Online Appendices: “Statistical Bias in Racial and Ethnic Disparity Estimates using BIFSG”

Elena Derby, Connor Dowd, and Jacob Mortenson

National Tax Journal, Volume 78, Number 4, December 2025

Online Appendix A: Matching LIHTC Residents to Tax Records

In this appendix, we describe our procedure for matching micro data from the LIHTC Tenant Form with tax data. The LIHTC data contain several information items that are also present in many tax filings: first name, surname, date of birth, address, and the last four digits of a social security number (SSN) or individual tax identification number (ITIN). Our approach is an iterative process, where we generate matches using strict criteria initially, and gradually loosen these criteria to match additional observations.

We begin by creating a matrix of individuals and individual characteristics found on tax documents in 2017. In these data (and in the LIHTC data) we exclude individuals younger than 18 years of age and older than 100 years of age at the start of 2017.[1] Tax data include the variables used in our analyses, as well as full 9-digit SSNs/ITINs. We join this matrix with the matrix of LIHTC individuals on date of birth and last four digits of SSN/ITIN, and then (if necessary) on surname, first three letters of first name, and zip code.

There are 2.6 million adults between the ages of 18 and 100 (in 2017) in the LIHTC Tenant Form data. Before we merge these data with tax information, we eliminate any individuals who do not have a valid entry for their date of birth and last four digits of their SSN (for example, individuals with “0000” as the last four digits of their SSN, or with missing birth dates), and who are missing race and ethnicity identifiers, which are necessary for our analyses of BIFSG. After eliminating individuals based on these criteria, we are left with 1.9 million observations. We link 1.3 million of these observations to unique tax individuals using partial SSNs, dates of birth, zip codes, and names.

Finally, we eliminate any individuals who match to a tax identification number, but do not have any tax information recorded for 2017 (meaning they did not file a tax return or receive any information return such as a W-2 or 1099). There are also some individuals who are matched to duplicate entries in the tax data (one individual matched to multiple SSNs or ITINs), and we use other variables to eliminate the duplicate matches. After this process

[1] Observations of individuals above age 100 are more common in the LIHTC data and are almost entirely attributable to erroneous data entry.

we end up with 1.1 million observations, which we use for our analyses. In particular, we focus on tax filers when evaluating estimates of EITC take up and audit rates. In our data tax filers make up about 60 percent of the full sample population.

While the LIHTC data are already non-random, we also care about the potential differential ability to match LIHTC respondents to tax data. **Table A.1** provides the racial and ethnic composition (as a percent of total population) for the 1.9 million cleaned observations from the 2017 LIHTC Tenant form, compared to the 1.1 million observations in the matched sample population that we use for our BIFSG validation analyses. The table also provides the estimated difference in means (as a 95 percent confidence interval) for each estimate of race as a percent of the population. As shown in the table, for most race/ethnicity groupings, the difference in means is statistically significant, although small (in most cases less than a one percent difference). The difference in means for Black and Hispanic individuals is slightly higher, at around 3 percent.

Table A.1: Race/Ethnicity Percentages – LIHTC Tenant Data vs. Matched Analysis Data

	LIHTC Tenant Form Data	Matched Analysis Data	Difference in Means (95% CI)	
White	34.5%	33.8%	0.5%	0.8%
Black	39.9%	36.8%	3.0%	3.3%
ANHPI	1.4%	2.0%	-0.6%	-0.5%
Native	1.2%	1.2%	-0.1%	-0.0%
Multiracial	1.3%	1.3%	-0.0%	0.0%
Other	2.1%	2.4%	-0.3%	-0.3%
Hispanic	19.6%	22.4%	-2.9	-2.7%

Notes: This table summarizes the racial and ethnic composition of the LIHTC Tenant Form data population compared to the sample population of LIHTC tenants matched with tax information.

Based on the results of the t-tests provided in **Table A.1**, we cannot rule out that there is some small amount of selection bias in the matched analysis data. While this means that the analyses in this paper may not be full representative of all LIHTC residents, this does not affect our ability to use these data to draw conclusions about the efficacy of BIFSG in estimating tax disparities between different groups within a specific population.

Online Appendix B: Audit Rates

In addition to conducting second-stage validation on the use of BIFSG in estimating differences in the EITC, we investigate how BIFSG affects estimates of differences in audit rates between groups by race and ethnicity. This is motivated by the findings of Elzayn et al. (2024), who use BIFSG to estimate whether a tax unit is non-Hispanic Black, or not non-Hispanic Black, and then estimate differences in audit rates between these two groups. They find that Black taxpayers are audited a higher rate compared to non-Black taxpayers, even when controlling for other factors like EITC claiming and type of audit.

