Appendix B

Dissecting IDDA vs. CPS Income Differences[1]

May 15, 2024

1 Introduction

In section 3 of the paper, we showed that income inequality trends differ between IDDA and the CPS. For instance, income changes over time and across states are flatter in the CPS, compared with IDDA. There are several reasons IDDA and CPS could give different patterns: from noise in survey income measures to "non-classical" sources of differences such as sample composition and selection, income definition and misclassification, or survey data biases due to imputation, proxy responses, or misreporting.

We compiled a supplementary dataset to further compare income distributions in IDDA with those in the restricted-use Current Population Survey Annual Social and Economic Supplement (CPS ASEC), a standard source for income measurement in the U.S. The IDDA CPS Supplement dataset contains percentiles of income in the CPS disaggregated by race/ethnicity, sex, and place of birth (foreign-born or U.S.-born) for a subset of tax years, as well as analysis of a linked CPS-IRS sample. These supplementary statistics help account for the differences between the IDDA statistics and their counterparts measured in the CPS.

Related literature The literature suggests pure classical errors—that is, random measurement errors in the survey data that are orthogonal to the true income—do not drive the divergence between survey and administrative income measures. Bollinger et al. (2019) link the restricted CPS ASEC data to SSA W-2 records and find that nonresponse is higher in both tails of the earnings distribution and varies across demographic groups. Differences in nonresponse rates tend to bias within-group income inequality downward and also affect estimates of inequality across

[1]The opinions and conclusions expressed here are those of the authors and should not be interpreted as reflecting the views of the Federal Reserve Board of Governors, the Federal Reserve Bank of Minneapolis, or any other person associated with the Federal Reserve System. Any opinions and conclusions expressed herein are those of the authors and do not represent the views of the U.S. Census Bureau. The Census Bureau has ensured appropriate access and use of confidential data and has reviewed these results for disclosure avoidance protection (Project 7511151; Disclosure Authorization Numbers CBDRB-FY23-0277, CBDRB-FY23-0373, CBDRB-FY23-CES014-019, CBDRB-FY23-CES014-016, and CBDRB-FY24-0131.)

groups on both ends of the income distribution.[2] Kim and Tamborini (2012) link the Survey of Income and Program Participation (SIPP) respondents to SSA W-2 records and find evidence of income overreporting by low earners and underreporting by top earners. This type of non-classical error leads to a downward bias in within-group earnings inequality in the SIPP survey. They also found that Black-White earnings disparities are misestimated in the SIPP because of both larger overreporting at the bottom and more underreporting at the top among Black respondents, compared with White respondents. Gideon et al. (2022) show that income reporting patterns in the CPS differ within and across race, gender, and education groups, and these differences influence inequality estimates. Murray-Close and Heggeness (2019) document that a woman's earnings are underreported in the CPS by her male partner when he earns just less than her. Tamborini and Kim (2013) also document a proxy-respondent downward bias against single women in the SIPP.

Using the restricted-use CPS ASEC data, we explore various non-classical sources of differences in our IDDA setting to help interpret the differences between IDDA and the CPS for IDDA potential users. Consistent with the granular theme of the paper, we highlight how IDDA and CPS incomes diverge across race/ethnicity groups, for different income concepts, and along the income distribution. For instance, we note that racial differences in the fraction of non-zero W-2 earners with zero CPS earnings account for most of the racial differences in overreporting by individuals with low earnings.

2 IDDA–CPS Income Comparison Approach

The restricted-use CPS microdata do not contain ranked proximity swaps of top incomes and feature higher topcodes than the public-use microdata. Hence, measured incomes in the survey data should be closer to those in the administrative data than they are to incomes from the public-use data.[3] We can link the primary samples of 1040 and W-2 records underlying IDDA to the restricted-use CPS data using PVS-assigned identifiers (PIKs).

[2] Hokayem et al. (2015) also document that a higher nonresponses rate among low income households leads to lower estimates of poverty rate in the CPS. See Meyer et al. (2015) on transfer income underreporting in surveys.

[3] As of data year 2019, the primary component of person-level wage and salary income was topcoded at just under $10 million in the internal files, compared with a proximity swap threshold of $360,000 in the public-use data. In addition, from data year 2010 onward, the public-use files (but not the internal files) use a ranked proximity swapping procedure to assign income values above the topcode. See topcodes documentation in Flood et al. (2023) for details.

Income samples We contrast income measures across four main samples: (a) the unlinked IDDA sample of incomes, (b) the unlinked, restricted-use CPS sample of incomes, (c) the linked IDDA sample of incomes for respondents or households present in both the restricted-use CPS and the primary IDDA sample, and (d) the linked CPS sample of incomes for respondents or households present in both the restricted-use CPS and the primary IDDA. To construct the linked datasets, we merge CPS ASEC records from a given survey year to the final IDDA W-2 and 1040 samples from the corresponding tax year via the PIK (the relevant tax year for CPS ASEC earnings data is the one before the survey collection date). The unlinked, restricted-use CPS samples are defined in the same way as in the public-use CPS analysis. At the individual level, the unlinked sample includes ASEC respondents aged at least 16 who report positive wage and salary income. At the household level, the unlinked sample includes all ASEC respondents aged at least 16, including those who report zero household income. This is because the 1040 records underlying IDDA include returns with zero adjusted gross income. The linked samples inherit IDDA sample selection criteria. (See section 2 in the main paper for the construction of the IDDA samples.)

