A Implementation Details for the BISG Adaptation of the Dual-Bootstrap

We outline our approach to a few issues that can arise when implementing our dual-bootstrap adaptation to BISG.

A.1 Zero Counts

In the ACS dataset, some geographic areas are estimated to have zero people of certain races. Intuitively, such "zero counts" are more common at the census block group level than at the ZIP code tabulation area level. In such cases, all 80 variance replicates also estimate zero people of that race, even though a full count may have revealed people of that race in that area. To nonetheless reflect this uncertainty, the Census Bureau reports a margin of error based on a different method from the successive differences replication method.

Whenever we encounter such instances of zero counts, we impute possible observed values for the 80 variance replicates from a discrete uniform distribution with minimum value 0 and maximum value determined by the margin of error. In particular, we derive the estimated variance from the reported margin of error using a formula prescribed by the Census Bureau. The variance, combined with the fact that observed counts can never be less than 0, allows us to derive the maximum possible value of the discrete uniform distribution. We sample from this distribution 80 times independently and replace the zero counts in the variance replicates with these values (and update the estimated total counts across all races in the variance replicates) for purposes of estimating the covariance matrix $\hat{\Sigma}$.

Our choice of parametric distribution here has little downstream impact since it is used only to recover the variance, which maps directly to the margin of error given by the Census Bureau. But our estimates for zero-count races in a geolocation have approximately zero covariance with the estimates for other races even though this might not be the case in reality. We leave an examination of the significance of this choice and possible improvements to it to future work.

A.2 Impermissible Sampled Probabilities

In some cases, the draws of vectors $\widehat{\Pr}_{\mathbb{P}}^{*b}(A \mid G = g) \sim \mathcal{N}\left(\hat{\mu}_g, \hat{\Sigma}_g\right)$ will include elements that are less than 0 or greater than 1. This arises because we assume that the sampling distribution of the conditional race-by-geolocation probability estimates is multivariate normal. In such instances, we simply round the elements to 0 or 1 accordingly. The rounding to 1 is not strictly necessary, since the normalization that occurs in Bayes' Theorem implicitly handles it; but the rounding to 0 appears to be necessary.

This problem likely can be avoided by imposing an alternative form on the sampling distribution. For example, there might exist a unique set of parameters that best fit $\hat{\mu}_g$ and $\hat{\Sigma}_g$ as a Dirichlet distribution. If so, then modeling the sampling distribution as a Dirichlet with those parameters instead would sidestep this issue. We also note here that the densities of the Dirichlet distribution and the multivariate normal distribution with the same means and covariances converge asymptotically (Ouimet, 2022). This might suggest that this problem is less significant in relatively large sample sizes like the ACS, but more research is needed to be sure.

A.3 Mutually Exclusive Conditional Probabilities

In some cases, the conditional surname-by-race probabilities and the conditional race-by-geolocation probabilities are incompatible. For example, an individual might have a surname that, according to the 2010 surname table, only White people have. But he or she might live in a census block group that, according to ACS estimates, has no White people. This problem is not unique to the dual-bootstrap; it can occur in any application of BISG. But it is more likely to occur when applying the dual-bootstrap, which calls for repeated computation of BISG probabilities on resamples of the training data.

Because this problem extends beyond the dual-bootstrap, we do not propose any particular solution. For purposes of the simulations in Section 5, however, our stopgap approach is to give primacy to the surname probabilities: If the conditional surname-by-race probabilities and the conditional race-by-geolocation probabilities are incompatible, we simply do not update the former with the latter.

B Additional Details on BISG Simulations

B.1 Additional State-by-State Results

Figure 7 shows the results of the state-by-state simulation in Section 5.2 for the remaining race categories that we study. Although the trends shown here are less pronounced, we believe they can be interpreted within the framework described in Section 5.2.