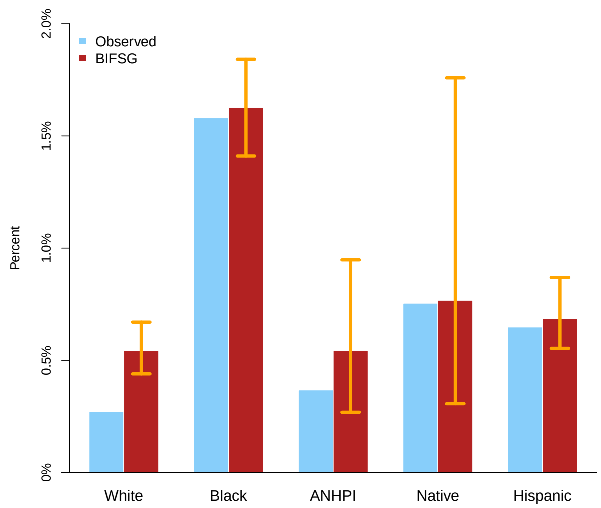
The audit data used in this paper consist of binary indicators for whether a taxpayer's 2017 tax return was audited (or not) between 2017 and 2021. These data come from the Audit Information Management System (AIMS), a database compiled by the IRS containing information related to the examination of taxpayer returns, including the outcomes of appeals, assessments, and adjustments of audits.

As shown in **Figure B.1**, our matched data supports the finding that Black taxpayers are audited at higher rates relative to non-Black taxpayers; however, we also find that BIFSG is not always accurate when estimating audit rates across racial and ethnic groups. As we find in our analysis of average tax rates and EITC claiming, BIFSG is better at predicting the audit rate of some taxpayers (in this case Black and Native taxpayers) and is not as accurate when it comes to other groups, particularly White taxpayers.

We further find that for all non-White taxpayers, confidence intervals constructed around point estimates of audit rates (calculated to incorporate uncertainty associated with BIFSG) also surround the observed outcomes. However, in line with our previous analyses, the observed audit rate for White taxpayers falls well below the estimated range of possible audit rates for this group. This makes sense considering that the BIFSG-predicted audit rate for White taxpayers is more than double the observed audit rate. Once again, this is suggestive of the introduction of bias in second stage BIFSG estimates, particularly through the miscategorization of non-White taxpayers as having high probability of being White.

Although estimated audit rates using BIFSG appear more accurate for non-White taxpayers, there are still some underlying patterns in the data that may raise concerns of potential bias in these estimates as well. Particularly among Black taxpayers, we find a strong correlation between the probability of being classified as the "correct" race/ethnicity and likelihood of being audited. This result is in line with the findings in Elzayn et al. (2024).

Figure B.1: Audit Rates by Race/Ethnicity (Observed and Predicted)



Notes: This figure graphs predicted audit rates by race and ethnicity for the sample of LIHTC residents. Blue bars show observed sample audit rates in the population and red bars are predicted audit rates using BIFSG. The orange bars represent 95 percent confidence intervals, which are bootstrapped and clustered by zip code, first name, and surname.