Income variables Next, we harmonize the income concepts across sources and the key variables we leverage to explore non-classical errors.[4]

Table B.1 describes the income variables used in the CPS and summarizes relevant differences from the tax measures used in IDDA. The individual-level earnings concepts and household wage and salary earnings align closely, but the household total income measures differ in a number of meaningful ways, highlighted in the right column.

Comparison flags We leverage various flags to inspect potential sources of divergence between IDDA and CPS incomes across race/ethnicity groups and along the income distribution. Imputation flags from the restricted-use CPS microdata allow us to assess how earnings nonresponse and survey data treatment of it contribute to the differences between IDDA and CPS incomes. At the individual level, we code an income concept as imputed for a respondent if any component of the income variable was imputed. At the household level, we code an income concept as imputed if it was imputed for the household member with the largest value for that income variable. For

[4]See Roemer (2000) for a comparison of incomes in the CPS and national accounts.

individual earners, we also use a flag for proxy-provided income responses to explore the role of proxy-induced differences between incomes in the CPS and tax data. To explore the role of income misclassification, we use a flag from the tax data indicating whether an individual received a Form 1099-MISC, a form typically received by self-employed workers.[5] Finally, we use the discrepancy between household size in the CPS and in IDDA to study whether the aggregation of tax units into households using Census address identifiers influences earnings measurement.

Table B.1: Income concepts in IDDA and the Current Population Survey

CPS Income Concept	Description	Types of Income Included	IDDA Income Concept	Major Differences
WSAL_VAL	Individual wage/salary earnings	Earnings from longest job if received wage/salary income in longest job + total wage or salary earnings from additional jobs. Includes tips, bonuses. Excludes self-employment, except if respondent owns an incorporated business and receives wages from it.	Individual-W2, wage compensation (WC) or total compensation (TC)	Neither the CPS nor IDDA wage income concept includes self-employment income. However, research has shown some CPS respondents misclassify self-employment income as wage income. TC includes elective deferrals reported in Box 12 of form W-2. In this section, individual earnings comparisons are based on WC.
HWSVAL	Household wage/salary income	Total of WSAL_VAL aggregated across all earners in a household	Household-1040, wage/salary income (WS)	See above row. Addresses may not align between the CPS and IRS data sources, causing household assignment to differ across the two sources for a given individual.
HTOTVAL	Total household income	Total income aggregated across all earners in a household. This includes wage and salary income and self-employment income, as well as non-wage income sources: *Social Security, SSI income, public assistance and welfare, disability income,* interest and dividends, rental income, veterans' benefits, *workers' compensation,* survivor's income, alimony, *child support payments,* distributions from pension or private retirement accounts, and unemployment compensation.	Household-1040, Adjusted gross income (GI)	CPS measure includes some types of nontaxable or partially taxable income that are excluded in IDDA (in italics). The CPS measure excludes above-the-line deductions on Form 1040, for example deductions from health savings accounts and student loan interest payments, which are subtracted from household AGI. The CPS measure excludes capital gains, which are included in household AGI.

[5] If individuals doing non-employee work receive an information return, it is most likely Form 1099-MISC. However, not all income earned by individuals doing non-employee work has an associated information return (Abraham et al., 2019). We do not observe the total earnings reported on 1099 forms, only whether an individual received one.

3 Individual Earnings Differences

Differences across unlinked data sources The analysis in section 3 of the main paper compared changes in distributional statistics over the full IDDA data period (1998-2018 for household income measures and 2005-2018 for individual earnings measures). Before delving into the sources of earnings differences, we compare earnings levels along the distribution for the overall sample and three selected race/ethnicity groups (Hispanic, non-Hispanic White, and non-Hispanic Black earners) in the CPS Annual Social and Economic Supplement and in IDDA. We focus on these groups to illustrate where and why results from IDDA might diverge from those in other income data sources, and to show that these differences vary across race and ethnicity even in relatively large samples.

Figure B.1 shows the 10th through 98th percentiles of individual earnings in 2012 (though similar patterns occur in 2005 and 2019). For all three groups, percentiles of earnings measured in the Current Population Survey are higher than in IDDA. In percentage terms, these differences are quite large at the bottom of the distribution (the 10th percentile of earnings in the CPS is close to double that in IDDA, and more for Black earners) but still meaningful at higher percentiles. However, magnitudes differ across the three race/ethnicity groups. The difference in earnings measured in the CPS and IDDA is higher for Black earners than for White earners at all percentiles. For Black earners, CPS earnings are around 30 percent higher than IDDA earnings at the median and 13 percent higher at the 98th percentile. For White earners, the difference falls from 17 percent at the median to 1 percent at the 98th percentile. As a result, relative to IDDA, the CPS understates the Black-White earnings gap, particularly in the tails of the distribution. Earnings percentiles measured in the CPS and IDDA are somewhat closer together for Hispanic earners.

Figure B.1 also suggests that measures of within-group inequality vary between the CPS and IDDA. For White and Black earners, both the 90/50 and 90/10 percentile ratios are lower in the CPS than in IDDA, since the difference in income values between the CPS and IDDA is larger in proportionate terms at the bottom and median than at the top. For Hispanic earners, the 90/10 ratio is substantially lower in the CPS than in IDDA, but not the 90/50 ratio.