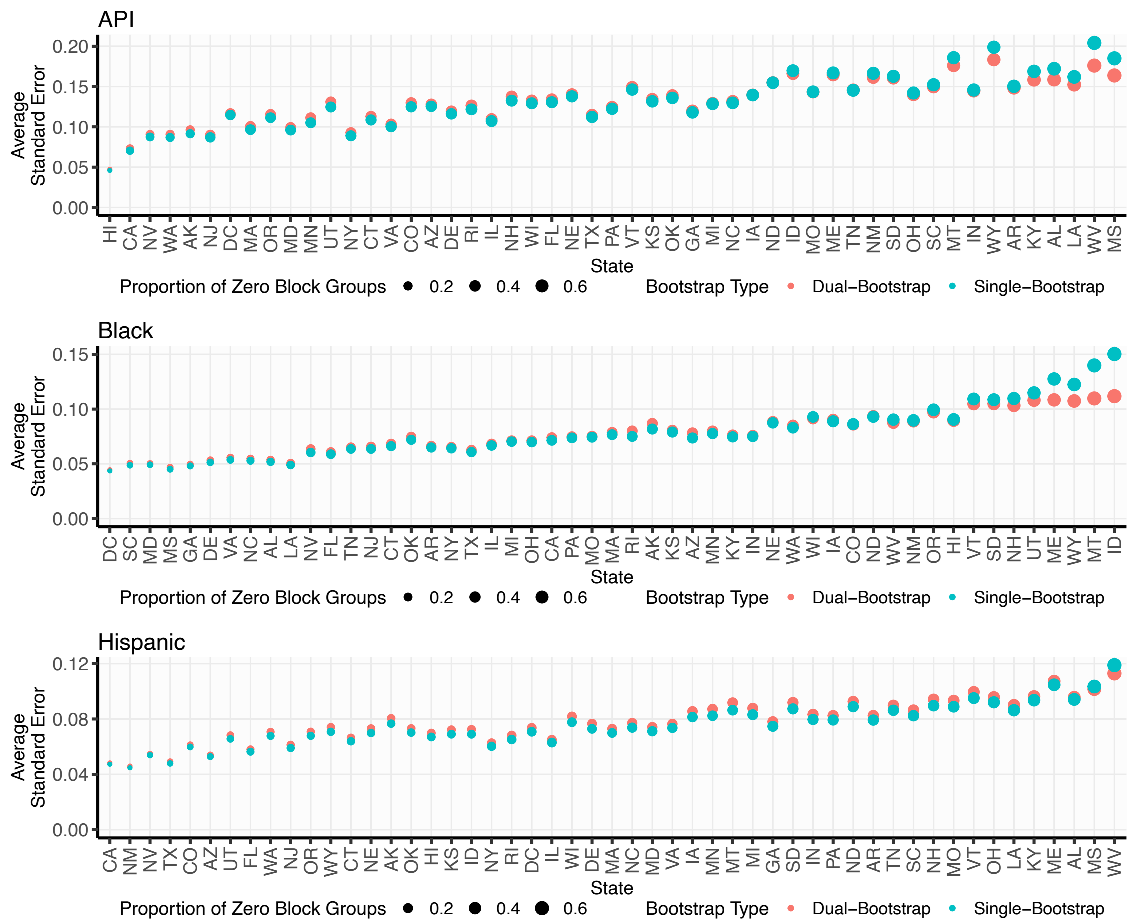

Figure 7: Dual-bootstrap and single-bootstrap standard errors of the estimated average outcome for the Asian and Pacific Islander (API), Black, and Hispanic race groups in each state. The states are ordered by the proportion of census block groups in which the American Community Survey estimates there are zero people of the given race.

B.2 Additional Details on New Mexico Simulation

In this section, we describe in more detail the New Mexico simulation reported in Section 5.2. We select New Mexico for illustrative purposes and focus on the effect of varying total size and census block group concentration of the American Indian and Alaska Native (AIAN) population in the state on the standard error of the average group outcome estimate. The following simulation can be conducted for any state and any race group.

The simulation follows the same general procedure as the state-by-state simulation in Section 5.2 for just New Mexico, except we modify the 2017-2021 ACS 5-year estimates of the AIAN composition of each census block group in the state. The actual estimated total population of AIAN in New Mexico is 181,021, and about 49% of census block groups are estimated to have zero AIAN people. We vary the proportion of census block groups with zero-count AIAN from 30% to 80%, and also vary the total population of AIAN among the values 50,000, 100,000, 200,000, and 400,000.