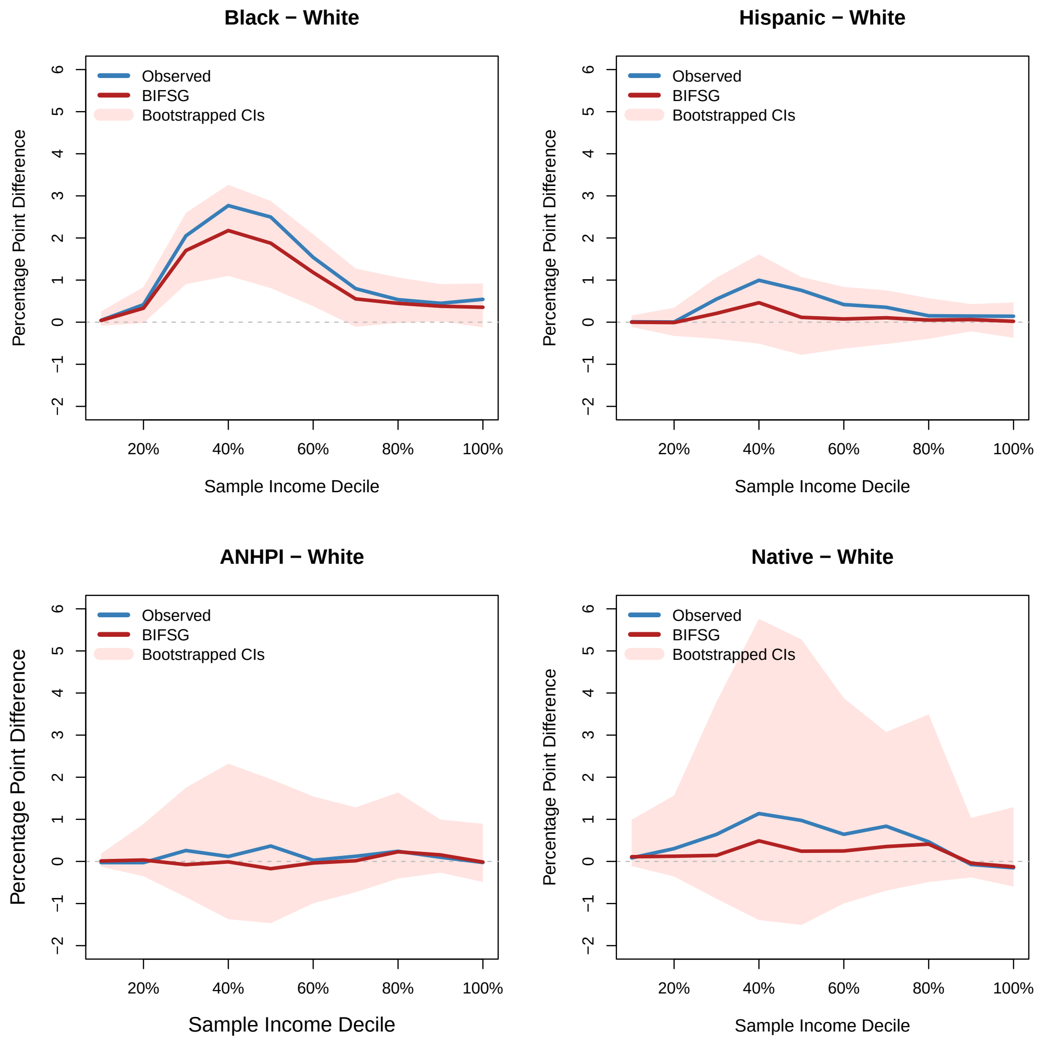
As shown in **Figures B.2** and **B.3**, the findings of our second-stage validity testing for audit rates are similar to the findings from our analyses of average tax rates and EITC claiming. BIFSG consistently underestimates differences in audit rates between non-White and White taxpayers. For Black taxpayers, BIFSG estimates that audit rates are higher compared to White taxpayers, but the observed difference is even greater than the predicted disparity for most income deciles. For Hispanic and Native taxpayers, BIFSG predicts a small or near-zero difference in audit rates when in reality both groups are audited at rates higher than White taxpayers. Finally, in some instances BIFSG predicts that White taxpayers are audited at higher rates than ANHPI taxpayers when the observed disparity is flipped: the audit rate for ANHPI taxpayers is slightly higher than that of White taxpayers for most sample income deciles.

Since estimates of audit rates are based on much smaller populations of taxpayers – relative to estimates of ATRs and EITC claiming – the confidence intervals surrounding the point estimates of disparities in audit rates do overlap with the observed differences for all groups. However, point estimates of audit rates are still consistently higher than predicted audit rates, and are close to the top of the possible range of estimated disparities for certain groups. For other groups, CIs incorporate a large range of possible audit rates – including zero – and as such, estimated disparities are not statistically different from zero.

We do not see the same pattern when comparing the audit rates of non-White taxpayers of different races and ethnicities. As shown in **Figure B.3**, the BIFSG estimates of audit differences between different non-White groups are very close to the observed differences by race and income level. This serves as further evidence that the primary bias introduced by BIFSG is the classification of non-White taxpayers as White. Once again, we also see that confidence intervals include observed differences between non-White taxpayers.