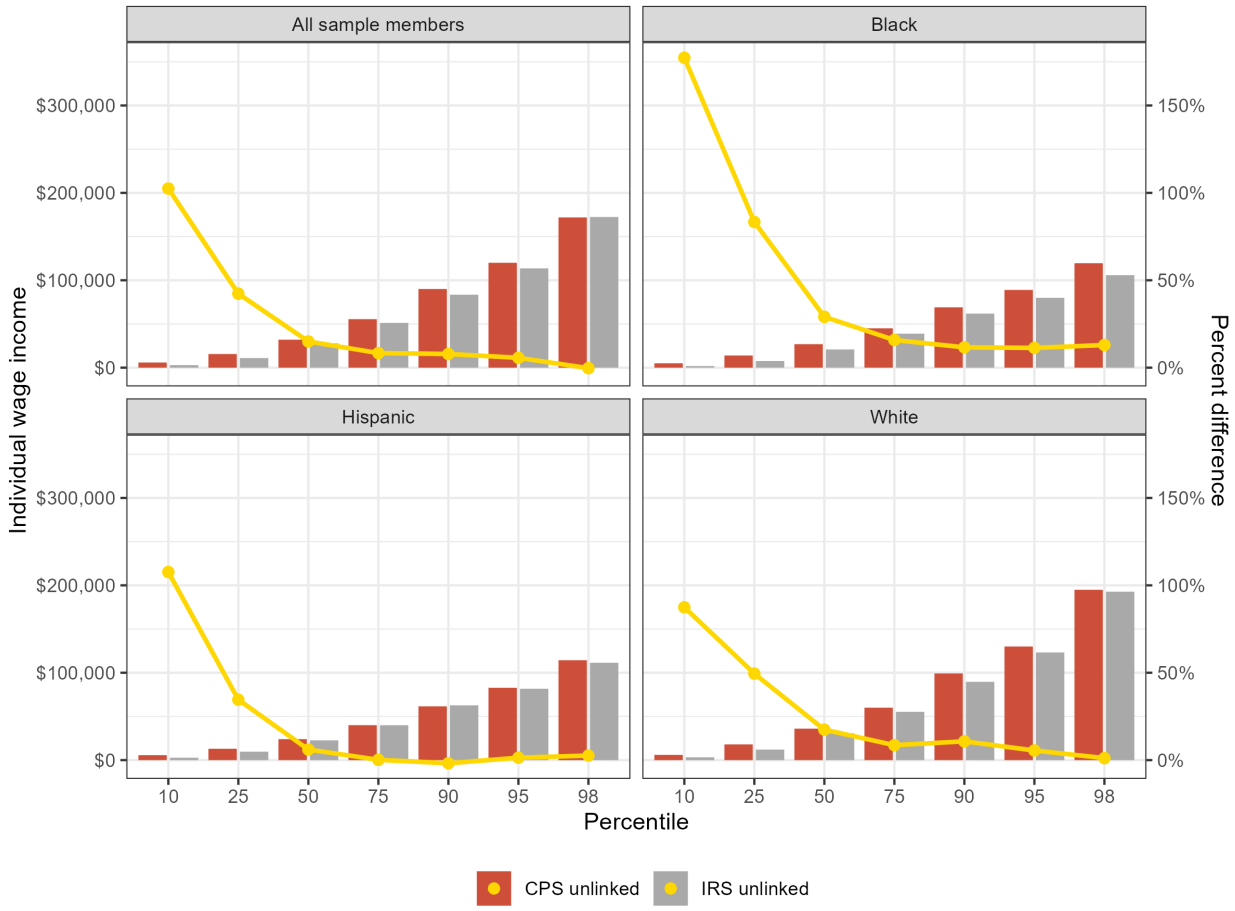

Figure B.1: Annual earnings in the CPS and in IDDA, by race and ethnicity, in 2012

Source: IDDA and Current Population Survey.
Note: Figure charts selected percentiles of the individual annual earnings distribution by race and ethnicity in 2012. Release authorizations CBDRB-FY23-0277 and CBDRB-FY23-0373.

Differences using linked samples Figure B.2 shows how these patterns change when we compare earnings measured in the CPS and administrative tax data for the same underlying sample. Among individuals we see in the W-2 data, does the distribution of CPS-reported earnings still look different from the distribution of earnings in the tax data? In the matched CPS-IDDA sample, we include respondents who do not report positive wage income in the CPS. This leaves open the possibility that both earnings and employment margin reporting differences influence distributional measures, including differences driven by earnings nonresponse in the CPS.

Restricting ourselves to the linked CPS-IDDA sample attenuates many of the differences observed in Figure B.1. At the bottom of the distribution, the large positive differences between earnings percentiles in the CPS and in IDDA become negative and much smaller in magnitude. In fact, the 10th percentile of CPS-reported earnings in the linked sample is zero for both Black and Hispanic earners and very small for White earners. In other words, at least 10 percent of Black and Hispanic earners in the linked sample do not report wage earnings to the CPS despite having received a W-2 showing positive wage compensation in 2019.[6] This employment margin difference suggests the IDDA W-2 data may be more likely than survey sources, to capture very small earnings values, such as earnings for very part-time workers or from short-term jobs, of note to data users.