When modifying the proportion of census block groups with zero-count AIAN, we start with the existing distribution of AIAN counts in census block groups. We decrease the proportion of zero-count AIAN census block groups by randomly selecting zero-count AIAN census block groups (without replacement) and assigning them all the AIAN information (including ACS margins of error and variance replicate estimates) of randomly selected non-zero-count AIAN census block groups (with replacement). Similarly, we increase the proportion of zero-count AIAN census block groups by randomly selecting non-zero-count AIAN census block groups (without replacement) and assigning them the AIAN information of zero-count AIAN census block groups (with replacement). To then achieve the desired total population of AIAN, we scale all non-zero counts (and margins of error) of AIAN proportionally up or down. The result is that each synthetically generated New Mexico has different total sizes and census block group concentrations of AIAN, but a similarly shaped distribution of AIAN among non-zero-count AIAN census block groups.

Having modified the records of the ACS table that report AIAN information, we simply follow through with the rest of the procedure described in Section 5 and Appendix A, including estimating the covariance matrix with zero-count adjustments, implementing Algorithm 2, and conducting the standard error estimation simulation with 1,000 tuples in our synthetically generated New Mexico.

C Asymptotic Normality of the Dual-Bootstrap

C.1 Proof of Asymptotic Normality for Logistic Regression

We prove that the dual-bootstrap produces asymptotically normal bootstrap statistics with properly calibrated variance under the simplifying assumption that the race probabilities obey the logistic regression model

$$\Pr(A = 1 \mid Z) = \frac{\exp(\alpha^\intercal Z)}{1 + \exp(\alpha^\intercal Z)}. \tag{1}$$

The proof can be readily extended to other race probability models that fall within the Z-estimation framework. The theorem may also hold for other race probability models as well, but we leave a proof of such results to future work.

For ease of notation, let $\mu_a \equiv \mathbb{E}[Y \mid A = a]$ for $a \in \{0, 1\}$.

Let

$$\psi_\theta(z, a, y) \equiv \begin{bmatrix} \psi_\alpha(z,a,y) \\ \psi_1(z,a,y) \\ \psi_0(z,a,y) \end{bmatrix} \equiv \begin{bmatrix} z\left\{a - \frac{\exp(\theta_\alpha^\intercal z)}{1+\exp(\theta_\alpha^\intercal z)}\right\} \\ \frac{\exp(\theta_\alpha^\intercal z)}{1+\exp(\theta_\alpha^\intercal z)} y - \frac{\exp(\theta_\alpha^\intercal z)}{1+\exp(\theta_\alpha^\intercal z)}\theta_1 \\ \frac{1}{1+\exp(\theta_\alpha^\intercal z)} y - \frac{1}{1+\exp(\theta_\alpha^\intercal z)}\theta_0, \end{bmatrix} \tag{2}$$

and assume that $\theta \equiv \begin{bmatrix}\theta_\alpha & \theta_1 & \theta_0\end{bmatrix}^\intercal \in \Theta \subset \mathbb{R}^p$ where Θ is open and $p < \infty$ is fixed.

Then, defining the map $\theta \mapsto \Psi(\theta) \equiv P\psi_\theta$, note that $\theta_0 \equiv \begin{bmatrix}\alpha & \mu_1 & \mu_0\end{bmatrix}^\intercal$ satisfies $\Psi(\theta_0) = 0$. We show this coordinate by coordinate. First,

$$\begin{aligned} \mathbb{E}\left[Z\left\{A - \frac{\exp(\alpha^\intercal Z)}{1+\exp(\alpha^\intercal Z)}\right\}\right] &= \mathbb{E}[ZA] - \mathbb{E}[Z\Pr(A=1 \mid Z)] & (3)\\ &= \mathbb{E}[Z\Pr(A=1 \mid Z)] - \mathbb{E}[Z\Pr(A=1 \mid Z)] & (4)\\ &= 0, & (5) \end{aligned}$$

where (4) follows from the tower property conditioning on Z and the fact that A is binary. Second,