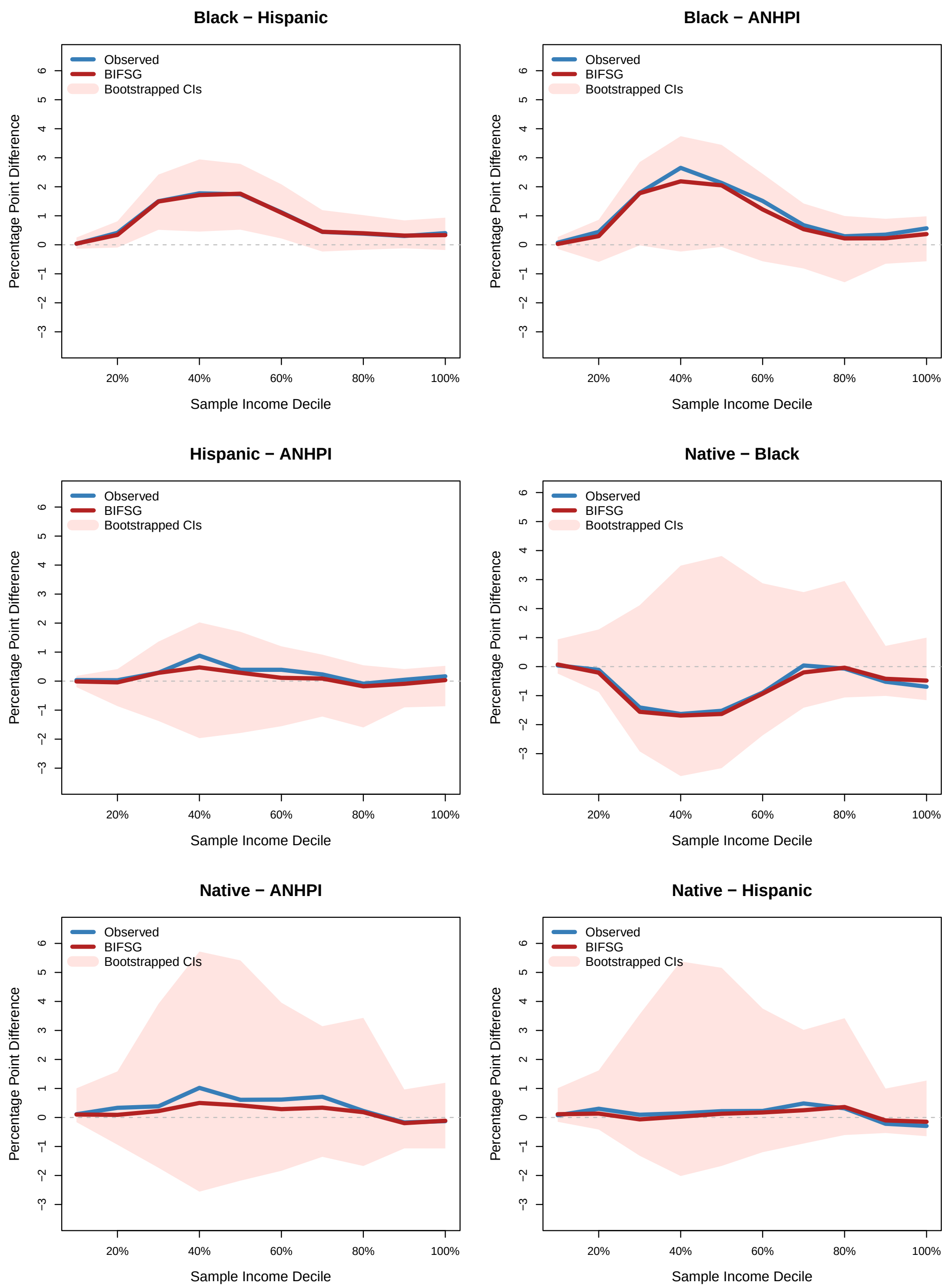
It is possible that BIFSG estimates of differences in audit rates between non-white taxpayers are more accurate because our sample population is socioeconomically homogeneous. If, for example, middle-income Black taxpayers are less likely to be categorized as Black compared to low-income Black taxpayers, and there are differences in audit rates between these two groups, then we would expect this to bias BIFSG estimates of differences in the average audit rate between Black taxpayers and other non-White groups, so long as the same patterns did not exist for these other non-Black groups. However, this is difficult to evaluate given our sample of taxpayers.

Figure B.2: Differences in Audit Rates between Non-White and White Taxpayers



Notes: This figure graphs the difference in audit rates between non-White and White taxpayers. Audit rates are calculated as a percent of the sub-population defined by race and ethnicity. Differences are calculated as the audit rate of non-White taxpayers minus the audit rate of White taxpayers. Differences are measured at each decile of the AGI distribution for the sample of LIHTC tenants. The shaded regions represent 95 percent confidence intervals, which are bootstrapped and clustered by zip code, first name, and surname.

Figure B.3: Differences in Audit Rates between Non-White Taxpayers of Different Race/Ethnicity



Notes: This figure graphs the difference in audit rates between non-White tax filers of different races/ethnicities. Audit rates are calculated as a percent of the sub-population defined by race and ethnicity. Differences are calculated as the audit rate of group A minus the audit rate of group B. Differences are measured at each decile of the income distribution for the sample of LIHTC tenants. The shaded regions represent 95 percent confidence intervals, which are bootstrapped and clustered by zip code, first name, and surname.

Online Appendix C: Bounding with Probabilistic and Linear Estimators

In addition to evaluating estimates using the probabilistic BIFSG methodology, we use our sample of LIHTC tenants to test a bounding method introduced in Elzayn et al. (2024). They find that if certain covariance conditions hold, the true disparity between two distinct racial/ethnic groups is bounded between the probabilistic and linear estimates of the disparity, using BIFSG to approximate race. The conditions are:

1. If $\mathbb{E}[\text{Cov}(Y, B|b)] \geq 0$ and $\mathbb{E}[\text{Cov}(Y, b|B)] \geq 0$ then $D_p \leq D \leq D_l$
2. If $\mathbb{E}[\text{Cov}(Y, B|b)] \leq 0$ and $\mathbb{E}[\text{Cov}(Y, b|B)] \leq 0$ then $D_l \leq D \leq D_p$

where D_p is the probabilistic estimator of the disparity, D_l is the linear estimator, B is observed race/ethnicity (binary), and b is the BIFSG-predicted probability of being that race/ethnicity. In short, in order for the two estimators to bound the true disparity two different correlations must have the same sign (both positive or both negative): the correlation of the outcome variable and predicted race, given true race, and the correlation of the outcome variable and true race, given predicted race (Elzayn et al., 2024).

