Bootstrapping with AI/ML-generated labels*

Timothy Christensen† Sílvia Gonçalves‡ Benoit Perron§

March 26, 2026

Abstract

A leading application of AI/ML methods in economics is the generation of binary variables (or labels) using classification algorithms. However, even small amounts of misclassification can produce large biases in OLS estimators and invalidate standard inference. We study whether the bootstrap can correct this bias and restore valid inference. This is a deceptively difficult problem because misclassification introduces nonclassical measurement error. We first show that a seemingly natural *fixed-label bootstrap*, which generates data using the estimated labels while relying on a corrupted version of the labels in estimation, is generally invalid unless a strong independence condition between the latent true labels and other covariates holds. We then introduce a *coupled-label bootstrap* that jointly resamples the pair of true and imputed labels, and we show that it is valid without such independence restrictions. Two finite-sample adjustments further improve coverage: a variance correction for uncertainty in the estimated misclassification rates and a Hessian rotation for near-singular designs. We illustrate the performance of the methods in simulations and apply them to estimate the wage premium for remote work.

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

*We thank Jaap Abbring, 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.

†Department of Economics, Yale University.

‡Department of Economics, McGill University.

§Department of Economics, Université de Montréal.

1 Introduction

It is now standard practice for researchers 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 missing gender of borrowers from borrower names, Bursztyn, Chaney, Hassan, and Rao (2024) impute the missing 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. The imputed variables are rarely of interest in their own right, but are rather introduced into regressions or other simple 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 are data constraints. A widely discussed approach for correcting bias requires access to large amounts of complete validation data, in which both the imputed variables and ground-truth labels are observed alongside other variables in the model.[1] In contemporary economic applications, however, researchers seldom have access to complete validation data. Instead, at best, one has an external sample from which the accuracy of the classifier can be

[1] 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.

measured. For instance, Bursztyn et al. (2024) estimate the accuracy of their classifier using an external sample of North Carolina voter registration data which contains self-reported ethnicity (but is missing data on charitable donations). Moreover, because collecting ground-truth labels is much more costly than generating labels with a classifier, the size m of this external sample is typically 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 investigate the use of the bootstrap, as opposed to analytical methods, to correct bias and restore valid inference. There is a long tradition of using the bootstrap to bias-correct estimators and improve the coverage properties of confidence intervals (Efron, 1981; Hall, 1992). But despite the simplicity of the model, this is a deceptively difficult problem in the binary labels case. Measurement error with imputed labels 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, we consider a regression model

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

where θ_i is a latent binary variable, Y_i and Z_i are observed, and the functional form of g is known. For instance, we might have $g(\theta_i, Z_i) = (\theta_i, Z_i')'$ for a model in which θ_i enters the regression additively, or $g(\theta_i, Z_i) = (\theta_i Z_i', Z_i')'$ for a model with interactions. In practice, an estimate $\hat{\theta}_i$ of the latent θ_i is generated using AI/ML methods. Then Y_i is regressed on $\hat{X}_i \equiv g(\hat{\theta}_i, Z_i)$, and confidence intervals and standard errors for β are reported, treating $\hat{X}_i$ as though it is the truth. Misclassification ($\hat{\theta}_i \neq \theta_i$ for some observations) introduces measurement error in the $\hat{X}_i$. As a consequence, regression of Y_i on $\hat{X}_i$ is biased and the usual approach to inference is 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 $(\theta_i, \hat{\theta}_i)_{i=1}^m$ to estimate classifier accuracy.[2] The question we address is: how should one bootstrap in this scenario?

Naively resampling $(Y_i, \hat{X}_i)$, or equivalent residual-based approaches such as the wild bootstrap, will not work. These do not introduce any measurement error in the

[2]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.

covariates in the bootstrap world, because the same covariates are used in data generation and parameter estimation. As a consequence, this approach fails to reproduce the bias of the OLS estimator and leads to invalid inference.

A seemingly natural approach is a *fixed-label bootstrap*, which is a variant of the wild bootstrap that builds-in measurement error. In this bootstrap, data are first 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 $\hat{\theta}_i$ such 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 scenarios with imputed covariates, such as factor-augmented regressions (Gonçalves and Perron, 2014; Gonçalves, Koh, and Perron, 2025) and various panel data settings (Gonçalves and Kaffo, 2015; Higgins and Jochmans, 2024; Li, Shen, and Zhou, 2024). However, we show that the fixed-label bootstrap is generally inconsistent: its asymptotic bias fails to replicate that of the OLS estimator unless certain independence conditions hold between θ_i and Z_i.