$$\begin{aligned} \mathbb{E}\left[\frac{\exp(\alpha^\intercal Z)}{1+\exp(\alpha^\intercal Z)}Y - \frac{\exp(\alpha^\intercal Z)}{1+\exp(\alpha^\intercal Z)}\mu_1\right] &= \mathbb{E}[\Pr(A=1 \mid Z)Y] - \mathbb{E}[\Pr(A=1 \mid Z)]\,\mu_1 & (6)\\ &= \mathbb{E}[\Pr(A=1 \mid Z)Y] - \mathbb{E}[\Pr(A=1 \mid Z)]\,\frac{\mathbb{E}(AY)}{\Pr(A=1)} & (7)\\ &= \mathbb{E}[\Pr(A=1 \mid Z)Y] - \mathbb{E}(AY) & (8)\\ &= \mathbb{E}[\Pr(A=1 \mid Z)\mathbb{E}(Y \mid Z)] - \mathbb{E}\{\mathbb{E}(AY \mid Z)\} & (9)\\ &= \mathbb{E}[\mathbb{E}(A \mid Z)\mathbb{E}(Y \mid Z) - \mathbb{E}(AY \mid Z)] & (10)\\ &= 0, & (11) \end{aligned}$$

where (7), (8), and (9) follow from the law of total expectation, and (11) follows from our identifying assumption of zero covariance. The proof of the third coordinate is analogous.

Let $\hat{\theta}_n$ be an approximate zero of the estimating equation $\theta \mapsto \Psi_n(\theta) \equiv \mathbb{P}_n \psi_\theta$, and let $\hat{\theta}^*_n$ be an approximate zero of the bootstrapped estimating equation $\theta \mapsto \Psi^*_n(\theta) \equiv \mathbb{P}^*_n \psi_\theta$. By Theorem 10.16 of Kosorok (2008),

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} Z \sim \mathcal{N}\left(0, V_{\theta_0}^{-1} P\left[\psi_{\theta_0}\psi_{\theta_0}^\intercal\right]\left(V_{\theta_0}^{-1}\right)^\intercal\right) \tag{12}$$

and

$$\sqrt{n}\left(\hat{\theta}^*_n - \hat{\theta}_n\right) \underset{*}{\overset{P}{\rightsquigarrow}} k_0 Z \tag{13}$$

if five conditions hold. We verify each condition in turn. As a preliminary matter, note that Exercise 10.5.5 of Kosorok (2008) already verifies each of the five conditions for the first coordinate of ψ_θ. So we verify them for the remaining two coordinates, focusing without loss of generality on the first of the two.

(A) For any sequence $\{\theta_n\} \in \Theta$, $\Psi(\theta_n) \to 0$ implies $\|\theta_n - \theta_0\| \to 0$.

Proof. By assumption,

$$\mathbb{E}\left[\frac{\exp(\theta_{\alpha n}^\intercal Z)}{1+\exp(\theta_{\alpha n}^\intercal Z)} Y - \frac{\exp(\theta_{\alpha n}^\intercal Z)}{1+\exp(\theta_{\alpha n}^\intercal Z)}\theta_{1n}\right] \to 0. \tag{14}$$

Distributing the expectation and dividing by both sides yields

$$\frac{\mathbb{E}\left[\frac{\exp(\theta_{\alpha n}^\intercal Z)}{1+\exp(\theta_{\alpha n}^\intercal Z)} Y\right]}{\mathbb{E}\left[\frac{\exp(\theta_{\alpha n}^\intercal Z)}{1+\exp(\theta_{\alpha n}^\intercal Z)}\right]} - \theta_{1n} \to 0, \tag{15}$$

so it suffices to show that

$$\frac{\mathbb{E}\left[\frac{\exp(\theta_{\alpha n}^\intercal Z)}{1+\exp(\theta_{\alpha n}^\intercal Z)} Y\right]}{\mathbb{E}\left[\frac{\exp(\theta_{\alpha n}^\intercal Z)}{1+\exp(\theta_{\alpha n}^\intercal Z)}\right]} \to \mu_1. \tag{16}$$