At the highest earnings percentiles, CPS-defined earnings measures again fall below those based on IDDA. Yet, splitting the sample by race/ethnicity shows this pattern is driven by White respondents – and it is limited to percentiles above the 95th, which are substantially higher for White earners than for other groups. So, Figure B.1 is consistent with racial differences in reporting patterns that are driven by differences in income levels across groups. These differences influence inequality measures across groups but primarily at the top of the distribution or as a result of the employment margin, in line with Chenevert et al. (2015), who find that replacing survey-reported earnings with administrative earnings has little impact on average Black-White earnings gaps.[7]

Respondent-level earnings differences and determinants Both Figures B.1 and B.2 report sample-level differences in earnings and inequality measurement between the CPS and IDDA.

[6] Chenevert et al. (2015) similarly find that 9 percent of respondents with positive administrative earnings had no survey-reported earnings in a linked SIPP-DER dataset.

[7] Other authors, including Kim and Tamborini (2012) and Gideon et al. (2022), find more salient differences across race/ethnicity in linked survey-administrative data, but look at more granular demographic groups than we include here.

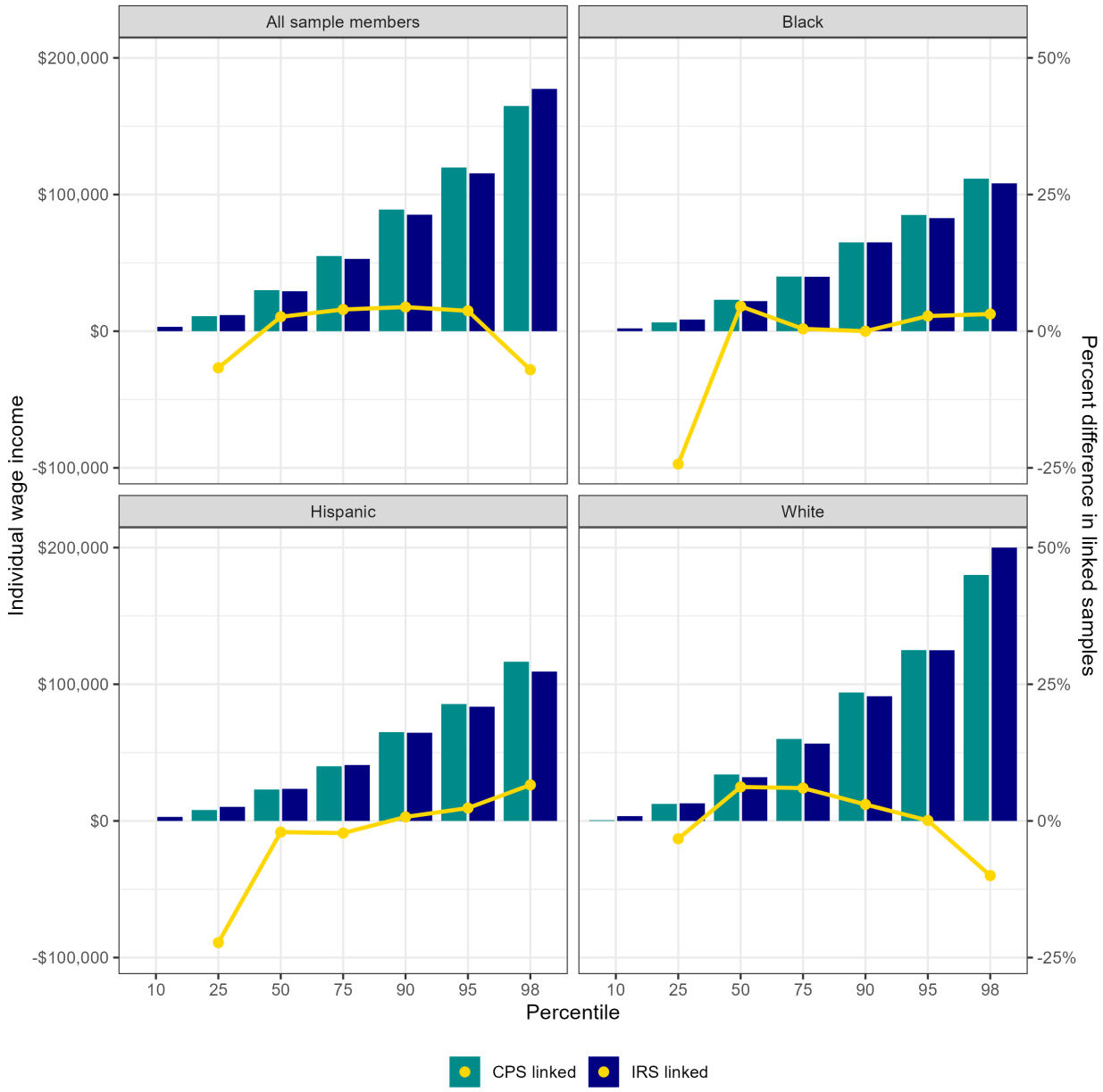

Figure B.2: Distribution of earnings in matched CPS-IDDA sample, by race and ethnicity

Source: IDDA and Current Population Survey.
Note: Figure charts selected percentiles of the individual annual earnings distribution by race and ethnicity in 2012. The linked sample includes CPS ASEC respondents aged 16 or over who can be matched to the IDDA W-2 sample via the PIK, which restricts the sample to individuals with positive wage compensation recorded on Form W-2. Release authorization CBDRB-FY23-0373.