To this end, we introduce a novel *coupled-label bootstrap*. In this bootstrap, data are generated as $Y_i^* = X_i^{*\prime}\hat{\beta} + u_i^*$, where $X_i^* = g(\theta_i^*, Z_i)$ with θ_i^* a bootstrap version of the true latent θ_i and $\hat{\beta}$ and u_i^* are as above, then Y_i^* is regressed on $\hat{X}_i^*$ as before. The novelty of this bootstrap is in generating the couples $(\theta_i^*, \hat{\theta}_i^*)$ so that the misclassification rates for predicting θ_i^* with $\hat{\theta}_i^*$ match those observed in the external sample. We show that this bootstrap is able to replicate the correct bias and variance of the asymptotic distribution of the OLS estimator, justifying the use of confidence sets and hypothesis tests constructed via this bootstrap.

While this theory establishes asymptotic bootstrap validity, we introduce two further modifications to the coupled-label bootstrap to improve its finite-sample performance. The first is a finite-sample variance correction that accounts for additional sampling error introduced by estimating the false positive and negative rates. BCHS develop analytic finite-sample corrections whereas we show how a similar correction can be achieved via simple modification of the bootstrap. The second is a rotation that adjusts the inverse Hessian. This is useful in near-singular designs where only a small fraction of θ_i's are 0 (or 1). In these settings, misclassification can introduce discrepancies between the bootstrap and estimated inverse Hessian that are asymptotically negligible but can be material in finite samples. Both are easy to implement.

Simulations calibrated to our empirical application confirm that these modifications substantially improve finite-sample performance, especially in near-singular designs. Accordingly, our recommendation for empirical practice is to use the coupled-label bootstrap with both of these modifications.

The remainder of the paper is structured as follows. Section 2 first outlines general, "high-level" theory for establishing asymptotic validity of bootstraps with generated covariates. Section 3 then develops theory specifically for the binary labels case. We develop general theory for the wild bootstrap with binary labels, show formally that the fixed-label bootstrap is inconsistent unless certain independence conditions hold, then introduce and establish consistency of the coupled bootstrap. In Section 4 we then introduce two modifications that improve the finite-sample performance of the bootstrap. Section 5 presents simulation evidence in a design calibrated to the application. Finally, Section 6 presents an application revisiting Hansen et al. (2026), who study the link 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 $\mathbb{E}^*$ denote expectation under $\mathbb{P}^*$ conditional on the data. For a sequence of bootstrap random vectors Y_n^*, we write $Y_n^* \to_{p^*} 0$ or $Y_n^* = o_{p^*}(1)$ to denote that for any $\epsilon > 0$, $\mathbb{P}^*(\|Y_n^*\| > \epsilon) \to_p 0$. Similarly, $Y_n^* = O_{p^*}(1)$ denotes that $\limsup_{n\to\infty} \mathbb{P}^*(\|Y_n^*\| > M) \to_p 0$ as $M \to \infty$. Finally, we write $Z_n^* \to_{d^*} Z$ to denote that for all continuity points z of the CDF F of Z, $\mathbb{P}^*(Z_n^* \leq z) - F(z) \to_p 0$.

2 General Case

In this section we present the general case in which Y is regressed on a vector of generated variables $\hat{X}$. We first briefly recap the theory from BCHS for why regression of Y on $\hat{X}$ leads to a bias in the asymptotic distribution of the OLS estimator and, consequently, shifts confidence intervals centered at the OLS estimator away from the truth. We will then present a set of high-level conditions under which the bootstrap can match the bias and variance of the asymptotic distribution of the OLS estimator. This justifies using the bootstrap percentile intervals to perform valid inference. The substantive question is then whether one can design a bootstrap that satisfies these high-level conditions in contexts of interest. In the next section, we will discuss some

specific bootstraps for the binary labels case and discuss when they lead to valid inference.

2.1 Asymptotic Distribution of the OLS Estimator

We wish to perform inference on the parameter vector β 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

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

of β 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) $\frac{1}{n} \sum_{i=1}^{n} \|\hat{X}_i - X_i\|^2 \to_p 0$;

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

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

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

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

Assumption 1 provides a set of conditions on the generated variable $\hat{X}_i$ while Assumption 2 is standard and only involves the latent variable X_i. Assumption 1 allows for both classical and non-classical error in $\hat{X}_i$. The key condition is Assumption 1(ii). In the case of classical measurement error, it is effectively assuming that the variance of the measurement error is $O(n^{-1/2})$, similar to the small measurement error asymptotic frameworks adopted in Chesher (1991) and Evdokimov and Zeleneev (2023). This condition ensures that measurement error remains present asymptotically, but does not overwhelm sampling error in the regression. This is designed to ensure asymptotic theory delivers useful approximations to the finite-sample distribution of $\hat{\beta}$, where both sources of error play roles. In the next section, we provide a set of 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* $n \to \infty$,

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

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 don't 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

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

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 $n \to \infty$.