Since we know from (1) and (6) that

$$\mu_1 = \frac{\mathbb{E}\left[\frac{\exp(\alpha^\intercal Z)}{1+\exp(\alpha^\intercal Z)} Y\right]}{\mathbb{E}\left[\frac{\exp(\alpha^\intercal Z)}{1+\exp(\alpha^\intercal Z)}\right]}, \tag{17}$$

it suffices to prove that the numerator and denominator of (16) each converge to their corresponding limit. We prove the denominator first. From Example 10.5.5 of Kosorok (2008), we can take as given that $\theta_{\alpha n} \to \alpha$. Since these are constants, this implies that $\theta_{\alpha n} \xrightarrow{p} \alpha$. Moreover, it is trivially true that an i.i.d. sequence $Z_1, Z_2, \ldots$ where each Z_i is distributed as Z satisfies $Z_n \xrightarrow{d} Z$. Then Slutsky's theorem implies that $\theta_{\alpha n} \odot Z \xrightarrow{d} \alpha \odot Z$. The continuous

mapping theorem then implies that $\theta_{\alpha n}^{\intercal} Z \xrightarrow{d} \alpha^{\intercal} Z$. Since the logistic function is bounded and continuous, convergence in distribution implies that

$$\mathbb{E}\left[\frac{\exp(\theta_{\alpha n}^{\intercal} Z)}{1+\exp(\theta_{\alpha n}^{\intercal} Z)}\right] \to \mathbb{E}\left[\frac{\exp(\alpha^{\intercal} Z)}{1+\exp(\alpha^{\intercal} Z)}\right]. \tag{18}$$

The proof for the numerator is similar but slightly more delicate. Again from Example 10.5.5 of Kosorok (2008), we can take as given that $\theta_{\alpha n} \to \alpha$. Since these are constants, this implies that $\theta_{\alpha n} \xrightarrow{p} \alpha$. Moreover, it is trivially true that an i.i.d. sequence $(Z_1, Y_1), (Z_2, Y_2), \ldots$ where each (Z_i, Y_i) is distributed as (Z, Y) satisfies $(Z_n, Y_n) \xrightarrow{d} (Z, Y)$. Since α is a constant, the portmanteau lemma implies that $(Z_n, Y_n, \theta_{\alpha n}) \xrightarrow{d} (Z, Y, \alpha)$. Then the continuous mapping theorem implies that

$$\frac{\exp(\theta_{\alpha n}^{\intercal} Z)}{1+\exp(\theta_{\alpha n}^{\intercal} Z)} Y \xrightarrow{d} \frac{\exp(\alpha^{\intercal} Z)}{1+\exp(\alpha^{\intercal} Z)} Y. \tag{19}$$

If Y has a finite second moment, then

$$\mathbb{E}\left[\left|\frac{\exp(\theta_{\alpha n}^{\intercal} Z)}{1+\exp(\theta_{\alpha n}^{\intercal} Z)} Y\right|^2\right] < \mathbb{E}\left[|Y|^2\right] < \infty \tag{20}$$

for all $n \in \mathbb{N}$, so

$$\frac{\exp(\theta_{\alpha n}^{\intercal} Z)}{1+\exp(\theta_{\alpha n}^{\intercal} Z)} Y \tag{21}$$

is uniformly integrable.[8] This, combined with convergence in distribution, implies that

$$\mathbb{E}\left[\frac{\exp(\theta_{\alpha n}^{\intercal} Z)}{1+\exp(\theta_{\alpha n}^{\intercal} Z)} Y\right] \to \mathbb{E}\left[\frac{\exp(\alpha^{\intercal} Z)}{1+\exp(\alpha^{\intercal} Z)} Y\right]. \tag{22}$$

Thus, the proof is complete. □

(B) The class $\{\psi_\theta : \theta \in \Theta\}$ is strong Glivenko-Cantelli.

Proof. As indicated by Van der Vaart (2000), it suffices to show separately that each coordinate class is strong Glivenko-Cantelli. This can be done under several different regularity conditions. We assume two regularity conditions. First, we assume that each coordinate of (Z, A, Y) is bounded almost surely—i.e., that $(Z, A, Y) \sim P$ where P has measure zero outside a bounded subset of $\mathbb{R}^{p+2}$. Second, we assume that Θ is bounded. Then let $\mathbb{R}^{p+2} = \cup_j I_j$ be a partition in cubes of volume 1. Since each ψ_1 in the class has partial derivatives up to order $\alpha > (p+2)/2$ that are bounded by constants M_j on each of the cubes I_j, Example 19.9 of Van der Vaart (2000) guarantees that, for any $V \geq (p+2)/\alpha$,

$$\log N_{[]}\left(\epsilon, \{\psi_\theta : \theta \in \Theta\}, L_2(P)\right) \leq K \left(\frac{1}{\epsilon}\right)^V \left(\sum_{j=1}^{\infty} \left(M_j^2 P(I_j)\right)^{\frac{V}{V+2}}\right)^{\frac{V+2}{2}}. \tag{23}$$