In their paper, Elzayn et al. evaluate the difference in audit rates between Black and non-Black taxpayers, so that the linear estimate of this disparity is simply the value of the coefficient from the regression of the incidence of being audited on P[Black]_i, the BIFSG-predicted probability of the individual being Black. Since we are evaluating disparities between six groups in our analysis, we take a slightly different approach. We use the following equation (1), regressing our outcome variable on all BIFSG-predicted probabilities except for P[White]_i.

$$Y_i = \alpha + \beta_1 \text{P[Black]}_i + \beta_2 \text{P[ANHPI]}_i + \beta_3 \text{P[Native]}_i + \beta_4 \text{P[Multi]}_i + \beta_5 \text{P[Hispanic]}_i + \epsilon_i \quad (1)$$

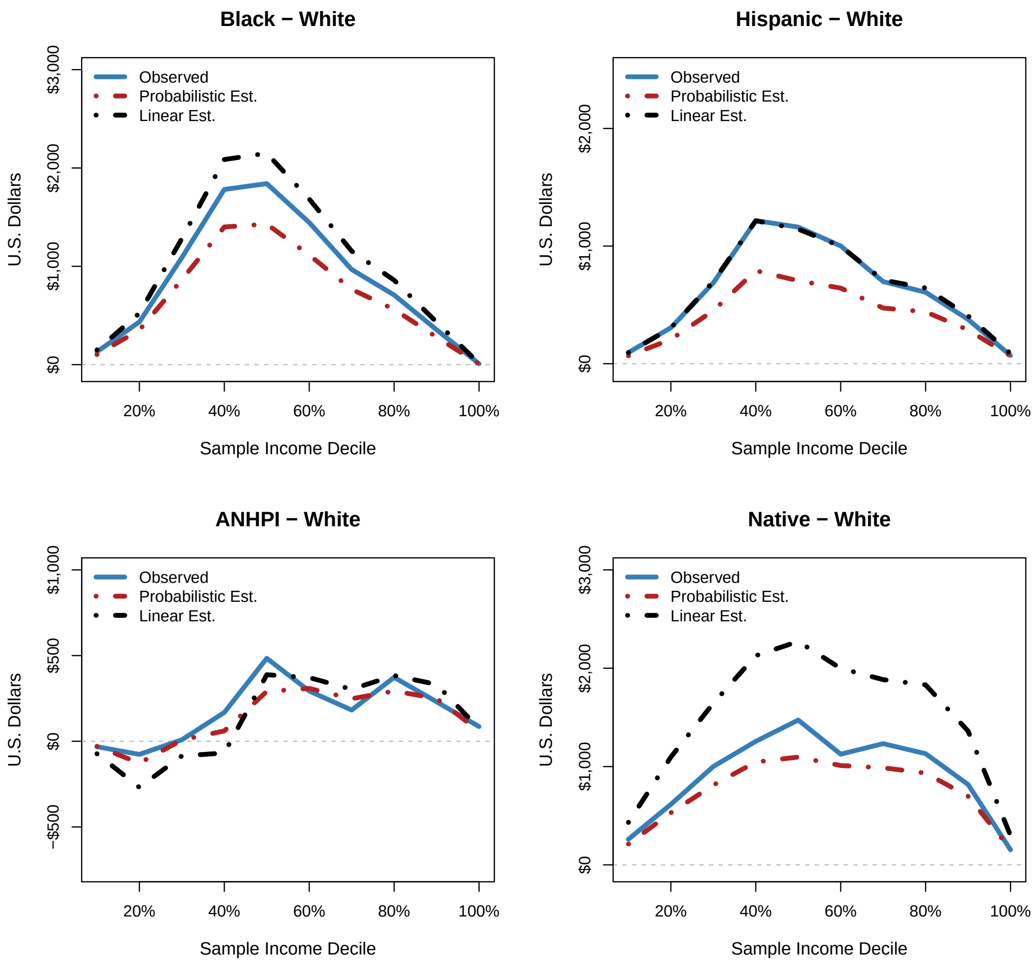
We then interpret each of the coefficients as the linear estimator of the difference in outcomes between the corresponding racial/ethnic group and White individuals. We use a similar approach for each of our non-White disparity estimates as well, excluding different probabilities from the regression depending on the comparison.

