt{n}\hat{F}_-\Gamma_-\beta\right) \rightarrow_p b.$$

Therefore, drawing  $\hat{F}_+^*$  and  $\hat{F}_-^*$  from a distribution centered at  $\hat{F}_+$  and  $\hat{F}_-$  does not affect the bootstrap's ability to reproduce the leading bias term.

To analyze the bootstrap variance of the bias term, first note that conditional on the data and  $\hat{F}_+^*, \hat{F}_-^*$ , the  $\hat{\theta}_i^*(1 - \theta_i^*)$  and  $\theta_i^*(1 - \hat{\theta}_i^*)$  are i.i.d. Bernoulli and satisfy

$$\begin{aligned} \mathbb{E}^*[\hat{\theta}_i^*(1 - \theta_i^*)|\hat{F}_+^*, \hat{F}_-^*] &= \hat{F}_+^*, & \mathbb{E}^*[\theta_i^*(1 - \hat{\theta}_i^*)|\hat{F}_+^*, \hat{F}_-^*] &= \hat{F}_-^*, \\ \text{Var}^*(\hat{\theta}_i^*(1 - \theta_i^*)|\hat{F}_+^*, \hat{F}_-^*) &= \hat{F}_+^*(1 - \hat{F}_+^*), & \text{Var}^*(\theta_i^*(1 - \hat{\theta}_i^*)|\hat{F}_+^*, \hat{F}_-^*) &= \hat{F}_-^*(1 - \hat{F}_-^*), \end{aligned}$$

and

$$\text{Cov}^*(\hat{\theta}_i^*(1 - \theta_i^*), \theta_i^*(1 - \hat{\theta}_i^*)|\hat{F}_+^*, \hat{F}_-^*) = -\hat{F}_+^*\hat{F}_-^*.$$

Using these expressions and the fact that  $m\hat{F}_+^*$  and  $m\hat{F}_-^*$  are independent Binomial( $m, \hat{F}_+$ ) and Binomial( $m, \hat{F}_-$ ) random variables conditional on the data, we have

$$\begin{aligned} \text{Var}^*\left(\mathbb{E}^*[b^*|\hat{F}_+^*, \hat{F}_-^*]\right) &= \text{Var}^*\left(\sqrt{n}\hat{F}_+^*\Gamma_+\beta + \sqrt{n}\hat{F}_-^*\Gamma_-\beta\right) \\ &= \frac{n\hat{F}_+(1 - \hat{F}_+)}{m}\Gamma_+\beta\beta'\Gamma_+' + \frac{n\hat{F}_-(1 - \hat{F}_-)}{m}\Gamma_-\beta\beta'\Gamma_-' . \end{aligned}$$

Moreover,

$$\begin{aligned} \text{Var}^*(b^*|\hat{F}_+^*, \hat{F}_-^*) &= \hat{F}_+^*(1 - \hat{F}_+^*)\Gamma_+\beta\beta'\Gamma_+' + \hat{F}_-^*(1 - \hat{F}_-^*)\Gamma_-\beta\beta'\Gamma_-' \\ &\quad - \hat{F}_+^*\hat{F}_-^*(\Gamma_+\beta\beta'\Gamma_-' + \Gamma_-\beta\beta'\Gamma_+'), \end{aligned}$$

from which we see that

$$\begin{aligned} \mathbb{E}^*\left[\text{Var}^*(b^*|\hat{F}_+^*, \hat{F}_-^*)\right] &= \left(1 - \frac{1}{m}\right)\left(\hat{F}_+(1 - \hat{F}_+)\Gamma_+\beta\beta'\Gamma_+' + \hat{F}_-(1 - \hat{F}_-)\Gamma_-\beta\beta'\Gamma_-' \right) \\ &\quad - \hat{F}_+\hat{F}_-(\Gamma_+\beta\beta'\Gamma_-' + \Gamma_-\beta\beta'\Gamma_+') \\ &= O_p(n^{-1/2}). \end{aligned}$$

Hence, by the law of total variance, we have,

$$\text{Var}^*(b^*) = \frac{n\hat{F}_+(1 - \hat{F}_+)}{m} \Gamma_+ \beta \beta' \Gamma'_+ + \frac{n\hat{F}_-(1 - \hat{F}_-)}{m} \Gamma_- \beta \beta' \Gamma'_- + O_p(n^{-1/2}).$$

And since  $\sqrt{n}\hat{F}_+ \rightarrow_p \kappa_+$  and  $\sqrt{n}\hat{F}_- \rightarrow_p \kappa_-$ , we have

$$\text{Var}^*(b^*) = \frac{nF_+(1 - F_+)}{m} \Gamma_+ \beta \beta' \Gamma'_+ + \frac{nF_-(1 - F_-)}{m} \Gamma_- \beta \beta' \Gamma'_- + o_p(\sqrt{n}/m).$$

These leading terms match  $O(\sqrt{n}/m)$  terms in (10), as required.

### 4.3 Rotation

As discussed previously, an additional source of finite-sample distortions associated with the coupled-label bootstrap (and its variance-corrected variant) is the discrepancy between the inverse bootstrap Hessian matrix  $\hat{Q}^{*-1}$  and its sample analogue  $\hat{Q}^{-1}$ . If this difference is large, the remainder term in the bootstrap expansion (7) is not negligible, introducing a bias in the bootstrap distribution.

One way to solve this problem is to bootstrap the distribution of  $\sqrt{n}\hat{R}^*(\hat{\beta}^* - \hat{\beta})$ , where  $\hat{R}^* = \hat{Q}^{-1}\hat{Q}^*$ . This is equivalent to bootstrapping the distribution of

$$\sqrt{n}(\tilde{\beta}^* - \hat{R}^*\hat{\beta}), \quad \tilde{\beta}^* = (\hat{X}'\hat{X})^{-1}\hat{X}'^*Y^*,$$

where  $\hat{X}^*$  and  $Y^*$  are obtained by the variance-corrected coupled-label bootstrap. With this rotation, the bias term of the bootstrap expansion in (6) becomes

$$\hat{R}^*\hat{b}^* = -\hat{Q}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\theta}_i^*(1 - \theta_i^*)D_{+i}\hat{\beta} + \theta_i^*(1 - \hat{\theta}_i^*)D_{-i}\hat{\beta},$$

thus no longer depending on  $\hat{Q}^{*-1}$ . Since  $\hat{R}^* \rightarrow_{p^*} I_k$ , the rotated coupled-label bootstrap with variance correction is asymptotically valid under the same conditions as those stated in Theorem 4.

## 5 Simulations

We analyze the finite-sample properties of the bootstrap methods by extending the simulation design of [BCHS](#) in two directions. First, we consider an interactions model in which a standard normal covariate  $Z_i$  is interacted with the binary label  $\theta_i$  so that the effect of  $Z_i$  on  $Y_i$  differs between  $\theta_i = 1$  and  $\theta_i = 0$ . Second, we allow  $p_i = \mathbb{P}(\theta_i = 1)$  to vary across observations to introduce correlation between  $\theta_i$  and the matrices  $D_{+i}$  and  $D_{-i}$ , which, as shown above, invalidates the fixed-label bootstrap. The model is:

$$\begin{aligned} Y_i &= 10 + \theta_i Z_i + Z_i + (0.3 + 0.2 \theta_i) u_i, \\ Z_i &\sim N(0, 1), \quad u_i \sim \text{i.i.d. } N(0, 1), \\ p_i &= F(Z_i^2), \end{aligned}$$

where  $F(\cdot)$  is the cumulative distribution function of a  $\chi^2(1)$  random variable, which implies that  $p_i \sim U[0, 1]$ . The joint distribution of the label pair  $(\theta_i, \hat{\theta}_i)'$  is given by

$$\begin{aligned} \mathbb{P}(\theta_i = 1, \hat{\theta}_i = 1) &= \tilde{p}_i - F_-, & \mathbb{P}(\theta_i = 1, \hat{\theta}_i = 0) &= F_-, \\ \mathbb{P}(\theta_i = 0, \hat{\theta}_i = 1) &= F_+, & \mathbb{P}(\theta_i = 0, \hat{\theta}_i = 0) &= 1 - \tilde{p}_i - F_+. \end{aligned}$$

We set  $\kappa_+ = \kappa_- = \kappa$  and  $F_+ = F_- = \kappa/\sqrt{n}$ . To ensure all probabilities remain non-negative, we map  $p_i$  into

$$\tilde{p}_i = p_i \cdot 2(\bar{p} - F_+) + F_+,$$

so that  $\tilde{p}_i$  is uniformly distributed between  $F_+$ , and  $2\bar{p} - F_+$ , where  $\bar{p}$  is the average probability that  $\theta_i = 1$ .