[8] https://www.randomservices.org/random/expect/Uniform.html

Since P has measure zero outside a bounded subset of $\mathbb{R}^{p+2}$, the series converges for any $V \geq (p+2)/\alpha$. Then, setting $V \in [(p+2)/\alpha, 2)$, we see that the function class has finite bracketing integral:

$$
\begin{aligned}
J_{[]}\left(1, \{\psi_\theta : \theta \in \Theta\}, L_2(P)\right) &\equiv \int_0^1 \sqrt{\log N_{[]}\left(\epsilon, \{\psi_\theta : \theta \in \Theta\}, L_2(P)\right)}\, d\epsilon && (24)\\
&\leq \int_0^1 \sqrt{K\left(\frac{1}{\epsilon}\right)^V \left(\sum_{j=1}^{\infty}\left(M_j^2 P(I_j)\right)^{\frac{V}{V+2}}\right)^{\frac{V+2}{2}}}\, d\epsilon && (25)\\
&= \sqrt{K\left(\sum_{j=1}^{\infty}\left(M_j^2 P(I_j)\right)^{\frac{V}{V+2}}\right)^{\frac{V+2}{2}}} \int_0^1 \sqrt{\left(\frac{1}{\epsilon}\right)^V}\, d\epsilon && (26)\\
&\leq C \int_0^1 \left(\frac{1}{\epsilon}\right)^{\frac{V}{2}} d\epsilon && (27)\\
&< \infty, && (28)
\end{aligned}
$$

where C is a constant. Thus, by Theorem 19.5 of Van der Vaart (2000), the function class is Donsker. Hence, it is strong Glivenko-Cantelli (Kosorok, 2008). Note that we can probably relax the assumption that P has measure zero if we instead assume it has a certain concentration. □

(C) For some $\eta > 0$, the class $\mathcal{F} \equiv \{\psi_\theta : \theta \in \Theta, \|\theta - \theta_0\| \leq \eta\}$ is Donsker and $P(\psi_\theta - \psi_{\theta_0})^2 \to 0$ as $\|\theta - \theta_0\| \to 0$.

Proof. The first statement follows immediately from our proof of (B). The second statement follows from similar logic to the proof of (A), assuming that $\mathbb{E}\left[|Y|^4\right] < \infty$. Observe that

$$
P(\psi_\theta - \psi_{\theta_0})^2 = \mathbb{E}\left[\left\{\frac{\exp(\theta_\alpha^\intercal Z)}{1+\exp(\theta_\alpha^\intercal Z)}(Y - \theta_1) - \frac{\exp(\alpha^\intercal Z)}{1+\exp(\alpha^\intercal Z)}(Y - \mu_1)\right\}^2\right]. \quad (29)
$$

Consider the first outer term obtained by squaring the inside:

$$
\begin{aligned}
&\mathbb{E}\left[\left\{\frac{\exp(\theta_\alpha^\intercal Z)}{1+\exp(\theta_\alpha^\intercal Z)}(Y - \theta_1)\right\}^2\right] && (30)\\
&= \mathbb{E}\left[\left\{\frac{\exp(\theta_\alpha^\intercal Z)}{1+\exp(\theta_\alpha^\intercal Z)}\right\}^2 Y^2\right] - 2\theta_1 \mathbb{E}\left[\left\{\frac{\exp(\theta_\alpha^\intercal Z)}{1+\exp(\theta_\alpha^\intercal Z)}\right\}^2 Y\right] + \theta_1^2 \mathbb{E}\left[\left\{\frac{\exp(\theta_\alpha^\intercal Z)}{1+\exp(\theta_\alpha^\intercal Z)}\right\}^2\right] && (31)\\
&\to \mathbb{E}\left[\left\{\frac{\exp(\alpha^\intercal Z)}{1+\exp(\alpha^\intercal Z)}\right\}^2 Y^2\right] - 2\mu_1 \mathbb{E}\left[\left\{\frac{\exp(\alpha^\intercal Z)}{1+\exp(\alpha^\intercal Z)}\right\}^2 Y\right] + \mu_1^2 \mathbb{E}\left[\left\{\frac{\exp(\alpha^\intercal Z)}{1+\exp(\alpha^\intercal Z)}\right\}^2\right], && (32)
\end{aligned}
$$