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

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

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

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

Assumption B.1 is the bootstrap analogue of Assumption 1. Part (i) is used to show that $n^{-1} \sum_{i=1}^{n} \hat{X}_i^* \hat{X}_i^{*\prime} \to_{p^*} Q$, given $n^{-1} \sum_{i=1}^{n} \hat{X}_i \hat{X}_i' \to_p Q$ under Assumptions 1 and 2. Part (iii) and Assumption B.2 imply that $n^{-1/2} \sum_{i=1}^{n} \hat{X}_i^* u_i^* \to_{d^*} N(0, \Sigma)$. 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 $n \to \infty$,*

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

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 β. Specifically, for $j = 1, \ldots, k$, a $100(1-\alpha)\%$ percentile interval for β_j 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 α-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

Now we turn our attention to the binary labels case. We first describe the setup and then specialize Theorem 1 to this case. We then consider two bootstrap methods, both of which are instances of the wild bootstrap. The first is a seemingly natural *fixed-label bootstrap*, which is analogous to the standard fixed-regressor bootstrap. We show that this bootstrap is generally inconsistent unless strong independence conditions hold between θ_i and Z_i. We then present a *coupled-label bootstrap* and show that it is consistent in scenarios where the fixed-label bootstrap is not. Accordingly,

our recommendation for empirical practice is to use the coupled-label bootstrap in conjunction with the finite-sample modifications described in the next section.

3.1 Setup

In this setting,

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

where $g : \{0,1\} \times \mathbb{R}^{d_z} \to \mathbb{R}^k$ is known, θ_i is a latent binary random variable, and $Z_i \in \mathbb{R}^{d_z}$ is an observed random vector. Since θ_i is latent, it is common practice to generate an estimate $\hat{\theta}_i$ of θ_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 θ_i, though it is straightforward to extend our analysis to the vector case.

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

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

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*

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

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 θ_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 $n \to \infty$, 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

$$D_{+i} = g(1,Z_i)(g(1,Z_i) - g(0,Z_i))',$$
$$D_{-i} = g(0,Z_i)(g(0,Z_i) - g(1,Z_i))'.$$

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 and $\mathbb{E}[G_i^{4+\delta}] < \infty$;

(iii) $\mathbb{E}[(\hat{X}_i - X_i)u_i] = o(n^{-1/2})$ as $n \to \infty$;

(iv) $\sqrt{n}\mathbb{E}[\hat{\theta}_i(1-\theta_i)] \to \kappa_+$ and $\sqrt{n}\mathbb{E}[\theta_i(1-\hat{\theta}_i)] \to \kappa_-$ as $n \to \infty$, for $0 \le \kappa_+, \kappa_- < \infty$;

(v) $\sqrt{n}(\mathbb{E}[\hat{\theta}_i(1-\theta_i)D_{+i} + \theta_i(1-\hat{\theta}_i)D_{-i}] - \mathbb{E}[\hat{\theta}_i(1-\theta_i)]\mathbb{E}[D_{+i}] - \mathbb{E}[\theta_i(1-\hat{\theta}_i)]\mathbb{E}[D_{-i}]) \to 0$ as $n \to \infty$.

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 θ_i and $\hat{\theta}_i$ are functions of Z_i and some

auxiliary information $\mathcal{U}_i$. Then by iterated expectations,

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

so it suffices that $\mathbb{E}[u_i|Z_i,\mathcal{U}_i] = 0$ for Assumption (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 κ_+ and κ_- 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 θ_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 in arbitrary ways, provided the unconditional moment conditions hold. 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

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

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,

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

It is important to note that we are not assuming the logistic model is the DGP for θ_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* γ_n *of the classifier maximize the population likelihood*

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

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 $\sqrt{n}\mathbb{E}[\hat{\theta}_i(1-\theta_i)] \to \kappa_+$ *implies that* $\sqrt{n}\mathbb{E}[\theta_i(1-\hat{\theta}_i)] \to \kappa_+$ *also holds, and thus* $\kappa_+ = \kappa_-$.