We consider two values of  $\bar{p} \in \{0.05, 0.5\}$ , matching the constant probabilities used by [BCHS](#). The correlation between  $\theta_i$  and  $Z_i^2$  depends strongly on  $\bar{p}$ : it is approximately 0.02 when  $\bar{p} = 0.05$  but rises to approximately 0.31 when  $\bar{p} = 0.5$ . As theory predicts, this has important consequences for the performance of the fixed-label bootstrap, which requires mean independence between  $\theta_i$  and  $(D_{+i}, D_{-i})$ .

The bootstrap DGP takes the form

$$\begin{aligned} Y_i^* &= \hat{\beta}' X_i^* + u_i^*, & u_i^* &= \hat{u}_i \eta_i, & \eta_i &\sim \text{i.i.d. } N(0, 1), \\ X_i^* &= (1, \theta_i^* Z_i, Z_i)', \end{aligned}$$

where  $\hat{\beta}$  is the sample OLS estimator and  $\hat{u}_i$  are the OLS residuals. In all experiments, we use the wild bootstrap with standard normal weights to generate  $u_i^*$ . Four methods for generating  $(\theta_i^*, \hat{\theta}_i^*)$  are considered. One is a naive bootstrap that sets  $\theta_i^* = \hat{\theta}_i^* = \hat{\theta}_i$ . This bootstrap method is denoted as “no-label resampling” in the tables. By not introducing any measurement error in the bootstrap world, it implicitly sets the bootstrap bias to zero. This method is only included as a benchmark to demonstrate the need to resample labels in the bootstrap world. We then consider the fixed-label and coupled-label methods from Section 3. Finally, we also include the variant of the coupled-label bootstrap with the finite-sample modifications discussed in Section 4. We use  $B = 499$  bootstrap replications and 10,000 Monte Carlo replications throughout. The sample sizes we consider are  $n \in \{8,000; 16,000; 32,000\}$ , and the misclassification rates are  $\kappa \in \{0.5, 1, 1.5\}$ . We vary the size of the external sample  $m$  to keep the ratio  $\sqrt{n}/m$  constant at 0.1265, corresponding to  $m = 707$  for  $n = 8,000$ ,  $m = 1,000$  for  $n = 16,000$ , and  $m = 1,414$  for  $n = 32,000$ .

We focus on inference for the slope coefficient on the interaction term  $\theta_i Z_i$ . For each configuration we report three statistics: the median bias, the empirical coverage rate of a nominal 95% confidence interval, and the median interval length. For bootstrap methods, we report equal-tailed percentile intervals, which dominate symmetric intervals in this setting.

Table 1 presents results for  $\bar{p} = 0.5$ . Each panel contains six estimators: OLS using the imputed regressor  $\hat{\theta}_i Z_i$ , the bias-corrected estimator of BCHS using their finite-sample variance adjustment, and the four bootstrap methods.

The use of an imputed label biases OLS downward, with the bias increasing in  $\kappa$ . The resulting undercoverage is severe: even in the mildest configuration ( $n = 8,000$ ,  $\kappa = 0.5$ ), OLS coverage is only 15.0%. The BCHS method removes virtually all the bias and delivers reliable inference across configurations, though coverage deteriorates modestly as  $\kappa$  rises, and the intervals are substantially wider than those based on OLS.

The no-label-resampling bootstrap simply mimics the OLS estimator, inheriting its bias and producing similarly poor coverage (at best 15.5%).

Table 1: Interactions model,  $\bar{p} = 0.5$ 

$\kappa_+ = \kappa_- =$	Median bias			Coverage (%)			Interval length		
	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5
<i>n</i> = 8,000									
OLS	-0.04	-0.07	-0.11	15.5	0.0	0.0	0.05	0.06	0.06
BCHS bias-corrected & var. adj.	-0.00	-0.01	-0.01	93.8	93.8	92.3	0.08	0.10	0.12
No-label resampling	-0.04	-0.07	-0.11	15.5	0.0	0.0	0.05	0.06	0.06
Fixed-label bootstrap	0.01	0.01	0.01	73.2	69.5	70.8	0.06	0.07	0.08
Coupled-label bootstrap	-0.00	-0.01	-0.02	85.1	80.6	72.6	0.06	0.07	0.07
Coupled-label, rotation & var. adj.	-0.00	-0.01	-0.01	94.5	94.6	92.7	0.08	0.11	0.12
<i>n</i> = 16,000									
OLS	-0.03	-0.05	-0.08	11.5	0.0	0.0	0.03	0.04	0.04
BCHS bias-corrected & var. adj.	-0.00	-0.00	-0.01	94.2	93.8	92.8	0.05	0.07	0.08
No-label resampling	-0.03	-0.05	-0.08	11.5	0.0	0.0	0.03	0.04	0.04
Fixed-label bootstrap	0.01	0.01	0.01	69.5	62.2	61.1	0.04	0.05	0.05
Coupled-label bootstrap	-0.00	-0.01	-0.01	83.6	78.4	72.4	0.04	0.04	0.05
Coupled-label, rotation & var. adj.	-0.00	-0.00	-0.01	94.7	94.7	93.3	0.06	0.07	0.08
<i>n</i> = 32,000									
OLS	-0.02	-0.04	-0.06	9.6	0.0	0.0	0.02	0.02	0.03
BCHS bias-corrected & var. adj.	-0.00	-0.00	-0.00	93.7	93.8	93.1	0.04	0.05	0.06
No-label resampling	-0.02	-0.04	-0.06	9.7	0.0	0.0	0.02	0.02	0.03
Fixed-label bootstrap	0.01	0.01	0.01	64.9	56.7	52.8	0.03	0.03	0.03
Coupled-label bootstrap	-0.00	-0.00	-0.01	80.6	75.5	71.0	0.02	0.03	0.03
Coupled-label, rotation & var. adj.	-0.00	-0.00	-0.00	94.2	94.5	93.8	0.04	0.05	0.06

Note: 10,000 Monte Carlo replications;  $B = 499$  bootstrap replications.

The fixed-label bootstrap overcorrects: it introduces bias in the opposite direction. Coverage remains well below nominal and far below the [BCHS](#) benchmark.

The coupled-label bootstrap, by contrast, correctly replicates the bias of the bias-corrected estimator. Coverage remains below nominal level, however, because the basic coupled-label bootstrap does not account for the additional sampling uncertainty from estimating  $F_+$  and  $F_-$ : interval lengths are close to those of OLS, and the improvement in coverage relative to no-label resampling comes entirely from correctly centering the bootstrap distribution. The coupled-label bootstrap with rotation and variance correction achieves coverage closest to the nominal 95% among all methods, slightly edging the [BCHS](#) analytic correction, while maintaining similar interval lengths.

Table 2 presents results for  $\bar{p} = 0.05$ . This near-singular design is very challenging: the sample contains very few observations with  $\theta_i = 1$ , and the majority of these are

Table 2: Interactions model,  $\bar{p} = 0.05$ 

$\kappa_+ = \kappa_- =$	Median bias			Coverage (%)			Interval length		
	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5
<i>n</i> = 8,000									
OLS	-0.08	-0.16	-0.25	15.0	0.0	0.0	0.10	0.13	0.14
BCHS bias-corrected & var. adj.	-0.01	-0.03	-0.07	93.6	91.4	84.6	0.18	0.25	0.31
No-label resampling	-0.08	-0.16	-0.25	15.4	0.0	0.0	0.10	0.12	0.14
Fixed-label bootstrap	-0.00	-0.02	-0.05	83.2	78.3	66.6	0.12	0.15	0.16
Coupled-label bootstrap	-0.00	-0.02	-0.05	82.6	77.5	65.9	0.12	0.15	0.16
Coupled-label, rotation & var. adj.	-0.01	-0.03	-0.07	93.2	90.4	79.4	0.17	0.22	0.25
<i>n</i> = 16,000									
OLS	-0.05	-0.11	-0.17	12.2	0.0	0.0	0.07	0.08	0.09
BCHS bias-corrected & var. adj.	-0.00	-0.01	-0.03	93.5	92.7	89.6	0.12	0.17	0.20
No-label resampling	-0.05	-0.11	-0.17	12.0	0.0	0.0	0.07	0.08	0.09
Fixed-label bootstrap	-0.00	-0.01	-0.02	81.5	76.9	71.7	0.08	0.10	0.11
Coupled-label bootstrap	-0.00	-0.01	-0.02	80.8	76.2	71.0	0.08	0.10	0.11
Coupled-label, rotation & var. adj.	-0.00	-0.01	-0.03	93.6	92.5	87.8	0.12	0.16	0.18
<i>n</i> = 32,000									
OLS	-0.04	-0.08	-0.12	9.4	0.0	0.0	0.05	0.05	0.06
BCHS bias-corrected & var. adj.	-0.00	-0.01	-0.02	92.6	92.5	90.9	0.08	0.11	0.14
No-label resampling	-0.04	-0.08	-0.12	9.5	0.0	0.0	0.05	0.05	0.06
Fixed-label bootstrap	-0.00	-0.00	-0.01	78.7	73.5	69.5	0.05	0.06	0.07
Coupled-label bootstrap	-0.00	-0.00	-0.01	78.0	72.6	68.6	0.05	0.06	0.07
Coupled-label, rotation & var. adj.	-0.00	-0.01	-0.02	93.0	92.5	90.3	0.08	0.11	0.13