where convergence occurs by application of the continuous mapping theorem and uniform integrability (again, assuming that the fourth moment of Y is finite) to the fact that $(\theta_\alpha, \theta_1) \to (\alpha, \mu_1)$. Since this just is the second outer term obtained by squaring the inside, it suffices to show that the inner term converges to twice it. Observe that

$$\mathbb{E}\left[\left\{\frac{\exp(\theta_\alpha^\intercal Z)}{1+\exp(\theta_\alpha^\intercal Z)}(Y-\theta_1)\right\}\left\{\frac{\exp(\alpha^\intercal Z)}{1+\exp(\alpha^\intercal Z)}(Y-\mu_1)\right\}\right] \tag{33}$$

$$= \mathbb{E}\left[\frac{\exp(\theta_\alpha^\intercal Z)}{1+\exp(\theta_\alpha^\intercal Z)}\frac{\exp(\alpha^\intercal Z)}{1+\exp(\alpha^\intercal Z)}\left\{Y^2-\theta_1 Y-\mu_1 Y+\theta_1\mu_1\right\}\right]. \tag{34}$$

Distributing and taking each term in turn, we have, by repeated application of the continuous mapping theorem,

$$\mathbb{E}\left[\frac{\exp(\theta_\alpha^\intercal Z)}{1+\exp(\theta_\alpha^\intercal Z)}\frac{\exp(\alpha^\intercal Z)}{1+\exp(\alpha^\intercal Z)}Y^2\right] \to \mathbb{E}\left[\left\{\frac{\exp(\alpha^\intercal Z)}{1+\exp(\alpha^\intercal Z)}\right\}^2 Y^2\right] \tag{35}$$

$$\theta_1\mathbb{E}\left[\frac{\exp(\theta_\alpha^\intercal Z)}{1+\exp(\theta_\alpha^\intercal Z)}\frac{\exp(\alpha^\intercal Z)}{1+\exp(\alpha^\intercal Z)}Y\right] \to \mu_1\mathbb{E}\left[\left\{\frac{\exp(\alpha^\intercal Z)}{1+\exp(\alpha^\intercal Z)}\right\}^2 Y\right] \tag{36}$$

$$\mu_1\mathbb{E}\left[\frac{\exp(\theta_\alpha^\intercal Z)}{1+\exp(\theta_\alpha^\intercal Z)}\frac{\exp(\alpha^\intercal Z)}{1+\exp(\alpha^\intercal Z)}Y\right] \to \mu_1\mathbb{E}\left[\left\{\frac{\exp(\alpha^\intercal Z)}{1+\exp(\alpha^\intercal Z)}\right\}^2 Y\right] \tag{37}$$

$$\theta_1\mu_1\mathbb{E}\left[\frac{\exp(\theta_\alpha^\intercal Z)}{1+\exp(\theta_\alpha^\intercal Z)}\frac{\exp(\alpha^\intercal Z)}{1+\exp(\alpha^\intercal Z)}\right] \to \mu_1^2\mathbb{E}\left[\left\{\frac{\exp(\alpha^\intercal Z)}{1+\exp(\alpha^\intercal Z)}\right\}^2\right]. \tag{38}$$

Combining terms completes the proof. □

(D) $P\|\psi_{\theta_0}\|^2 < \infty$ and $\Psi(\theta)$ is differentiable at θ_0 with nonsingular derivative matrix V_{θ_0}.