Example 3. *Suppose that in a sample of size* n, *the* $(\theta_i, \hat{\theta}_i)$ *satisfy*

$$\Pr(\theta_i = \theta_1, \hat{\theta}_i = \theta_2 \mid Z_i = z) = \begin{cases} p_n(z) - F_{n-} + o(n^{-1/2}) & \text{if } \theta_1 = 1, \theta_2 = 1, \\ F_{n-} + o(n^{-1/2}) & \text{if } \theta_1 = 1, \theta_2 = 0, \\ F_{n+} + o(n^{-1/2}) & \text{if } \theta_1 = 0, \theta_2 = 1, \\ 1 - p_n(z) - F_{n+} + o(n^{-1/2}) & \text{if } \theta_1 = 0, \theta_2 = 0, \end{cases}$$

for $p_n : \mathcal{Z} \to [0,1]$ *with* $\mathcal{Z}$ *the support of* Z_i, *and* F_{n+}, F_{n-} *are such that* $\sqrt{n}F_{n+} \to \kappa_+$, $\sqrt{n}F_{n-} \to \kappa_-$. *The* $o(n^{-1/2})$ *terms may depend on* z. *Then*

$$\mathbb{E}[\hat{\theta}_i(1-\theta_i)|Z_i] = \Pr(\theta_i = 0, \hat{\theta}_i = 1, |Z_i) = F_{n+} + o(n^{-1/2}),$$

so $\sqrt{n}\mathbb{E}[\hat{\theta}_i(1-\theta_i)D_{+i}] = \sqrt{n}F_{n+}\mathbb{E}[D_{+i}] + o(1) \to \kappa_+\mathbb{E}[D_{+i}]$. *An analogous result holds for the* D_{-i} *terms.*

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, to understand when the bootstrap methods that follow work, it is helpful to characterize the bias 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 the bias term simplifies to

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

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 η_i are i.i.d. $(0, 1)$ and generated independently from the bootstrap labels $(\theta_i^*, \hat{\theta}_i^*)$, conditionally on the original sample. To understand when these methods are consistent or inconsistent, we use the following set of conditions on $(\theta_i^*, \hat{\theta}_i^*)$.

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

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

(iii) $\mathrm{Var}^*\big(\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{\theta}_i^*(1-\theta_i^*)\big) \to_p 0$ and $\mathrm{Var}^*\big(\frac{1}{\sqrt{n}}\sum_{i=1}^n \theta_i^*(1-\hat{\theta}_i^*)\big) \to_p 0$.

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

Assumption L(i) requires the bootstrap "true" labels θ_i^* to match the estimated labels $\hat{\theta}_i$ on average with probability converging to one. This condition is automatically satisfied if we set $\theta_i^* = \hat{\theta}_i$. Assumption L(ii) is the bootstrap analogue of Assumption 3(iv). It requires the average bootstrap false positive and negative rates to converge in probability to κ_+ and κ_-, respectively, at the $\sqrt{n}$ rate. This assumption (jointly with Assumptions L(iii) and L(iv)) imply that the bootstrap is able to mimic 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 η_i are i.i.d. $(0,1)$ independently from $(\theta_i^*, \hat{\theta}_i^*)$ 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 the bootstrap labels $(\theta_i^*, \hat{\theta}_i^*)$. The first is a standard application of a fixed-regressor bootstrap which sets $\theta_i^* = \hat{\theta}_i$. Its validity requires the assumption that $\mathbb{E}[D_{+i}|\theta_i = 0] = \mathbb{E}[D_{+i}]$ and $\mathbb{E}[D_{-i}|\theta_i = 1] = \mathbb{E}[D_{-i}]$, which is satisfied if D_{+i} and D_{-i} are mean independent of θ_i, but not necessarily otherwise. For this reason, we introduce a coupled-label bootstrap which resamples jointly $(\theta_i^*, \hat{\theta}_i^*)$ independently of the original sample so as 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, conditionally 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 $\sqrt{n}\hat{F}_+ \to_p \kappa_+$ and $\sqrt{n}\hat{F}_- \to_p \kappa_-$, 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 $n, m \to \infty$ with $n/m^2 \to 0$ 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 $\theta_i^* = \hat{\theta}_i$. Although we can easily show that Assumptions L(ii)-(iii) are satisfied provided $\sqrt{n}\hat{F}_+ \to_p \kappa_+$ and $\sqrt{n}\hat{F}_- \to_p \kappa_-$, 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

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

Fixing the "latent" bootstrap labels θ_i^* at $\hat{\theta}_i$ implies that the bootstrap expectation of $\hat{B}^*$ is equal to:

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

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 θ_i. We summarize this result next.