Figures C.1 through **C.4** provide the results of the bounding exercise: estimating differences in both EITC amounts and audit rates between different racial and ethnic groups using both probabilistic and linear BIFSG estimating procedures. As shown in the figures, the validation exercise has mixed results. There are instances in which the probabilistic estimate serves as a lower bound and the linear estimate serves as an upper bound for observed values (or vice versa). In line with the findings in Elzayn et al. (2024), this is the case when estimating the difference in audit rates between Black and White taxpayers.

However, there are some instances in which either the linear or probabilistic estimator crosses the observed values (as you go up the sample income distribution), and others in which both estimates are consistently higher or lower than the observed values. For example, we find that although the bounding procedure works when predicting the difference in audit rates between Black and White taxpayers, the two estimates are at times either both higher or both lower than the observed disparity in audit rates between Black taxpayers and different non-White groups (or between White and non-Black taxpayers in some instances).

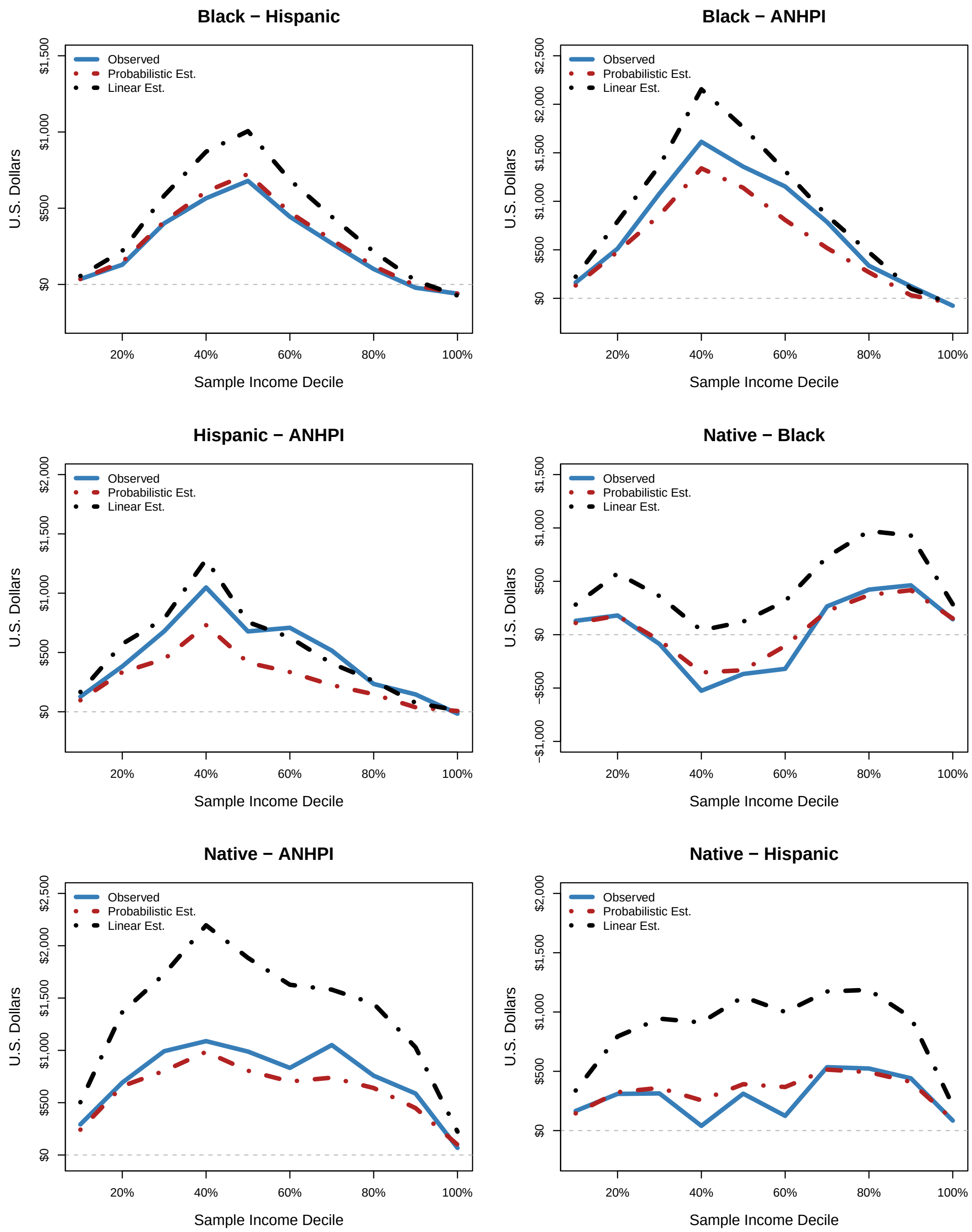
There are several possible explanations for this, two of which we discuss here. First, Elzayn et al. compare a single group (non-Hispanic Black taxpayers) to the rest of the population. Our setting, by contrast, compares six binary sets of groupings to each other. Second, the main result of the paper relies on certain covariance assumptions, which may fail when comparing different sub-groups in our setting. Presumably for at least some comparisons, this happens because the outcome and race/ethnicity variables do not meet the covariance conditions in Elzayn et al. (2024) for these sub-populations. It is important to note though, that in many settings of interest, race cannot be observed, which makes it impossible to calculate these covariances in order to determine how the bounding exercise will perform.