Another approach is to assess individual-level differences in earnings reported to the CPS and in the tax data for earners in the matched sample. We summarize individual-level reporting differences within quantiles of the W-2 earnings distribution, and further split these between self-reported, proxy reported, and imputed earnings values. Additionally, we consider earners with a 1099-MISC information return (which identifies likely self-employed workers).

The results are shown in Figure B.3. The boxplots show the 25th percentile, median, 75th percentile, and mean earnings difference between the CPS and W-2 records, for individuals in the top and bottom quantiles of the W-2 earnings distribution. The comparison includes both earnings and employment margin differences (an individual with zero CPS-reported earnings and $\$x$ in W-2 earnings receives an earnings difference value of $-\$x$). Quantiles of the W-2 distribution are defined in the overall linked sample, across demographic groups.

Individual-level differences in CPS-reported earnings and W-2 earnings are highest in absolute terms at the top of the distribution, consistent with previous findings linking the CPS and SIPP to administrative records (Kim and Tamborini, 2012; Bollinger et al., 2019). The median earnings difference among individuals in the top 10 percent of the W-2 earnings distribution was just under –$7,000 (i.e., the earner with the median reporting difference reported $7,000 less in income to the CPS than is recorded on their W-2 record). The mean was close to $40,000, suggesting a small number of large underreports in the CPS relative to the W-2 data. Among those in the bottom quartile of W-2 earnings, the median earnings difference was close to zero, but the mean was nearly $5,000, a meaningful gap—the 25th percentile of individual W-2 earnings in the linked sample was $11,790. These patterns may reflect social desirability bias (respondents may report an income they perceive as closer to average), differential nonresponse, participation in informal labor market activities, or income misclassification (Kim and Tamborini, 2012; Abraham et al., 2021; Imboden et al., 2023). They could also include misreporting of earnings by employers (Gideon et al., 2022).

Within quantiles of the overall W-2 earnings distribution, the IQR of individual-level earnings differences were largely similar for Black, Hispanic, and White respondents. The median earnings difference among top White earners was somewhat smaller than the one for other groups, but the mean was much larger: White earners appear overrepresented among large CPS-underreporters at the top of the distribution.

Breaking out our comparison by response status, we note that reporting differences were larger among high earners whose CPS earnings were given by a proxy respondent.[8] Earners with a 1099-MISC information return also had greater mismatch. At the bottom of the distribution, this likely reflects misclassification of self-employment income as wage income, which research has documented in the CPS and is especially relevant for gig workers (Abraham et al., 2019, 2021). The largest systematic mismatch occurs for individuals whose wage income was imputed in the CPS. This result bears out the finding that earnings nonresponse is nonrandom in the CPS ASEC (Bollinger et al., 2019). Overall, very high and very low earnings are not well observed in the CPS even without topcoding, and this problem is worse for individuals who do not respond to earnings items or have a more complicated mix of income sources. This has been shown previously in linked CPS-administrative data, and it is true for the IDDA samples as well. An implication for data users is that the CPS tends to give lower readings of inequality both within and across groups than the IDDA W-2 sample when those measures rely on the tails of the distribution. Similarly, small nominal changes in low earnings percentiles in IDDA can substantially move percentile ratio measures.

4 Household Income Differences

Household income concepts in IDDA provide measures of total resources available to individuals sharing a common residence. Understanding how these income concepts and household aggregation methods may differ from other sources can help users determine the best applications of the IDDA data to their research question. For instance, household income measures in IDDA are not well suited to studying poverty, as they do not contain transfers or other after-tax income sources.

Here, we compare the distribution of household income measures in the linked CPS-IDDA 1040 sample. This restricts the ASEC sample to individuals who are listed as a primary or secondary filer on a Form 1040 filed for the year 2012. The distribution of household income is somewhat higher in the linked sample than in the unlinked CPS, in line with this restriction.

[8] Close to half of ASEC earnings reports are given by proxy respondents, and more than a quarter of earnings reports are imputed (Bollinger et al., 2019).