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

3.4.2 Coupled-Label Bootstrap

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

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

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 consequence, the bootstrap false positive rate $\mathbb{E}^*[\hat{\theta}_i^*(1 - \theta_i^*)] = \hat{F}_+$ and the bootstrap false negative rate $\mathbb{E}^*[\theta_i^*(1 - \hat{\theta}_i^*)] = \hat{F}_-$ are both the same for every i. Jointly drawing *pairs* of labels $(\theta_i^*, \hat{\theta}_i^*)$ is crucial to ensuring that the bootstrap correctly

replicates the asymptotic bias B without requiring independence between θ_i and Z_i, as it guarantees that 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 $\mathbb{E}^*[\hat{\theta}_i^*]$. Given the restrictions imposed by setting the discordant pair probabilities equal to $\hat{F}_+$ and $\hat{F}_-$, the key requirement is that $\mathbb{P}^*(\theta_i^* = 1, \hat{\theta}_i^* = 1 | \hat{\theta}_i = 1) \to_p 1$ and $\mathbb{P}^*(\theta_i^* = 1, \hat{\theta}_i^* = 1 | \hat{\theta}_i = 0) \to_p 0$, both of which are satisfied by the proposed distribution. 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 guarantees 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 η_i are i.i.d. $(0,1)$ independently from $(\theta_i^*, \hat{\theta}_i^*)$ such that $\mathbb{E}^*[|\eta_i|^{2+\delta}] < \infty$ for $\delta > 0$. If, in addition, $\sqrt{n}\hat{F}_+ \to_p \kappa_+$ and $\sqrt{n}\hat{F}_- \to_p \kappa_-$, then the coupled-label bootstrap satisfies the conditions of Theorem 2 and, consequently,*

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

as $n \to \infty$, 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,

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

where $Z^* \to_{d^*} N(0, V)$ 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 $\hat{Q}^* = n{-}1\sum_{i=1}^{n}\hat{X}_i^*\hat{X}_i^{*\prime}$ is the bootstrap Hessian matrix. To show that $\hat{b}^* \to_{p^*} b$, we add and subtract appropriately to obtain

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

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, the additional noise from estimating $\hat{F}_+$ and $\hat{F}_-$ may inflate the variance of b^*, and require a finite-sample variance correction to improve coverage of bias-corrected confidence intervals when m/n is small. The same is true here. Fortunately, a similar finite-sample variance correction can easily be integrated into the coupled-label bootstrap. In addition, the $o_{p^*}(1)$ term representing the difference between $\hat{b}^*$ and b^* may not be negligible in finite samples. This term involves the difference between $\hat{Q}^{*-1}$ and $\hat{Q}^{-1}$, which can be substantially different from zero in finite samples (this is particularly true in near-singular designs where the proportion of $\hat{\theta}_i$'s equal to one or zero is small). When this occurs, the bootstrap mean of $\hat{b}^*$ is different from the bootstrap mean of b^*, introducing a location bias in the bootstrap bias. 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-bootstrap, show that the bootstrap is valid, then show it is able to replicate the two most important highest-order variance terms. We conclude by describing the rotation that solves 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 $i = 1, \ldots, n$, independently across i and independently of $u_i^* = \hat{u}_i \eta_i$, according to the following categorical distribution:

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

and

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

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 $i = 1, \ldots, n$.

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 η_i are i.i.d. $(0, 1)$ independently from $(\theta_i^*, \hat{\theta}_i^*)$ such that $\mathbb{E}^*[|\eta_i|^{2+\delta}] < \infty$ for $\delta > 0$. If, in addition, $\sqrt{n}\hat{F}_+ \to_p \kappa_+$ and $\sqrt{n}\hat{F}_- \to_p \kappa_-$ and $n/m^2 \to 0$, then the coupled-label bootstrap with variance correction satisfies the conditions of Theorem 2 and, consequently,*

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

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

Theorem 4 shows that the coupled-label bootstrap with variance correction is able to reproduce the asymptotic bias of the OLS estimator and the asymptotic variance term V. We now show it is able to match key higher-order 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 $\hat{Q} = \frac{1}{n}\sum_{i=1}^n \hat{X}_i \hat{X}_i'$, $\hat{\Gamma}_+ = \hat{Q}^{-1}\left(\frac{1}{n}\sum_{i=1}^n D_{+i}\right)$, 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 $\mathrm{Var}(\sqrt{n}(\hat{\beta} - \beta)) \approx V$, we have

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

Moreover,

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

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

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

Under the framework adopted in BCHS where $m/n \to 0$ with $n/m^2 \to 0$, 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 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

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

Therefore, drawing $\hat{F}_+^*$ and $\hat{F}_+^*$ from a distribution centered at $\hat{F}_+$ and $\hat{F}_-$ doesn't 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