*Note:* See notes to Table 1. With  $\bar{p} = 0.05$ , the correlation between  $\theta_i$  and  $(D_{+i}, D_{-i})$  is low ( $\approx 0.02$ ), so the fixed-label bootstrap performs relatively well.

misclassified when  $\kappa \geq 1$ . As a consequence, the OLS bias is roughly twice as large as in the  $\bar{p} = 0.5$  case. Because the correlation between  $\theta_i$  and  $(D_{+i}, D_{-i})$  is low, the fixed-label bootstrap performs well and even slightly better than the coupled-label bootstrap. However, both maintain coverage below the [BCHS](#) benchmark. The variance correction and Hessian rotation improve coverage, and the gap with the analytic correction from [BCHS](#) decreases with sample size.

We conclude that the coupled-label bootstrap with finite-sample modifications is the preferred bootstrap method in all situations as it delivers coverage close to nominal coverage while producing confidence intervals of competitive length.

Table 3: Estimated Coefficient of Remote Work Status and 95% CIs

Method	No FE		With FE	
	Estimate	95% CI	Estimate	95% CI
OLS	0.649	[0.600, 0.697]	0.364	[0.322, 0.406]
BCHS bias-corrected & var. adj.	0.897	[0.668, 1.126]	0.521	[0.366, 0.677]
No-label resampling	0.648	[0.599, 0.695]	0.363	[0.322, 0.408]
Fixed-label bootstrap	0.898	[0.849, 0.944]	0.522	[0.482, 0.563]
Coupled-label bootstrap	0.896	[0.846, 0.944]	0.510	[0.473, 0.549]
Coupled-label, rotation & var. adj.	0.899	[0.752, 1.062]	0.520	[0.413, 0.643]

*Note:* Point estimates and 95% confidence intervals for the slope coefficient in a regression of log wages on remote work status, with and without occupation (SOC2) and full-time/part-time fixed effects (FE). Results are presented with  $\hat{F}_+ = \hat{F}_- = 0.009$ .

## 6 Application

As an illustration, we revisit BCHS’ regressions investigating the wage premium for remote work. The data are from Hansen et al. (2026), who construct a dataset that imputes remote work status for a corpus of online job postings provided by Lightcast.<sup>3</sup> For each job posting  $i$ , they impute a binary variable  $\hat{\theta}_i$  representing whether or not the job offers remote work. The variable is generated using DistilBERT, a LLM, applied to metadata on occupation, firm location, job title, posted wage, and a textual job description. As in BCHS, we regress log posted wage on  $\hat{\theta}_i$ , with and without occupation (SOC2) and full-time/part-time fixed effects (FEs). We focus on a sample of  $n = 16,315$  job postings from 2022-2023 for the accommodation and food services sector in San Diego, CA. This is a near-singular design: very few jobs in this industry offer remote work ( $\hat{\pi} = 0.024$ ), making bias correction more challenging. We take  $\hat{F}_+ = 0.009$ , as estimated by BCHS using  $m = 1,000$  postings. Since there is no reason to expect the classifier to systematically over- or under-predict, we set  $\hat{F}_- = 0.009$ . As a robustness check, we present a second set of results with  $\hat{F}_- = 0.018$ .

The first set of results with  $\hat{F}_+ = \hat{F}_- = 0.009$  is reported in Table 3. We present OLS estimates and 95% confidence intervals for the coefficient of  $\theta_i$ , alongside results based on the analytical correction in BCHS, and bias-corrected estimates and confidence intervals using the naive no-label resampling bootstrap, the fixed-label bootstrap, and the coupled-label bootstrap with and without modifications. OLS estimates are around 30% smaller than those using the analytical correction. As ex-

<sup>3</sup>See <https://wfhmap.com/>.

Table 4: Estimated Coefficient of Remote Work Status and 95% CIs

Method	No FE		With FE	
	Estimate	95% CI	Estimate	95% CI
OLS	0.649	[0.600, 0.697]	0.364	[0.322, 0.406]
BCHS bias-corrected & var. adj.	0.903	[0.673, 1.134]	0.525	[0.368, 0.682]
No-label resampling	0.648	[0.599, 0.695]	0.364	[0.321, 0.410]
Fixed-label bootstrap	1.048	[0.986, 1.108]	0.603	[0.556, 0.647]
Coupled-label bootstrap	1.047	[0.984, 1.107]	0.591	[0.546, 0.638]
Coupled-label, rotation & var. adj.	0.905	[0.762, 1.068]	0.519	[0.418, 0.640]

*Note:* See notes to Table 3. Results are presented with  $\hat{F}_+ = 0.009$ ,  $\hat{F}_- = 0.018$ .

pected, the naive no-label resampling method produces estimates and CIs that are near identical to those produced by OLS. In the baseline specification without fixed effects, both the fixed-label and coupled-label bootstrap are valid. The estimates and CIs they produce are very similar, as predicted. Our preferred coupled-label bootstrap with variance correction and rotation produces estimates and CIs that are close to those produced by the analytical correction. Similar findings are obtained when FEs are included.

As a robustness check, Table 4 presents results with  $\hat{F}_-$  doubled. The estimates and CIs are largely similar to those presented in Table 3, except for the fixed-label and coupled-label bootstrap (without rotation). For those methods, estimates and CIs are shifted to the right, well above those obtained using analytical methods. This illustrates the importance of using the rotation in near-singular designs.

## 7 Conclusion

In this paper, we study whether the bootstrap can correct bias and deliver valid inference in regressions with binary labels that have been generated by AI or ML methods. We show that a seemingly natural *fixed-label bootstrap*, which generates data using the estimated labels while relying on a corrupted version in estimation, is generally invalid unless a strong independence condition between the latent true labels and other covariates holds. We propose a *coupled-label bootstrap* that jointly resamples the pair of true and imputed labels and show it is valid without such independence restrictions. We also introduce two further finite-sample modifications: a variance correction and a Hessian rotation for near-singular designs. Our recommen-

dation for empirical practice is to use the coupled-label bootstrap with both of these modifications.

## References

- ADAMS-PRASSL, A., T. WATERS, M. BALGOVA, AND M. QIAN (2023): “Firm Concentration & Job Design: The Case of Schedule Flexible Work Arrangements,” Tech. rep., Institute for Fiscal Studies.
- AIGNER, D. J. (1973): “Regression with a Binary Independent Variable Subject to Errors of Observation,” *Journal of Econometrics*, 1, 49–59.
- ANGELOPOULOS, A. N., S. BATES, C. FANNJIANG, M. I. JORDAN, AND T. ZRNIC (2023a): “Prediction-Powered Inference,” *Science*, 382, 669–674.
- ANGELOPOULOS, A. N., J. C. DUCHI, AND T. ZRNIC (2023b): “PPI++: Efficient Prediction-Powered Inference,” *arXiv:2311.01453 [stat.ML]*.
- BATTAGLIA, L., T. CHRISTENSEN, S. HANSEN, AND S. SACHER (2025): “Inference for Regression with Variables Generated by AI or Machine Learning,” *arXiv:2402.15585*.
- BOUND, J., C. BROWN, G. J. DUNCAN, AND W. L. RODGERS (1994): “Evidence on the Validity of Cross-Sectional and Longitudinal Labor Market Data,” *Journal of Labor Economics*, 12, 345–368.
- BOUND, J. AND A. B. KRUEGER (1991): “The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?” *Journal of Labor Economics*, 9, 1–24.
- BURSZTYN, L., T. CHANEY, T. A. HASSAN, AND A. RAO (2024): “The Immigrant Next Door,” *American Economic Review*, 114, 348–384.
- CARLSON, J. AND M. DELL (2026): “A Unifying Framework for Robust and Efficient Inference with Unstructured Data,” *arXiv:2505.00282*.
- CARROLL, R. J. AND M. P. WAND (1991): “Semiparametric Estimation in Logistic Measurement Error Models,” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 53, 573–585.
- CHEN, X., H. HONG, AND E. TAMER (2005): “Measurement Error Models with Auxiliary Data,” *Review of Economic Studies*, 72, 343–366.
- CHEN, X., H. HONG, AND A. TAROZZI (2008): “Semiparametric Efficiency in GMM Models with Auxiliary Data,” *The Annals of Statistics*, 36, 808–843.