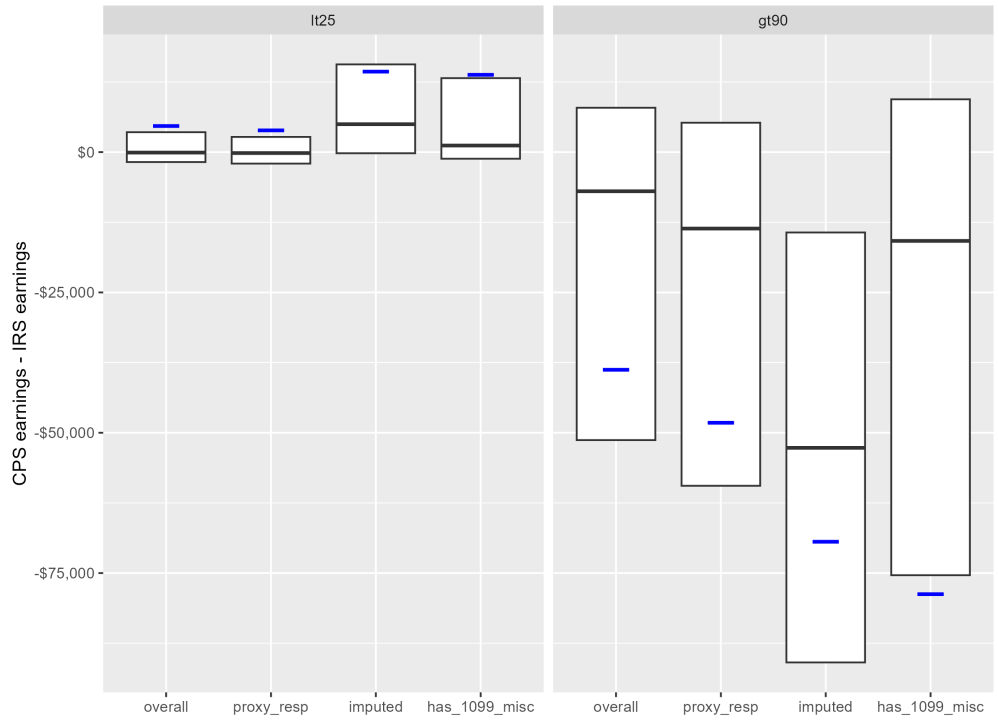

Figure B.3: Summary of earnings differences in top 10 percent and bottom 25 percent of W-2 earnings

Source: IDDA and Current Population Survey.
Note: Boxplots show the 25th, 50th, and 75th percentiles of earnings differences between individual wage earnings reported in the CPS and IDDA W-2 data. These are reported for respondents in the bottom 25 percent of individual earnings (left) or the top 10 percent (right), defined using W-2 wage compensation in the linked sample. On the horizontal axis at the bottom of the figure, "proxy_resp" indicates whether CPS earnings were given by a proxy respondent; "imputed" flags respondents whose CPS wage earnings were imputed; "has_1099_misc" indicates whether the respondent received a 1099-MISC information return. The blue bars show mean earnings differences for each group. Release authorization CBDRB-FY23-0373.

Differences using linked samples Percentiles of household wage and salary income in the CPS and IDDA are shown in Figure B.4. As with individual earnings, the overall pattern is an inverted-U shape: percentiles of income as measured in the CPS are lower than those in IDDA in the tails of the distribution, and similar to those in IDDA around the median (shown in the top left panel). The other three panels show that differences at the highest-income percentiles are more pronounced for White respondents, while those at lower percentiles are more pronounced for Black and Hispanic respondents. This influences relative income measures only slightly at the top: for example, at the 98th percentile, the Black-White income ratio is 66 percent using CPS household wage and salary income and 63 percent using IDDA. The effect is larger in ratio terms at the 25th percentile.[9] Still, we conclude that along much of the distribution, percentiles of household wage and salary income are closely aligned in the linked sample.[10]

In both the CPS and IDDA, household wage and salary income is the sum of earnings for individuals sharing a common address. Both Figure B.2 and Figure B.4 show relatively small differences between the CPS and IDDA, and similar patterns across race/ethnicity. However, percentiles of household wage and salary income are lower relative to IDDA than percentiles of individual earnings, particularly in the middle of the distribution. This could stem from a number of factors. Individuals who over- or underreport income in the CPS, relative to IDDA, could reside with earners who show the same, opposite, or no reporting difference. Such sorting patterns within households could vary across the income distribution and/or by race and ethnicity. It's also possible that the aggregation of income across 1040 forms captures different earners than the aggregation of incomes across workers in the CPS. Table B.2 includes a more detailed description of households in the CPS and IDDA, and finds a high rate of overlap between household constructs in the linked sample.

On the other hand, we do find notable differences in measures of total household income. These patterns are consistent with differences in the types of income captured by household AGI in

[9]Yet, respondent-level differences in the linked sample show that low-income Black and Hispanic filers do not appear more likely than low-income White respondents to underreport household wage and salary income in the CPS, relative to IDDA (see Figure B.9).

[10]This analysis includes incomes reported on 1040 forms filed for the tax year 2012. One way to explore whether patterns change as the universe of filers changes is to consider 2019, as many households with low or zero income filed taxes for that year to receive Covid-19 stimulus payments. Figure B.8 shows that percentiles of household wage/salary income are higher in the CPS, relative to IDDA in the linked sample in that year – the distribution of income in IDDA is pushed to the right as a result of filing changes. The effect is strongest at low percentiles, creating a downward sloping trend in CPS-IDDA differences along the distribution, as opposed to an inverted U.