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

and

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

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

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

Moreover,

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

from which we see that

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

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

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

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

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

These 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

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

where $\hat{X}^*$ and Y^* are obtained by the variance-corrected coupled-label bootstrap. With the 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 $\hat{R}^* \to_{p^*} I_k$, 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 adapting the simulation design of BCHS. We consider an interactions model in which a standard normal covariate Z_i is interacted with the binary label θ_i. We allow $p_i = \mathbb{P}(\theta_i = 1)$ to vary with Z_i to introduces correlation between θ_i and the matrices D_{+i} and D_{-i}, which, as shown above, invalidates the fixed-label bootstrap. The regression model is:

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

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 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 those used by BCHS. The correlation between θ_i and Z_i^2 depends strongly on $\bar{p}$: it is approximately 0.11 when $\bar{p} = 0.05$ but rises to approximately 0.57 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 θ_i and $(D_{+i},\, D_{-i})$.

The bootstrap DGP takes the form

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

where $\hat{\beta}$ is the 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. 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. Sample sizes are $n \in \{8{,}000;\ 16{,}000;\ 32{,}000\}$, misclassification rates $\kappa \in \{0.5,\ 1,\ 1.5\}$, and the external sample has size $m = 1{,}000$.

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 five estimators: OLS using the imputed regressor $\hat{\theta}_i Z_i$, the feasible bias-corrected estimator of BCHS, and the four bootstrap methods.

The use of an imputed label biases OLS downward, with the bias increasing in κ. The resulting undercoverage is severe: even in the mildest configuration ($n = 8{,}000$, $\kappa = 0.5$), OLS coverage is only 11.2%. The BCHS bias-corrected estimator removes virtually all the bias and delivers reliable inference across configurations, though coverage deteriorates modestly as κ 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 11.5%).

The fixed-label bootstrap overcorrects: it introduces bias in the opposite direction and this bias converges to zero very slowly with n. 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

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

	Median bias			Coverage (%)			Interval length		
$\kappa_+ = \kappa_- =$	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5
$n = 8{,}000$									
OLS	−0.04	−0.08	−0.12	11.2	0.0	0.0	0.05	0.06	0.07
BCHS bias-corrected	0.00	−0.01	−0.02	94.7	94.0	91.9	0.08	0.11	0.12
No-label resampling	−0.04	−0.08	−0.12	11.5	0.0	0.0	0.05	0.06	0.07
Fixed-label bootstrap	0.02	0.03	0.03	67.4	61.0	62.9	0.07	0.08	0.08
Coupled-label bootstrap	0.00	−0.01	−0.03	87.8	81.1	66.3	0.06	0.07	0.08
Coupled-label, rotation & var. adj.	0.00	−0.01	−0.02	95.6	95.1	92.9	0.09	0.11	0.12
$n = 16{,}000$									
OLS	−0.03	−0.06	−0.09	8.9	0.0	0.0	0.04	0.04	0.05
BCHS bias-corrected	0.00	−0.01	−0.01	93.8	93.4	92.4	0.06	0.08	0.10
No-label resampling	−0.03	−0.06	−0.09	8.6	0.0	0.0	0.04	0.04	0.05
Fixed-label bootstrap	0.02	0.02	0.03	56.6	48.3	45.3	0.05	0.05	0.06
Coupled-label bootstrap	0.00	−0.01	−0.01	82.5	76.8	68.8	0.04	0.05	0.05
Coupled-label, rotation & var. adj.	0.00	0.00	−0.01	94.5	94.3	93.1	0.06	0.08	0.09
$n = 32{,}000$									
OLS	−0.02	−0.04	−0.06	6.1	0.0	0.0	0.02	0.03	0.03
BCHS bias-corrected	0.00	0.00	0.00	92.2	92.6	92.6	0.05	0.06	0.08
No-label resampling	−0.02	−0.04	−0.06	6.5	0.0	0.0	0.02	0.03	0.03
Fixed-label bootstrap	0.01	0.02	0.02	49.5	39.2	36.1	0.03	0.03	0.04
Coupled-label bootstrap	0.00	0.00	−0.01	74.7	68.4	63.3	0.03	0.03	0.03
Coupled-label, rotation & var. adj.	0.00	0.00	−0.01	92.8	93.5	93.4	0.05	0.06	0.08

Note: 10,000 Monte Carlo replications; $B = 499$ bootstrap replications; external sample $m = 1{,}000$.