- CHESHER, A. (1991): “The Effect of Measurement Error,” *Biometrika*, 78, 451–462.
- DUPAS, P., A. HANDLAN, A. S. MODESTINO, M. NIEDERLE, M. SERÉ, H. SHENG, J. WOLFERS, AND THE SEMINAR DYNAMICS COLLECTIVE (2026): “Gender Differences in Economics Seminars,” *American Economic Review*, 116, 749–789.
- EFRON, B. (1981): “Nonparametric Standard Errors and Confidence Intervals,” *Canadian Journal of Statistics*, 9, 139–158.
- EGAMI, N., M. HINCK, B. M. STEWART, AND H. WEI (2024): “Using Large Language Model Annotations for the Social Sciences: A General Framework of Using Predicted Variables in Downstream Analyses,” Working Paper, Columbia University.
- EVDOKIMOV, K. S. AND A. ZELENEEV (2023): “Simple Estimation of Semiparametric Models with Measurement Errors,” *arXiv:2306.14311 [econ.EM]*.
- FONG, C. AND M. TYLER (2021): “Machine Learning Predictions as Regression Covariates,” *Political Analysis*, 29, 467–484.
- GOLDSMITH-PINKHAM, P. AND K. SHUE (2023): “The Gender Gap in Housing Returns,” *The Journal of Finance*, 78, 1097–1145.
- GONÇALVES, S. AND M. KAFFO (2015): “Bootstrap Inference for Linear Dynamic Panel Data Models with Individual Fixed Effects,” *Journal of Econometrics*, 186, 407–426.
- GONÇALVES, S., J. KOH, AND B. PERRON (2025): “Bootstrap Inference for Group Factor Models,” *Journal of Financial Econometrics*, 23, nbae020.
- GONÇALVES, S. AND B. PERRON (2014): “Bootstrapping Factor-Augmented Regression Models,” *Journal of Econometrics*, 182, 156–173.
- HALL, P. (1992): *The Bootstrap and Edgeworth Expansion*, Springer Series in Statistics, New York, NY: Springer New York.
- HANSEN, S., P. J. LAMBERT, N. BLOOM, S. J. DAVIS, R. SADUN, AND B. TASKA (2026): “Remote Work across Jobs, Companies, and Space,” Working Paper 31007, NBER.
- HIGGINS, A. AND K. JOCHMANS (2024): “Bootstrap Inference for Fixed-Effect Models,” *Econometrica*, 92, 411–427.
- LEE, L.-F. AND J. H. SEPANSKI (1995): “Estimation of Linear and Nonlinear Errors-in-Variables Models Using Validation Data,” *Journal of the American Statistical Association*, 90, 130–140.

LI, X., Y. SHEN, AND Q. ZHOU (2024): “Confidence Intervals of Treatment Effects in Panel Data Models with Interactive Fixed Effects,” *Journal of Econometrics*, 240, 105684.

LUDWIG, J., S. MULLAINATHAN, AND A. RAMBACHAN (2025): “Large Language Models: An Applied Econometric Framework,” *arXiv:2412.07031 [econ]*.

SEPANSKI, J. AND R. CARROLL (1993): “Semiparametric Quasilielihood and Variance Function Estimation in Measurement Error Models,” *Journal of Econometrics*, 58, 223–256.

## A Appendix

### A.1 Proofs for Section 2

*Proof of Theorem 2.* We can write

$$Y_i^* = \hat{\beta}' X_i^* + u_i^* = \hat{\beta}' \hat{X}_i^* - \hat{\beta}' (\hat{X}_i^* - X_i^*) + u_i^*,$$

from which we obtain  $\sqrt{n}(\hat{\beta}^* - \hat{\beta}) = \mathcal{I}_{1n}^* + \mathcal{I}_{2n}^*$ , where

$$\begin{aligned} \mathcal{I}_{1n}^* &= \left( n^{-1} \sum_{i=1}^n \hat{X}_i^* \hat{X}_i^{*'} \right)^{-1} n^{-1/2} \sum_{i=1}^n \hat{X}_i^* u_i^*, \\ \mathcal{I}_{2n}^* &= - \left( n^{-1} \sum_{i=1}^n \hat{X}_i^* \hat{X}_i^{*'} \right)^{-1} n^{-1/2} \sum_{i=1}^n \hat{X}_i^* (\hat{X}_i^* - X_i^*)' \hat{\beta}. \end{aligned}$$

To study the first term, let  $\hat{X}_i^* = \hat{X}_i + (X_i^* - \hat{X}_i) + (\hat{X}_i^* - X_i^*)$ , and write

$$n^{-1} \sum_{i=1}^n \hat{X}_i^* \hat{X}_i^{*'} = n^{-1} \sum_{i=1}^n \hat{X}_i \hat{X}_i' + \mathcal{A}_{1n}^* + \mathcal{B}_{1n}^* + \mathcal{C}_{1n}^* + \mathcal{A}_{2n}^* + \mathcal{B}_{2n}^* + \mathcal{C}_{2n}^* + \mathcal{A}_{3n}^* + \mathcal{B}_{3n}^*,$$

where

$$\begin{aligned} \mathcal{A}_{1n}^* &= \frac{1}{n} \sum_{i=1}^n \hat{X}_i (X_i^* - \hat{X}_i)', & \mathcal{B}_{1n}^* &= (\mathcal{A}_{1n}^*)', & \mathcal{C}_{1n}^* &= \frac{1}{n} \sum_{i=1}^n (X_i^* - \hat{X}_i) (X_i^* - \hat{X}_i)', \\ \mathcal{A}_{2n}^* &= \frac{1}{n} \sum_{i=1}^n \hat{X}_i (\hat{X}_i^* - X_i^*)', & \mathcal{B}_{2n}^* &= (\mathcal{A}_{2n}^*)', & \mathcal{C}_{2n}^* &= \frac{1}{n} \sum_{i=1}^n (\hat{X}_i^* - X_i^*) (\hat{X}_i^* - X_i^*)', \end{aligned}$$

and

$$\mathcal{A}_{3n}^* = \frac{1}{n} \sum_{i=1}^n (X_i^* - \hat{X}_i)(\hat{X}_i^* - X_i^*)', \quad \mathcal{B}_{3n}^* = (\mathcal{A}_{3n}^*)'.$$

We can show that  $n^{-1} \sum_{i=1}^n \hat{X}_i \hat{X}_i'$  converges in probability to  $Q$  by Assumptions 1(i) and 2(i), whereas the remainder terms converge to 0 by applying Cauchy–Schwarz inequality and Assumption B.1(i). Hence,  $n^{-1} \sum_{i=1}^n \hat{X}_i^* \hat{X}_i^{*'} \rightarrow_{p^*} Q$ . Since

$$n^{-1/2} \sum_{i=1}^n \hat{X}_i^* u_i^* = n^{-1/2} \sum_{i=1}^n \hat{X}_i u_i^* + n^{-1/2} \sum_{i=1}^n (\hat{X}_i^* - X_i^*) u_i^* + n^{-1/2} \sum_{i=1}^n (X_i^* - \hat{X}_i) u_i^*,$$

where the last two terms converge to zero by Assumption B.1(iii), we can then use Assumption B.2 to obtain  $\mathcal{I}_{1n}^* \rightarrow_{d^*} N(0, V)$ , with  $V = Q^{-1} \Sigma Q^{-1}$ .