Proof. The first part holds for ψ_1 and ψ_0 if Y has finite second moment. To verify the second part, observe that

$$\frac{\partial}{\partial\theta_\alpha}\psi_1 = \frac{\exp(\theta_\alpha^\intercal z)}{\left(1+\exp(\theta_\alpha^\intercal z)\right)^2}(y-\theta_1)\,z, \tag{39}$$

$$\frac{\partial}{\partial\theta_1}\psi_1 = \frac{\exp(\theta_\alpha^\intercal z)}{1+\exp(\theta_\alpha^\intercal z)}. \tag{40}$$

These continuous partial derivatives are uniformly bounded within a neighborhood of θ_0:

$$\left|\frac{\partial}{\partial\theta_\alpha}\psi_1\right| \le (|y|+|\mu_1|+\epsilon)\,|z|, \tag{41}$$

$$\left|\frac{\partial}{\partial\theta_1}\psi_1\right| \le 1. \tag{42}$$

Moreover, these upper bounds are integrable if we assume that $\mathbb{E}[|Y|] < \infty$, $\mathbb{E}[|Z|] < \infty$, and Z and Y have finite variances. Thus, the Leibniz integral rule (applying the dominated

convergence theorem and mean value theorem) implies that

$$\frac{\partial}{\partial \theta_\alpha}\mathbb{E}\left[\psi_1\right]\Big|_{\theta=\theta_0} = \mathbb{E}\left[\frac{\partial}{\partial \theta_\alpha}\psi_1\right]\Big|_{\theta=\theta_0} = \mathbb{E}\left[\frac{\exp(\alpha^\intercal Z)}{\left(1+\exp(\alpha^\intercal Z)\right)^2}\left(Y-\mu_1\right)Z\right], \tag{43}$$

$$\frac{\partial}{\partial \theta_1}\mathbb{E}\left[\psi_1\right]\Big|_{\theta=\theta_0} = \mathbb{E}\left[\frac{\partial}{\partial \theta_1}\psi_1\right]\Big|_{\theta=\theta_0} = \mathbb{E}\left[\frac{\exp(\alpha^\intercal Z)}{1+\exp(\alpha^\intercal Z)}\right]. \tag{44}$$

A similar argument holds for the partial derivatives of ψ_0. Notably, one regularity condition for the derivative matrix to be nonsingular is that the expected value of the race probabilities must be bounded away from 0 and 1. □

(E) $\Psi_n(\hat{\theta}_n) = o_P(n^{-1/2})$ **and** $\Psi_n^\circ(\hat{\theta}_n^\circ) = o_P(n^{-1/2})$**.**

Proof. This follows for the last two coordinates of $\Psi_n(\hat{\theta}_n)$ because $\hat{\theta}_{1n}$ and $\hat{\theta}_{0n}$ are exact zeros of the estimating equation. The same is true of the last two coordinates of $\Psi_n^\circ(\hat{\theta}_n^\circ)$. □

C.2 Extending the Z-Estimator Framework to Other Models

In this section, we briefly describe how the linear disparity estimator that Elzayn et al. (2023) consider might fit into the Z-estimator framework used to prove asymptotic normality of the dual-bootstrap in Appendix C.1. The linear disparity estimator $\hat{\delta}_l$ is given by the estimated slope coefficient in the linear regression of Y on the estimated probability $\widehat{\Pr}(A = 1|Z)$ plus an intercept term.

We can formulate $\hat{\delta}_l$ as a Z-estimator. Specifically, let

$$\psi_\theta(z, a, y) \equiv \begin{bmatrix}\psi_\alpha(z,a,y)\\ \psi_l(z,a,y)\end{bmatrix} \equiv \begin{bmatrix} z\left\{a - \frac{\exp(\theta_\alpha^\intercal z)}{1+\exp(\theta_\alpha^\intercal z)}\right\} \\ \frac{\exp(\theta_\alpha^\intercal z)}{1+\exp(\theta_\alpha^\intercal z)}\left(y - \frac{\exp(\theta_\alpha^\intercal z)}{1+\exp(\theta_\alpha^\intercal z)}\theta_l\right)\end{bmatrix} \tag{45}$$

and assume that $\theta \equiv \begin{bmatrix}\theta_\alpha & \theta_l\end{bmatrix}^\intercal \in \Theta \subset \mathbb{R}^p$ where Θ is open and $p < \infty$ is fixed. Then, defining the map $\theta \mapsto \Psi(\theta) \equiv P\psi_\theta$, we can show that $\theta_0 \equiv \begin{bmatrix}\alpha & \delta_l\end{bmatrix}^\intercal$ satisfies $\Psi(\theta_0) = 0$, where δ_l is the true disparity. If the same five conditions discussed in Appendix C.1 also hold here, then the dual-bootstrap is asymptotically normal for the linear disparity estimator as well. We leave verification of these conditions to future work.