Figure C.1: Bounding Estimates for White and Non-White Taxpayers: EITC



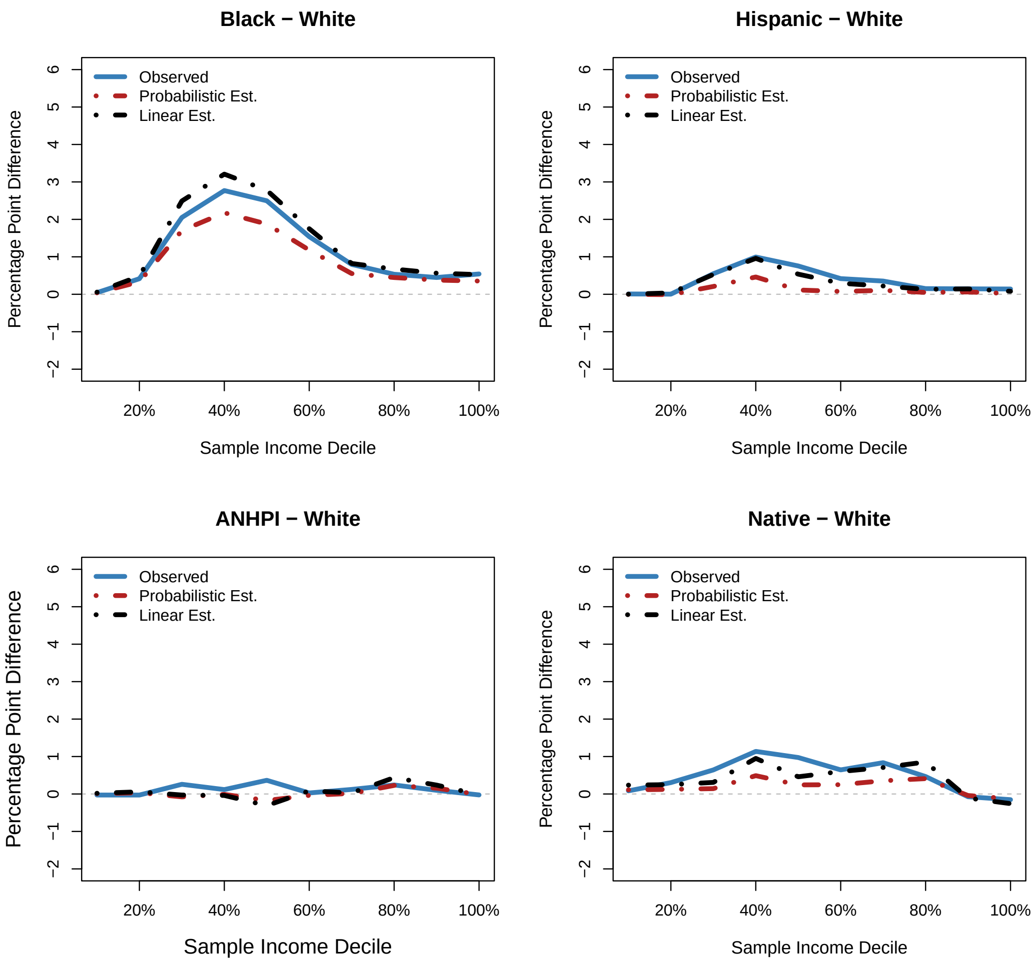
Notes: This figure graphs the difference in unconditional dollars of earned income tax credit claimed between non-White and White taxpayers (mean for non-White group minus mean for White group). Differences are measured for at each decile of the AGI distribution for the sample of LIHTC tenants. The dotted red line is the BIFSG probabilistic estimator and the dotted black line is the BIFSG linear estimator.