For the second term, we can use Assumption B.1(ii) to obtain  $\mathcal{I}_{2n}^* \rightarrow_{p^*} b$  with  $b = -Q^{-1} B \beta$ .  $\square$

## A.2 Proofs for Section 3

*Proof of Lemma 1.* First, Assumption 2 holds under Assumptions 3(i)–(ii). To verify Assumption 1(i), we first note by Assumption 3(iv) that

$$\mathbb{E}[\mathbb{I}[\hat{\theta}_i \neq \theta_i]] = \mathbb{E}[\hat{\theta}_i(1 - \theta_i)] + \mathbb{E}[\theta_i(1 - \hat{\theta}_i)] \rightarrow 0. \quad (11)$$

We now use Markov’s inequality, noting that

$$\begin{aligned} \mathbb{E}[\|\hat{X}_i - X_i\|^2] &= \mathbb{E}[\|g(1, Z_i) - g(0, Z_i)\|^2 \mathbb{I}[\hat{\theta}_i \neq \theta_i]] \\ &\leq 4\mathbb{E}[G_i^2 \mathbb{I}[\hat{\theta}_i \neq \theta_i]] \\ &\leq 4\mathbb{E}[G_i^4]^{\frac{1}{2}} \mathbb{E}[\mathbb{I}[\hat{\theta}_i \neq \theta_i]]^{\frac{1}{2}} \\ &\rightarrow 0, \end{aligned}$$

where the first equality is by definition, the second and third lines are by the triangle and Cauchy–Schwarz inequalities, and the final line is by Assumption 3(ii) and (11). For Assumption 1(iii), first note by Assumption 3(iii) that  $\sqrt{n} \mathbb{E}[(\hat{X}_i - X_i) u_i] \rightarrow 0$ .

Then

$$\begin{aligned}
& \mathbb{E} \left[ \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( (\hat{X}_i - X_i)u_i - \mathbb{E}[(\hat{X}_i - X_i)u_i] \right) \right\|^2 \right] \\
& \leq \mathbb{E} \left[ \|(\hat{X}_i - X_i)u_i\|^2 \right] \\
& = \mathbb{E} \left[ \|g(1, Z_i) - g(0, Z_i)\|^2 u_i^2 \mathbb{I}[\hat{\theta}_i \neq \theta_i] \right] \\
& \leq 4\mathbb{E}[G_i^2 u_i^2 \mathbb{I}[\hat{\theta}_i \neq \theta_i]] \\
& \leq 4\mathbb{E}[G_i^{4+\delta}]^{\frac{2}{4+\delta}} \mathbb{E}[|u_i|^{4+\delta}]^{\frac{2}{4+\delta}} \mathbb{E}[\mathbb{I}[\hat{\theta}_i \neq \theta_i]]^{\frac{\delta}{4+\delta}} \rightarrow 0,
\end{aligned}$$

again by Assumption 3(i)–(ii) and (11).

It remains to verify Assumption 1(ii). To this end, we first write

$$\mathcal{I}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{X}_i (\hat{X}_i - X_i)' = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( \hat{\theta}_i (1 - \theta_i) D_{+i} + \theta_i (1 - \hat{\theta}_i) D_{-i} \right).$$

Taking expectations, we have by Assumption 3(iv)–(v) that

$$\begin{aligned}
\mathbb{E}[\mathcal{I}_n] &= \sqrt{n} \mathbb{E}[\hat{\theta}_i (1 - \theta_i) D_{+i} + \theta_i (1 - \hat{\theta}_i) D_{-i}] \\
&\rightarrow \kappa_+ \mathbb{E}[D_{+i}] + \kappa_- \mathbb{E}[D_{-i}].
\end{aligned}$$

Finally, with  $\|\cdot\|_F$  denoting the Frobenius norm, we have

$$\begin{aligned}
\mathbb{E} \left[ \|\mathcal{I}_n - \mathbb{E}[\mathcal{I}_n]\|_F^2 \right] &\leq \mathbb{E} \left[ \|D_{+i}\|_F^2 \mathbb{I}[\hat{\theta}_i = 1, \theta_i = 0] + \|D_{-i}\|_F^2 \mathbb{I}[\theta_i = 1, \hat{\theta}_i = 0] \right] \\
&\leq \mathbb{E}[G_i^4 \mathbb{I}[\hat{\theta}_i \neq \theta_i]] \\
&\leq \mathbb{E}[G_i^{4+\delta}]^{\frac{4}{4+\delta}} \mathbb{E}[\mathbb{I}[\hat{\theta}_i \neq \theta_i]]^{\frac{\delta}{4+\delta}} \rightarrow 0,
\end{aligned}$$

by Assumption 3(i)–(ii) and (iv). Hence,  $\mathcal{I}_n \rightarrow_p \kappa_+ \mathbb{E}[D_{+i}] + \kappa_- \mathbb{E}[D_{-i}]$ .  $\square$

To prove Lemma 2 we rely on Lemmas A.1–A.3 below. Lemma A.1 verifies Assumption B.2 for the wild bootstrap. Lemma A.2 verifies Assumptions B.1(i) and (iii), and Lemma A.3 verifies Assumption B.1(ii). For Lemma A.1, Assumption 3 and standard conditions on the wild bootstrap suffice. No assumptions on  $\hat{\theta}_i^*$  are required.

**Lemma A.1.** *Suppose that Assumption 3 holds. Let  $u_i^* = \hat{u}_i \eta_i$  where  $\eta_i$  are i.i.d.  $(0, 1)$  such that  $\mathbb{E}^*[\|\eta_i\|^{2+\delta}] < C < \infty$  for  $\delta > 0$ . Then Assumption B.2 holds.*

*Proof of Lemma A.1.* We can write

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{X}_i u_i^* = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i u_i^* + \frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{X}_i - X_i) u_i^* = \mathcal{Z}_{1n}^* + \mathcal{Z}_{2n}^*. \quad (12)$$

We will show (i)  $\mathcal{Z}_{1n}^* \rightarrow_{d^*} N(0, \Sigma)$ , and (ii)  $\mathcal{Z}_{2n}^* \rightarrow_{p^*} 0$ . Then Assumption B.2 follows from (i) and (ii). We start with (i). Write

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i u_i^* = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i u_i \eta_i + \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i (\hat{u}_i - u_i) \eta_i. \quad (13)$$

Next we show that the first term in (13) follows a bootstrap CLT. Note that  $\mathbb{E}^*[X_i u_i \eta_i] = 0$  and  $\text{Var}^*(n^{-1/2} \sum_{i=1}^n X_i u_i \eta_i) = n^{-1} \sum_{i=1}^n X_i X_i' u_i^2 \rightarrow_p \Sigma \equiv \mathbb{E}[X_i X_i' u_i^2]$  under Assumption 3(i)–(ii). Since, conditional on the original sample, the draws  $X_i u_i \eta_i$  are independent but non-identically distributed, we verify Lyapunov's condition using the Cramer–Wold device. Let  $\alpha \in \mathbb{R}^k$  such that  $\alpha' \alpha = 1$ . We show that

$$n^{-(1+\delta/2)} \sum_{i=1}^n \mathbb{E}^*[|\alpha' X_i u_i \eta_i|^{2+\delta}] \rightarrow_p 0 \quad (14)$$

for  $\delta > 0$ . Since  $\mathbb{E}^*[|\alpha' X_i u_i \eta_i|^{2+\delta}] \leq \|X_i\|^{2+\delta} |u_i|^{2+\delta} \mathbb{E}^*[|\eta_i|^{2+\delta}]$ , where  $\mathbb{E}^*[|\eta_i|^{2+\delta}] < C$  by assumption, an application of the Cauchy–Schwarz inequality yields

$$\begin{aligned} \frac{1}{n^{1+\delta/2}} \sum_{i=1}^n \mathbb{E}^*[|\alpha' X_i u_i \eta_i|^{2+\delta}] &\leq \frac{C}{n^{\delta/2}} \left( \frac{1}{n} \sum_{i=1}^n \|X_i\|^{2(2+\delta)} \right)^{1/2} \left( \frac{1}{n} \sum_{i=1}^n |u_i|^{2(2+\delta)} \right)^{1/2} \\ &= O_p(n^{-\delta/2}), \end{aligned}$$

provided  $\mathbb{E}[\|X_i\|^{2(2+\delta)}] < \infty$  and  $\mathbb{E}[|u_i|^{2(2+\delta)}] < \infty$ , which is implied by Assumption 3(i)–(ii) with a suitable redefinition of  $\delta$  above. This proves (14).