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 in several configurations, while maintaining competitive 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 misclassified when $\kappa \geq 1$. The OLS bias is roughly twice as large as in the $\bar{p} = 0.5$ case. Because the correlation between θ_i and (D_{+i}, D_{-i}) is low, the fixed-label bootstrap performs well and comparably to the coupled-label bootstrap. Both maintain coverage below the BCHS benchmark. However, the variance correction and Hessian rotation improve coverage, and the gap with the analytic correction from BCHS decreases

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

	Median bias			Coverage (%)			Interval length		
$\kappa_+ = \kappa_- =$	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5
$n = 8{,}000$									
OLS	−0.08	−0.16	−0.25	15.0	0.0	0.0	0.10	0.12	0.14
BCHS bias-corrected	−0.01	−0.03	−0.06	94.8	91.8	84.4	0.16	0.22	0.27
No-label resampling	−0.08	−0.16	−0.25	15.6	0.0	0.0	0.10	0.12	0.14
Fixed-label bootstrap	0.00	0.01	−0.03	87.2	83.0	73.0	0.12	0.15	0.17
Coupled-label bootstrap	0.00	−0.01	−0.03	86.5	82.4	72.9	0.12	0.15	0.17
Coupled-label, rotation & var. adj.	0.00	−0.02	−0.06	94.4	90.2	78.0	0.16	0.20	0.22
$n = 16{,}000$									
OLS	−0.05	−0.11	−0.17	13.6	0.0	0.0	0.07	0.08	0.09
BCHS bias-corrected	0.00	−0.01	−0.03	93.2	92.2	89.1	0.12	0.16	0.20
No-label resampling	−0.05	−0.11	−0.17	13.3	0.0	0.0	0.07	0.08	0.09
Fixed-label bootstrap	0.00	0.00	−0.01	80.6	77.7	74.1	0.08	0.09	0.11
Coupled-label bootstrap	0.00	−0.01	−0.01	79.8	77.6	72.9	0.08	0.09	0.11
Coupled-label, rotation & var. adj.	0.00	−0.01	−0.03	93.1	91.8	87.6	0.12	0.15	0.18
$n = 32{,}000$									
OLS	−0.04	−0.07	−0.12	10.5	0.0	0.0	0.04	0.05	0.06
BCHS bias-corrected	0.00	−0.01	−0.02	91.4	92.1	91.3	0.09	0.13	0.15
No-label resampling	−0.04	−0.08	−0.12	10.1	0.0	0.0	0.04	0.05	0.06
Fixed-label bootstrap	0.00	0.00	0.00	76.1	66.6	65.8	0.05	0.05	0.07
Coupled-label bootstrap	0.00	0.00	−0.01	74.6	65.7	63.9	0.05	0.06	0.07
Coupled-label, rotation & var. adj.	0.00	−0.01	−0.02	91.5	92.5	91.1	0.09	0.12	0.14

Note: See notes to Table 1. With $\bar{p} = 0.05$, the correlation between θ_i and (D_{+i}, D_{-i}) is low (≈ 0.11), so the fixed-label bootstrap performs comparably to the coupled-label bootstrap.

with sample size. Nevertheless, our findings are otherwise qualitatively similar to the $\bar{p} = 0.5$ case.

We conclude that the coupled-label bootstrap with finite-sample modifications is the preferred method in all situations. It delivers coverage close to nominal while producing confidence intervals of competitive length unless $\bar{p}$ and n are small and κ is large.

6 Application

TBA

7 Conclusion

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

gression 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 Quasilikelihood 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

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

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

$$n^{-1}\sum_{i=1}^n \hat{X}_i^*\hat{X}_i^{*\prime} = n^{-1}\sum_{i=1}^n \hat{X}_i\hat{X}_i' + (a_1^*) + (b_1^*) + (c_1^*) + (a_2^*) + (b_2^*) + (c_2^*) + (a_3^*) + (b_3^*)$$

$$(a_1^*) = \frac{1}{n}\sum_{i=1}^n \hat{X}_i(X_i^* - \hat{X}_i)',\ (b_1^*) = (a_1^*)',$$

$$(c_1^*) = \frac{1}{n}\sum_{i=1}^n (X_i^* - \hat{X}_i)(X_i^* - \hat{X}_i)',$$

$$(a_2^*) = \frac{1}{n}\sum_{i=1}^n \hat{X}_i(\hat{X}_i^* - X_i^*)',\ (b_2^*) = (a_2^*)',$$

$$(c_2^*) = \frac{1}{n}\sum_{i=1}^n (\hat{X}_i^* - X_i^*)(\hat{X}_i^* - X_i^*)',$$

$$(a_3^*) = \frac{1}{n}\sum_{i=1}^n (X_i^* - \hat{X}_i)(\hat{X}_i^* - X_i^*)',\ (b_3^*) = (a_3^*)'.$$

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, $n^{-1}\sum_{i=1}^n \hat{X}_i^*\hat{X}_i^{*\prime} \to_{p^*} Q$. Since