Figure C.2: Bounding Estimates for Non-White Taxpayers of Different Race/Ethnicity: EITC



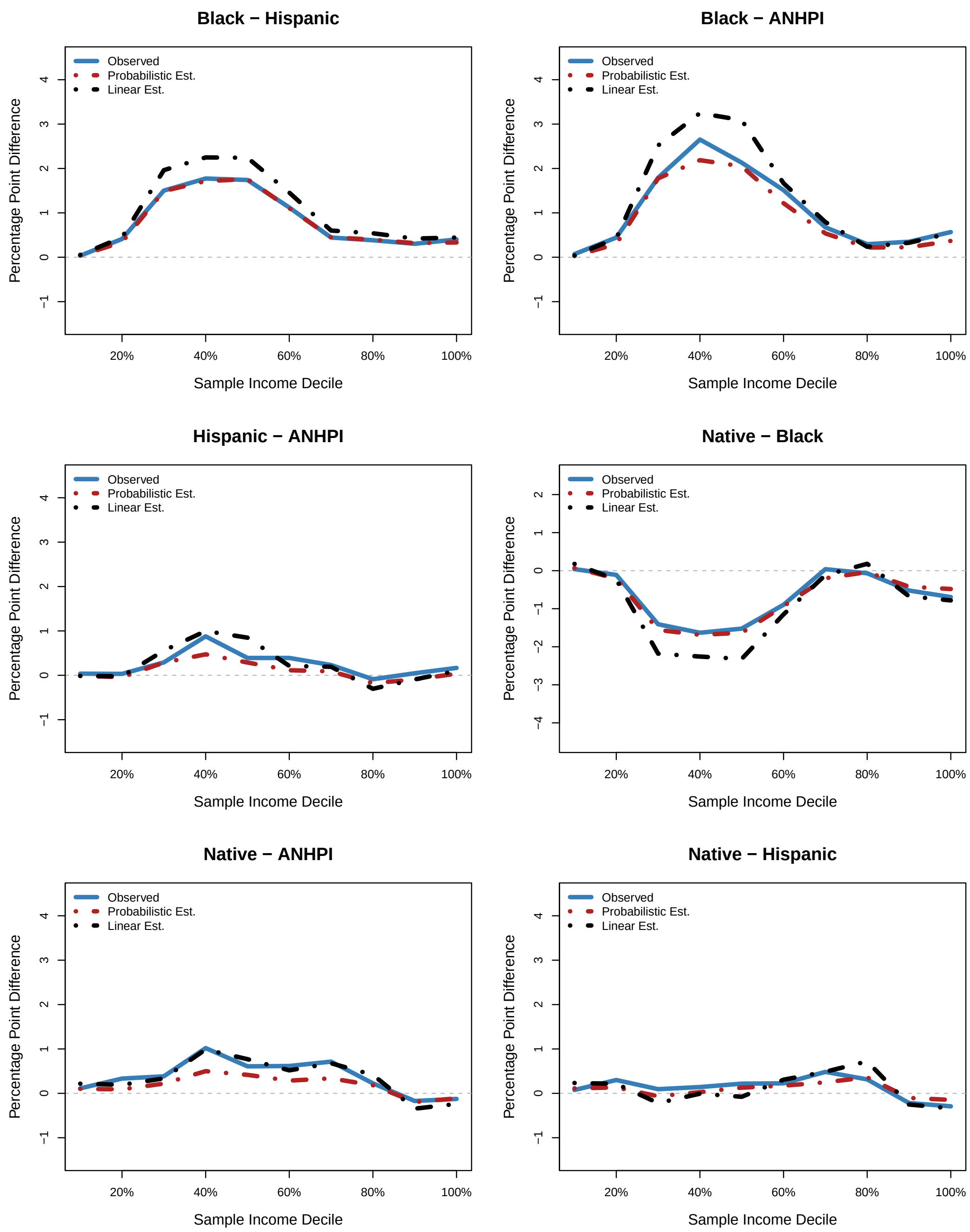
Notes: This figure graphs the difference in unconditional dollars of earned income tax credit claimed between non-White tax filers of different races/ethnicities (mean for group A minus the mean for group B). Differences are measured at each decile of the AGI distribution for the sample of LIHTC tenants. The dotted red line is the BIFSG probabilistic estimator and the dotted black line is the BIFSG linear estimator.

Figure C.3: Bounding Estimates for White and Non-White Taxpayers: EITC



Notes: This figure graphs the difference in audit rates between non-White and White taxpayers. Audit rates are calculated as a percent of the sub-population defined by race and ethnicity. Differences are calculated as the audit rate of non-White taxpayers minus the audit rate of White taxpayers. Differences are measured for at each decile of the AGI distribution for the sample of LIHTC tenants. The dotted red line is the BIFSG probabilistic estimator and the dotted black line is the BIFSG linear estimator.

Figure C.4: Bounding Estimates for Non-White Taxpayers of Different Race/Ethnicity: EITC



Notes: This figure graphs the difference in audit rates between non-White tax filers of different races/ethnicities. Audit rates are calculated as a percent of the sub-population defined by race and ethnicity. Differences are calculated as the audit rate of group A minus the audit rate of group B. Differences are measured at each decile of the AGI distribution for the sample of LIHTC tenants. The dotted red line is the BIFSG probabilistic estimator and the dotted black line is the BIFSG linear estimator.