Next, we show that the second term in (13) is asymptotically negligible, i.e.,  $n^{-1/2} \sum_{i=1}^n X_i (\hat{u}_i - u_i) \eta_i \rightarrow_{p^*} 0$ . Since  $\mathbb{E}^*[\eta_i] = 0$ , this term is centered at zero, and it suffices to prove that its bootstrap variance converges to zero in probability. Since

$\mathbb{E}^*[\eta_i^2] = 1$ , we have that

$$\begin{aligned} \text{Var}^* \left( n^{-1/2} \sum_{i=1}^n X_i (\hat{u}_i - u_i) \eta_i \right) &= \frac{1}{n} \sum_{i=1}^n X_i X_i' (\hat{u}_i - u_i)^2 \\ &\leq \frac{3}{n} \sum_{i=1}^n X_i X_i' [X_i' (\hat{\beta} - \beta)]^2 + \frac{3}{n} \sum_{i=1}^n X_i X_i' [(\hat{X}_i - X_i)' \beta]^2 \\ &\quad + \frac{3}{n} \sum_{i=1}^n X_i X_i' [(\hat{X}_i - X_i)' (\hat{\beta} - \beta)]^2 = 3(\mathcal{A}_{1n} + \mathcal{A}_{2n} + \mathcal{A}_{3n}). \end{aligned} \quad (15)$$

We can bound the Frobenius norm of  $\mathcal{A}_{1n}$  as

$$\left\| \frac{1}{n} \sum_{i=1}^n X_i X_i' [X_i' (\hat{\beta} - \beta)]^2 \right\| \leq \frac{1}{n} \sum_{i=1}^n \|X_i X_i'\| |X_i' (\hat{\beta} - \beta)|^2 \leq \frac{1}{n} \sum_{i=1}^n \|X_i\|^4 \|\hat{\beta} - \beta\|^2 = o_p(1),$$

since  $\mathbb{E}[\|X_i\|^4] < \infty$  and  $\|\hat{\beta} - \beta\|^2 = O_p(n^{-1}) = o_p(1)$  under Assumption 3. To prove that  $\mathcal{A}_{2n} = o_p(1)$ , we can use Cauchy–Schwarz inequality to bound its Frobenius norm as

$$\begin{aligned} \left\| \frac{1}{n} \sum_{i=1}^n X_i X_i' [(\hat{X}_i - X_i)' \beta]^2 \right\| &\leq \frac{1}{n} \sum_{i=1}^n \|X_i\|^2 \|\hat{X}_i - X_i\|^2 \|\beta\|^2 \\ &\leq \left( \frac{1}{n} \sum_{i=1}^n \|X_i\|^4 \right)^{1/2} \left( \frac{1}{n} \sum_{i=1}^n \|\hat{X}_i - X_i\|^4 \right)^{1/2} \|\beta\|^2, \end{aligned}$$

where the first factor is  $O_p(1)$  if  $\mathbb{E}[\|X_i\|^4] < \infty$  and the second term is  $o_p(1)$  under Assumption 3. To see this, note that  $\hat{X}_i - X_i = (\hat{\theta}_i - \theta_i)(g(1, Z_i) - g(0, Z_i))$ . By Hölder's inequality with  $p, q > 1$  such that  $p^{-1} + q^{-1} = 1$ ,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \|\hat{X}_i - X_i\|^4 &\leq \frac{1}{n} \sum_{i=1}^n \mathbb{I}[\hat{\theta}_i \neq \theta_i] \|g(1, Z_i) - g(0, Z_i)\|^4 \\ &\leq \left( \frac{1}{n} \sum_{i=1}^n \mathbb{I}[\hat{\theta}_i \neq \theta_i] \right)^{1/p} \left( \frac{1}{n} \sum_{i=1}^n \|g(1, Z_i) - g(0, Z_i)\|^{4q} \right)^{1/q} \\ &\leq \left( \frac{1}{n} \sum_{i=1}^n \mathbb{I}[\hat{\theta}_i \neq \theta_i] \right)^{\delta/(4+\delta)} \left( \frac{1}{n} \sum_{i=1}^n \|g(1, Z_i) - g(0, Z_i)\|^{4+\delta} \right)^{4/(4+\delta)}, \end{aligned}$$

setting  $q = 1 + \delta/4$  and  $p = (4 + \delta)/\delta$ . Under Assumption 3(ii),  $\mathbb{E}[G_i^{4+\delta}] < \infty$ ,

which implies that the second factor is  $O_p(1)$ , whereas the first factor is  $o_p(1)$  by Assumption 3(iv). The proof that  $\mathcal{A}_{3n} = o_p(1)$  follows similarly.

We end the proof of the lemma by showing that  $\mathcal{Z}_{2n}^* \rightarrow_{p^*} 0$ . Since  $\mathbb{E}^*[u_i^*] = 0$ , we have  $\mathbb{E}^*[\mathcal{Z}_{2n}^*] = 0$  and the result follows by showing that  $\text{Var}^*(\mathcal{Z}_{2n}^*) \rightarrow_p 0$ . Using the fact that  $u_i^* = \hat{u}_i \eta_i$  with  $\eta_i$  i.i.d.  $(0, 1)$  yields

$$\text{Var}^*(\mathcal{Z}_{2n}^*) = \frac{1}{n} \sum_{i=1}^n (\hat{X}_i - X_i)(\hat{X}_i - X_i)' \hat{u}_i^2,$$

whose Frobenius norm is bounded by

$$\frac{1}{n} \sum_{i=1}^n \|\hat{X}_i - X_i\|^2 |\hat{u}_i|^2 \leq \left( \frac{1}{n} \sum_{i=1}^n \|\hat{X}_i - X_i\|^4 \right)^{1/2} \left( \frac{1}{n} \sum_{i=1}^n |\hat{u}_i|^4 \right)^{1/2}.$$

We have shown above that the first factor is  $o_p(1)$  under Assumption 3. For the second, note  $|\hat{u}_i|^4 \leq 8|u_i|^4 + 8|u_i - \hat{u}_i|^4$ , where  $\frac{1}{n} \sum_{i=1}^n |u_i|^4 = O_p(1)$  by Assumption 3(i), and  $\frac{1}{n} \sum_{i=1}^n |\hat{u}_i - u_i|^4 = o_p(1)$  can be shown by similar arguments to those used in (15).  $\square$

Lemma A.2 requires additional conditions. In particular, we assume that  $u_i^*$  is obtained by the wild bootstrap independently of  $(\theta_i^*, \hat{\theta}_i^*)$ , conditional on the data, and impose the following high-level condition on the bootstrap labels  $(\theta_i^*, \hat{\theta}_i^*)$ .

**Assumption L.1.** (i)  $\frac{1}{n} \sum_{i=1}^n \mathbb{E}^*[\mathbb{I}[\hat{\theta}_i^* \neq \theta_i^*]] \rightarrow_p 0$ .

(ii)  $\frac{1}{n} \sum_{i=1}^n \mathbb{E}^*[\mathbb{I}[\theta_i^* \neq \hat{\theta}_i^*]] \rightarrow_p 0$ .

Assumption L.1(i) requires the overall misclassification rate of the bootstrap labels to converge to zero in probability. This condition is implied by Assumption L(ii). Assumption L.1(ii) reproduces Assumption L(i).

**Lemma A.2.** *Suppose that Assumption 3 holds. Let  $u_i^* = \hat{u}_i \eta_i$  where  $\eta_i$  are i.i.d.  $(0, 1)$  independently of  $(\theta_i^*, \hat{\theta}_i^*)$  and assume that  $\theta_i^*$  and  $\hat{\theta}_i^*$  satisfy Assumption L.1. Then Assumptions B.1(i) and B.1(iii) hold.*

*Proof of Lemma A.2.* The first part of Assumption B.1(i) follows by noting that

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \|\hat{X}_i^* - X_i^*\|^2 &= \frac{1}{n} \sum_{i=1}^n \mathbb{I}[\hat{\theta}_i^* \neq \theta_i^*] \|g(1, Z_i) - g(0, Z_i)\|^2 \\ &\leq \left( \frac{1}{n} \sum_{i=1}^n \mathbb{I}[\hat{\theta}_i^* \neq \theta_i^*] \right)^{1/2} \left( \frac{1}{n} \sum_{i=1}^n \|g(1, Z_i) - g(0, Z_i)\|^4 \right)^{1/2}, \end{aligned}$$

which is  $o_{p^*}(1)$  provided the first factor is  $o_{p^*}(1)$  (since the second factor is  $O_p(1)$  under Assumption 3(ii)). Assumption L.1(i) suffices for this given Markov's inequality, noting that  $\mathbb{I}(\hat{\theta}_i^* \neq \theta_i^*)$  is non-negative. An analogous argument, in conjunction with Assumption L.1(ii), verifies the second part of Assumption B.1(i).