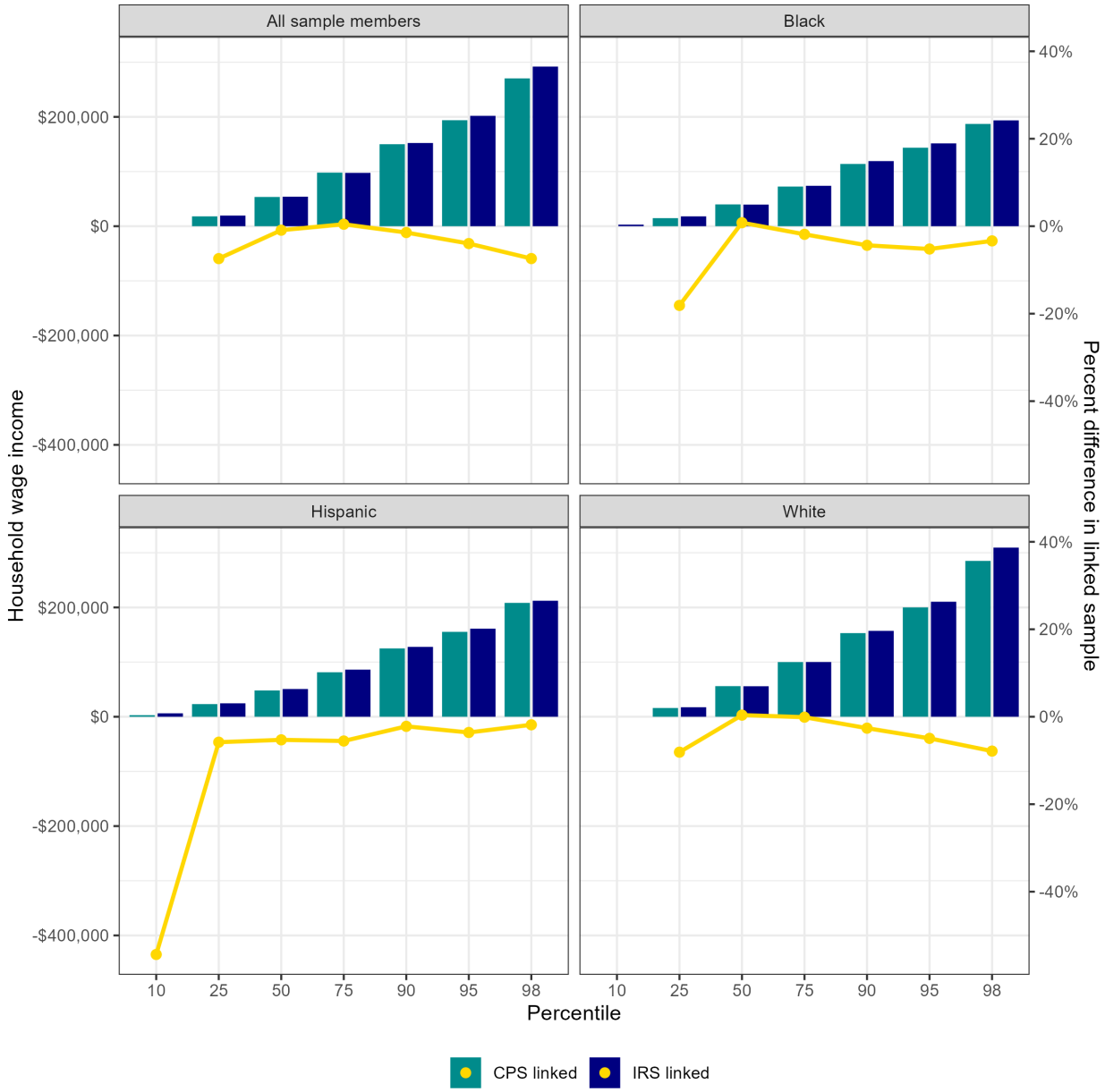

Figure B.4: Distribution of household wage and salary income in matched CPS-IDDA sample, by race and ethnicity

Source: IDDA and Current Population Survey.
Note: Figure charts selected percentiles of the distribution of household wage and salary income by race and ethnicity in 2012. The linked sample includes CPS ASEC respondents aged 16 or over who can be matched to the IDDA 1040 sample via the PIK, which restricts it to individuals listed as a primary or secondary filer on a Form 1040 filed for tax year 2012. Release authorization CBDRB-FY23-0373.

IDDA and the closest CPS variable, but they vary across race/ethnicity. This analysis is shown in Figure B.5. For White filers and in the overall linked sample, the distribution of CPS total household income is compressed relative to the distribution of household AGI: CPS incomes are both larger than IDDA AGI at the bottom and much smaller than IDDA AGI at the top. For Black and Hispanic respondents, this pattern is much less stark. For Hispanic respondents in particular, percentiles of income measured in the CPS hover at about 5 percent lower than IDDA at the median and above. Therefore, the CPS documents smaller income gaps between Black and Hispanic respondents and White respondents than IDDA does at the top of the distribution.

Racial differences in the measurement of total household income may reflect the role of capital gains, a growing source of top incomes that has disproportionately accrued to White households and is not captured in the CPS total household income variable (Derenoncourt et al., 2023). At the bottom of the distribution, the CPS household income concept includes sources of transfer income that are nontaxable and so are not captured in IDDA. We are not able to disentangle these differences in variable definition from reporting patterns in non-labor income sources (such as transfer or retirement income, as in Meyer and Mittag (2019) and Bee and Mitchell (2017)), since we cannot isolate individual components of non-wage income in the tax data.