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

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

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

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

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

We now use Markov's inequality, noting that

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

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 $\sqrt{n}\mathbb{E}[(\hat{X}_i - X_i)u_i] \to 0$. Then

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

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

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

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

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

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

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

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

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

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 η_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

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

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

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

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 $\mathrm{Var}^*(n^{-1/2} \sum_{i=1}^n X_i u_i \eta_i) = n^{-1} \sum_{i=1}^n X_i X_i' u_i^2 \to_p \Sigma \equiv \mathbb{E}[X_i X_i' u_i^2]$ under Assumption 3(i)–(ii). Since, conditionally 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

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

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

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

provided $\mathbb{E}[\|X_i\|^{2(2+\delta)}] < \infty$ and $\mathbb{E}[|u_i|^{2(2+\delta)}] < \infty$, which is implied by Assump-

tion 3(i)–(ii) with a suitable redefinition of δ above. This proves (14).

Next, we show that the second term in (13) is asymptotically negligible, i.e., $n^{-1/2}\sum_{i=1}^n X_i(\hat{u}_i - u_i)\eta_i \to_{p^*} 0$. 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

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

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

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

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

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

where the first factor is $O_p(1)$ if $\mathbb{E}[\|X_i\|^4] < \infty$ andthe 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 $p^{-1} + q^{-1} = 1$,

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

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 $\mathcal{Z}_{2n}^* \to_{p^*} 0$. Since $\mathbb{E}^*[u_i^*] = 0$, we have $\mathbb{E}^*[\mathcal{Z}_{2n}^*] = 0$ and the result follows by showing that $\mathrm{Var}^*(\mathcal{Z}_{2n}^*) \to_p 0$. Using the fact that $u_i^* = \hat{u}_i\eta_i$ with η_i i.i.d. $(0, 1)$ yields

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

whose Frobenius norm is bounded by

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

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). □

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

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

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

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 η_i are i.i.d. $(0,1)$ independently from $(\theta_i^*, \hat{\theta}_i^*)$ and assume that θ_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

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

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) verify 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 $\mathbb{E}^*[(\hat{X}_i^* - X_i^*)u_i^*] = \mathbb{E}^*[(\hat{X}_i^* - X_i^*)]\mathbb{E}^*[u_i^*] = 0$ given $u_i^* = \hat{u}_i\eta_i$, where η_i is independent of $\hat{u}_i$ and $\hat{X}_i^* - X_i^*$, conditionally on the sample, and given that $\mathbb{E}^*[\eta_i] = 0$. Hence, it suffices to show that $\mathrm{Var}^*(n^{-1/2}\sum_{i=1}^n(\hat{X}_i^* - X_i^*)u_i^*) \to_p 0$. To verify this condition, we can use the independence between η_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

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

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

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

for any $p, q > 1$ such that $p^{-1} + q^{-1} = 1$. Note $\big(\mathbb{E}^*[\mathbb{I}(\hat{\theta}_i^* \neq \theta_i^*)]\big)^p \leq \mathbb{E}^*[\mathbb{I}(\hat{\theta}_i^* \neq \theta_i^*)]$ 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:

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

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,

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

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 $|\hat{u}_i|^{4+\delta} \leq 2^{3+\delta}(|u_i|^{4+\delta} + |\hat{u}_i - u_i|^{4+\delta})$, 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). □

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

Lemma A.3. *Suppose that Assumption 3 holds. If θ_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

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

with B given Lemma 1. Adding and subtracting appropriately yields $B_n^* = B_{1n}^* + B_{2n}^*$, where

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

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

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). □

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

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

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

where we used $\sqrt{n}\hat{F}_+ \to_p \kappa_+$ and $\sqrt{n}\hat{F}_- \to_p \kappa_-$. 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 i.i.d. Bernoulli$(\hat{F}_+^*)$

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

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

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

For part (iii), consider $\mathrm{Var}^*\big(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\theta}_i^*(1-\theta_i^*)\big)$. Again using the fact that $\hat{\theta}_i^*(1-\theta_i^*)$ are i.i.d. Bernoulli($\hat{F}_+^*$) conditional on the data and $\hat{F}_+^*$, we have

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

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

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

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

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

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

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

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

where the first factor is because $\sqrt{n}\hat{F}_+ \to_p \kappa_+$ 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 i.i.d. Bernoulli($\hat{F}_+^*$), we have

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

and, assuming D_{+i} is scalar (otherwise, the argument applies to a generic element of this matrix),

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

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

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

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

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

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

proof that $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \theta_i^*(1-\hat{\theta}_i^*)\big(D_{-i} - \mathbb{E}[D_{-i}]\big) \to_{p^*} 0$ is analogous. $\square$