Next, we verify Assumption B.1(iii). We focus on showing the first part, since the second part follows by an analogous argument. Note that  $\mathbb{E}^*[(\hat{X}_i^* - X_i^*)u_i^*] = \mathbb{E}^*[(\hat{X}_i^* - X_i^*)\mathbb{E}^*[u_i^*]] = 0$  since  $u_i^* = \hat{u}_i\eta_i$ , where  $\eta_i$  is independent of  $\hat{u}_i$  and  $\hat{X}_i^* - X_i^*$ , conditional on the sample, and  $\mathbb{E}^*[\eta_i] = 0$ . Hence, it suffices to show that  $\text{Var}^*(n^{-1/2} \sum_{i=1}^n (\hat{X}_i^* - X_i^*)u_i^*) \rightarrow_p 0$ . To verify this condition, we can use the independence between  $\eta_i$  and  $\hat{X}_j^* - X_j^*$  for all  $i, j$  and the fact that  $u_i^*$  is independent across  $i$  (by the wild bootstrap property). These assumptions imply that

$$\begin{aligned} &\text{Var}^* \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{X}_i^* - X_i^*)u_i^* \right) \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}^*[(\hat{X}_i^* - X_i^*)u_i^* (\hat{X}_j^* - X_j^*)' u_j^*] \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}^*[(\hat{X}_i^* - X_i^*)(\hat{X}_i^* - X_i^*)'] \mathbb{E}^*[u_i^{*2}] \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}^*[(\hat{\theta}_i^* - \theta_i^*)^2] (g(1, Z_i) - g(0, Z_i))(g(1, Z_i) - g(0, Z_i))' \hat{u}_i^2 \end{aligned}$$

where the last equality follows because  $\hat{X}_i^* - X_i^* = (\hat{\theta}_i^* - \theta_i^*)(g(1, Z_i) - g(0, Z_i))$ . Noting that  $\mathbb{E}^*[(\hat{\theta}_i^* - \theta_i^*)^2] = \mathbb{E}^*[\mathbb{I}[\hat{\theta}_i^* \neq \theta_i^*]]$ , we can bound the Frobenius norm of the

bootstrap variance above by

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \mathbb{E}^* [\mathbb{I}[\hat{\theta}_i^* \neq \theta_i^*]] \|g(1, Z_i) - g(0, Z_i)\|^2 |\hat{u}_i^2| \\ & \leq \left( \frac{1}{n} \sum_{i=1}^n (\mathbb{E}^* [\mathbb{I}[\hat{\theta}_i^* \neq \theta_i^*]])^p \right)^{1/p} \left( \frac{1}{n} \sum_{i=1}^n (\|g(1, Z_i) - g(0, Z_i)\|^{2q} |\hat{u}_i^2|^q) \right)^{1/q} \end{aligned}$$

for any  $p, q > 1$  such that  $p^{-1} + q^{-1} = 1$ . Note  $(\mathbb{E}^* [\mathbb{I}[\hat{\theta}_i^* \neq \theta_i^*]])^p \leq \mathbb{E}^* [\mathbb{I}[\hat{\theta}_i^* \neq \theta_i^*]]$  by Jensen's inequality. Taking  $q = 1 + \delta/4$  for  $\delta > 0$  as in Assumption 1 implies  $p^{-1} = \delta/(4 + \delta)$ , yielding the following bound:

$$\left( \frac{1}{n} \sum_{i=1}^n \mathbb{E}^* [\mathbb{I}[\hat{\theta}_i^* \neq \theta_i^*]] \right)^{\delta/(4+\delta)} \left( \frac{1}{n} \sum_{i=1}^n \|g(1, Z_i) - g(0, Z_i)\|^{2+\delta/2} |\hat{u}_i^2|^{2+\delta/2} \right)^{4/(4+\delta)}.$$

Assumption L.1(i) implies that the first factor is  $o_p^*(1)$ . It remains to show the second factor is  $O_p(1)$ . By the Cauchy–Schwarz inequality,

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \|g(1, Z_i) - g(0, Z_i)\|^{2+\delta/2} |\hat{u}_i^2|^{2+\delta/2} \\ & \leq \left( \frac{1}{n} \sum_{i=1}^n \|g(1, Z_i) - g(0, Z_i)\|^{4+\delta} \right)^{1/2} \left( \frac{1}{n} \sum_{i=1}^n |\hat{u}_i^2|^{4+\delta} \right)^{1/2}, \end{aligned}$$

where the first factor on the right-hand side is  $O_p(1)$  by Assumption 3(ii). The second factor can be shown to be  $O_p(1)$ , noting  $|\hat{u}_i^2|^{4+\delta} \leq 2^{3+\delta} (|u_i|^{4+\delta} + |\hat{u}_i - u_i|^{4+\delta})$ , where  $\frac{1}{n} \sum_{i=1}^n |u_i|^{4+\delta} = O_p(1)$  follows by Assumption 3(i), and  $\frac{1}{n} \sum_{i=1}^n |\hat{u}_i - u_i|^{4+\delta} = O_p(1)$  follows by similar arguments to those used in (15).  $\square$

Finally, Lemma A.3 verifies Assumption B.1(ii) under Assumption L(ii)–(iv).

**Lemma A.3.** *Suppose that Assumption 3 holds. If  $\theta_i^*$  and  $\hat{\theta}_i^*$  satisfy Assumption L(ii)–(iv), then Assumption B.1(ii) holds.*

*Proof of Lemma A.3.* Since  $\hat{X}_i^* - X_i^* = (\hat{\theta}_i^* - \theta_i^*)[g(1, Z_i) - g(0, Z_i)]$ , we can write

$$\hat{X}_i^* (\hat{X}_i^* - X_i^*)' = \hat{\theta}_i^* (1 - \theta_i^*) D_{+i} + \theta_i^* (1 - \hat{\theta}_i^*) D_{-i}.$$

Assumption [B.1\(ii\)](#) can be rewritten as

$$B_n^* \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{\theta}_i^*(1 - \theta_i^*)D_{+i} + \theta_i^*(1 - \hat{\theta}_i^*)D_{-i}) \xrightarrow{p^*} \kappa_+ \mathbb{E}[D_{+i}] + \kappa_- \mathbb{E}[D_{-i}] \equiv B,$$

with  $B$  given in [Lemma 1](#). Adding and subtracting yields  $B_n^* = B_{1n}^* + B_{2n}^*$ , where

$$\begin{aligned} B_{1n}^* &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\theta}_i^*(1 - \theta_i^*) \mathbb{E}[D_{+i}] + \theta_i^*(1 - \hat{\theta}_i^*) \mathbb{E}[D_{-i}], \\ B_{2n}^* &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\theta}_i^*(1 - \theta_i^*) (D_{+i} - \mathbb{E}[D_{+i}]) + \theta_i^*(1 - \hat{\theta}_i^*) (D_{-i} - \mathbb{E}[D_{-i}]). \end{aligned}$$

Assumption [L\(iv\)](#) implies that  $B_{2n}^* \xrightarrow{p^*} 0$ . Next, we show that  $B_{1n}^* \xrightarrow{p^*} B$ . Since  $\mathbb{E}[D_{+i}]$  and  $\mathbb{E}[D_{-i}]$  are independent of  $i$ , the result follows by noting that Assumptions [L\(ii\)](#)–[L\(iii\)](#) imply that  $\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\theta}_i^*(1 - \theta_i^*) \xrightarrow{p^*} \kappa_+$  and  $\frac{1}{\sqrt{n}} \sum_{i=1}^n \theta_i^*(1 - \hat{\theta}_i^*) \xrightarrow{p^*} \kappa_-$ , by Chebyshev's inequality.  $\square$

*Proof of [Lemma 2](#).* This follows from [Lemmas A.1–A.3](#) given our assumptions on  $u_i^*$  and [Assumption L](#), and noting that [Assumption L\(ii\)](#) implies [Assumption L.1\(i\)](#).  $\square$

*Proof of [Theorem 3](#).* This is a special case of the proof of [Theorem 4](#) below, setting  $\hat{F}_+^* = \hat{F}_+$  and  $\hat{F}_-^* = \hat{F}_-$ .  $\square$

*Proof of [Theorem 4](#).* In view of [Theorem 2](#) and [Lemma 2](#), it is enough to show that [Assumption L](#) holds.

[Assumption L\(i\)](#) requires  $\frac{1}{n} \sum_{i=1}^n \mathbb{E}^*[\mathbb{I}[\theta_i^* \neq \hat{\theta}_i]] \rightarrow_p 0$ . From the DGP for  $(\theta_i^*, \hat{\theta}_i^*)$  in [\(8\)](#) and [\(9\)](#), we see that if  $\hat{\theta}_i = 1$  then  $\theta_i^* = 0$  with probability  $\hat{F}_+ + \hat{F}_- \frac{1-\hat{\pi}}{\hat{\pi}}$  because  $\mathbb{E}^*[\hat{F}_+^*] = \hat{F}_+$  and  $\mathbb{E}^*[\hat{F}_-^*] = \hat{F}_-$ . Similarly, if  $\hat{\theta}_i = 0$  then  $\theta_i^* = 1$  with probability  $\hat{F}_- + \hat{F}_+ \frac{\hat{\pi}}{1-\hat{\pi}}$ . Therefore,

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E}^*[\mathbb{I}[\theta_i^* \neq \hat{\theta}_i]] = 2\hat{\pi}\hat{F}_+ + 2(1 - \hat{\pi})\hat{F}_- = O_p(n^{-1/2}) \rightarrow_p 0,$$

where we used  $\sqrt{n}\hat{F}_+ \rightarrow_p \kappa_+$  and  $\sqrt{n}\hat{F}_- \rightarrow_p \kappa_-$ . Hence, [Assumption L\(i\)](#) holds.

For part [\(ii\)](#), consider  $\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{E}^*[\hat{\theta}_i^*(1 - \theta_i^*)]$ . As  $\hat{\theta}_i^*(1 - \theta_i^*)$  are independent Bernoulli random variables with common success probability  $\hat{F}_+^*$  conditional on the

data and  $\hat{F}_+^*$ , and  $\mathbb{E}^*[\hat{F}_+^*] = \hat{F}_+$ , we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{E}^*[\hat{\theta}_i^*(1 - \theta_i^*)] = \sqrt{n} \mathbb{E}^*[\hat{F}_+^*] = \sqrt{n} \hat{F}_+ \rightarrow_p \kappa_+,$$

again by  $\sqrt{n} \hat{F}_+ \rightarrow_p \kappa_+$ . The result  $\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{E}^*[\theta_i^*(1 - \hat{\theta}_i^*)] \rightarrow_p \kappa_-$  follows similarly.

For part (iii), consider  $\text{Var}^*\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\theta}_i^*(1 - \theta_i^*)\right)$ . Using the fact that  $\hat{\theta}_i^*(1 - \theta_i^*)$  are independent Bernoulli random variables conditional on the data and  $\hat{F}_+^*$ , we have

$$\text{Var}^*\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\theta}_i^*(1 - \theta_i^*) \middle| \hat{F}_+^*\right) = \text{Var}^*(\hat{\theta}_i^*(1 - \theta_i^*) | \hat{F}_+^*) = \hat{F}_+^*(1 - \hat{F}_+^*).$$

It follows by the fact that  $m\hat{F}_+^* \sim \text{Binomial}(m, \hat{F}_+)$ , that

$$\mathbb{E}^*\left[\text{Var}^*\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\theta}_i^*(1 - \theta_i^*) \middle| \hat{F}_+^*\right)\right] = \hat{F}_+(1 - \hat{F}_+) \left(1 - \frac{1}{m}\right),$$

which is  $O_p(n^{-1/2})$  because  $\sqrt{n} \hat{F}_+ \rightarrow_p \kappa_+$ . Similarly,

$$\mathbb{E}^*\left[\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\theta}_i^*(1 - \theta_i^*) \middle| \hat{F}_+^*\right] = \sqrt{n} \hat{F}_+^*,$$

and so

$$\text{Var}^*\left(\mathbb{E}^*\left[\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\theta}_i^*(1 - \theta_i^*) \middle| \hat{F}_+^*\right]\right) = \frac{n\hat{F}_+(1 - \hat{F}_+)}{m} \rightarrow_p 0,$$

because  $\sqrt{n} \hat{F}_+ \rightarrow_p \kappa_+$  and  $n/m^2 \rightarrow 0$ . It follows by the law of total variance that  $\text{Var}^*\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\theta}_i^*(1 - \theta_i^*)\right) \rightarrow_p 0$ , as required. The result  $\text{Var}^*\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \theta_i^*(1 - \hat{\theta}_i^*)\right) \rightarrow_p 0$  follows similarly.

For part (iv), consider  $\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\theta}_i^*(1 - \theta_i^*)(D_{+i} - \mathbb{E}[D_{+i}])$ . By similar arguments to part (iii), we have

$$\begin{aligned} \mathbb{E}^*\left[\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\theta}_i^*(1 - \theta_i^*)(D_{+i} - \mathbb{E}[D_{+i}])\right] &= \hat{F}_+ \times \frac{1}{\sqrt{n}} \sum_{i=1}^n (D_{+i} - \mathbb{E}[D_{+i}]) \\ &= O_p(n^{-1/2}) \times O_p(1) \rightarrow_p 0, \end{aligned}$$

where the first factor is because  $\sqrt{n} \hat{F}_+ \rightarrow_p \kappa_+$  and the second is by the CLT and

Assumption 3(ii). We now calculate the bootstrap variance of this term. Again using the fact that conditional on  $\hat{F}_+^*$  and the data, the  $\hat{\theta}_i^*(1 - \theta_i^*)$  are independent Bernoulli random variables, we have

$$\mathbb{E}^* \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\theta}_i^*(1 - \theta_i^*)(D_{+i} - \mathbb{E}[D_{+i}]) \middle| \hat{F}_+^* \right] = \hat{F}_+^* \times \frac{1}{\sqrt{n}} \sum_{i=1}^n (D_{+i} - \mathbb{E}[D_{+i}]),$$

and, for notational simplicity, assuming  $D_{+i}$  is scalar (otherwise, the argument applies element-wise):

$$\begin{aligned} \text{Var}^* \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\theta}_i^*(1 - \theta_i^*)(D_{+i} - \mathbb{E}[D_{+i}]) \middle| \hat{F}_+^* \right) \\ = \hat{F}_+^*(1 - \hat{F}_+^*) \times \frac{1}{n} \sum_{i=1}^n (D_{+i} - \mathbb{E}[D_{+i}])^2. \end{aligned}$$

Now using the fact that  $m\hat{F}_+^*$  is Binomial( $m, \hat{F}_+^*$ ) conditional on the data, we have

$$\begin{aligned} \text{Var}^* \left( \mathbb{E}^* \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\theta}_i^*(1 - \theta_i^*)(D_{+i} - \mathbb{E}[D_{+i}]) \middle| \hat{F}_+^* \right] \right) \\ = \frac{\hat{F}_+(1 - \hat{F}_+)}{m} \times \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n (D_{+i} - \mathbb{E}[D_{+i}]) \right)^2 = O_p(1/(m\sqrt{n})) \times O_p(1) \rightarrow_p 0, \end{aligned}$$

where the first factor is because  $\sqrt{n}\hat{F}_+ \rightarrow_p \kappa_+$  and the second is by the CLT and Assumption 3(ii). Similarly,

$$\begin{aligned} \mathbb{E}^* \left[ \text{Var}^* \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\theta}_i^*(1 - \theta_i^*)(D_{+i} - \mathbb{E}[D_{+i}]) \middle| \hat{F}_+^* \right) \right] \\ = \hat{F}_+(1 - \hat{F}_+) \left( 1 - \frac{1}{m} \right) \times \frac{1}{n} \sum_{i=1}^n (D_{+i} - \mathbb{E}[D_{+i}])^2 = O_p(n^{-1/2}) \times O_p(1) \rightarrow_p 0, \end{aligned}$$

where the first factor is because  $\sqrt{n}\hat{F}_+ \rightarrow_p \kappa_+$  and the second is by the LLN and Assumption 3(ii). Therefore, we have shown that both the bootstrap expectation and bootstrap variance of  $\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\theta}_i^*(1 - \theta_i^*)(D_{+i} - \mathbb{E}[D_{+i}])$  converge in probability to zero. It follows that  $\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\theta}_i^*(1 - \theta_i^*)(D_{+i} - \mathbb{E}[D_{+i}]) \rightarrow_{p^*} 0$ , as required. The proof that  $\frac{1}{\sqrt{n}} \sum_{i=1}^n \theta_i^*(1 - \hat{\theta}_i^*)(D_{-i} - \mathbb{E}[D_{-i}]) \rightarrow_{p^*} 0$  is analogous.  $\square$