Household-level earnings differences and determinants Finally, we look at household-level differences between household wage and salary income values observed in the CPS and the 1040-derived tax data for individual respondents in the matched sample. We explore the role of household characteristics at the top and at the bottom of the earnings distribution. We consider the following household variables: whether the household contains any dependents, whether the highest earner's household wage and salary income was imputed in the CPS, and whether the highest earner in the household received a form 1099-MISC. We also consider whether household size in the CPS exactly matches the size of the MAFID-based household in the IDDA sample. We report our findings in Figure B.6.

At the bottom of the distribution, we find that CPS-reported wage and salary income is higher than household WSI measured in the tax sample for most respondents. This is especially true for households with dependents, with a 1099-MISC, and with imputed wage and salary income. In contrast, at the top of the distribution, CPS wage and salary income is less than household WSI, as

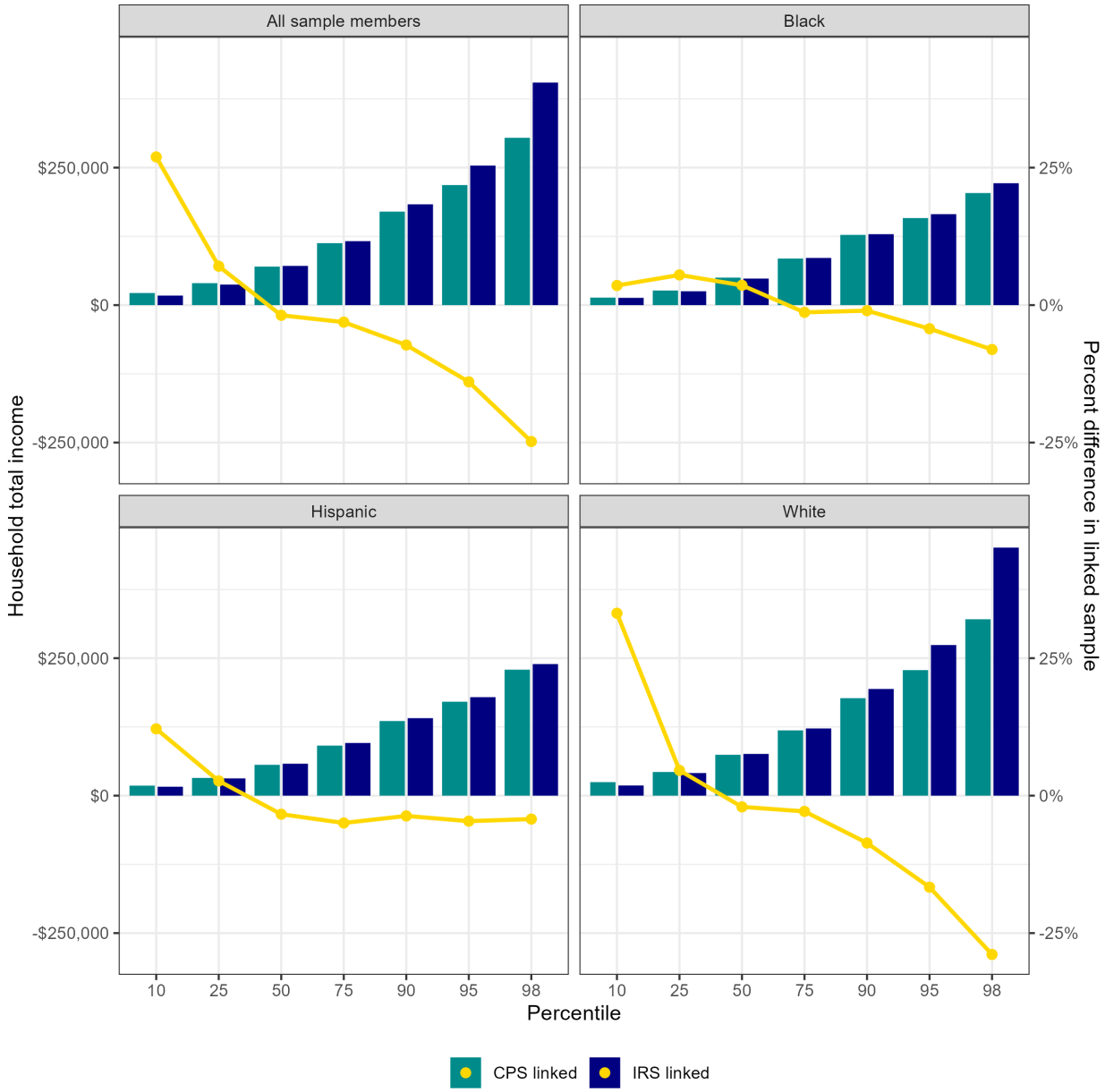

Figure B.5: Distribution of total household income in matched CPS-IDDA sample, by race and ethnicity

Source: IDDA and Current Population Survey.
Note: Figure charts selected percentiles of the distribution of total household income by race and ethnicity in 2019. The linked sample includes CPS ASEC respondents aged 16 or over who can be matched to the IDDA 1040 sample via the PIK, which restricts to individuals listed as a primary or secondary filer on a Form 1040 filed for tax year 2012. Release authorization CBDRB-FY23-0373.

expected. Households whose size matches exactly in the CPS data and in the IRS data have slightly less mismatch, and households with imputed wage and salary income have much more mismatched incomes.

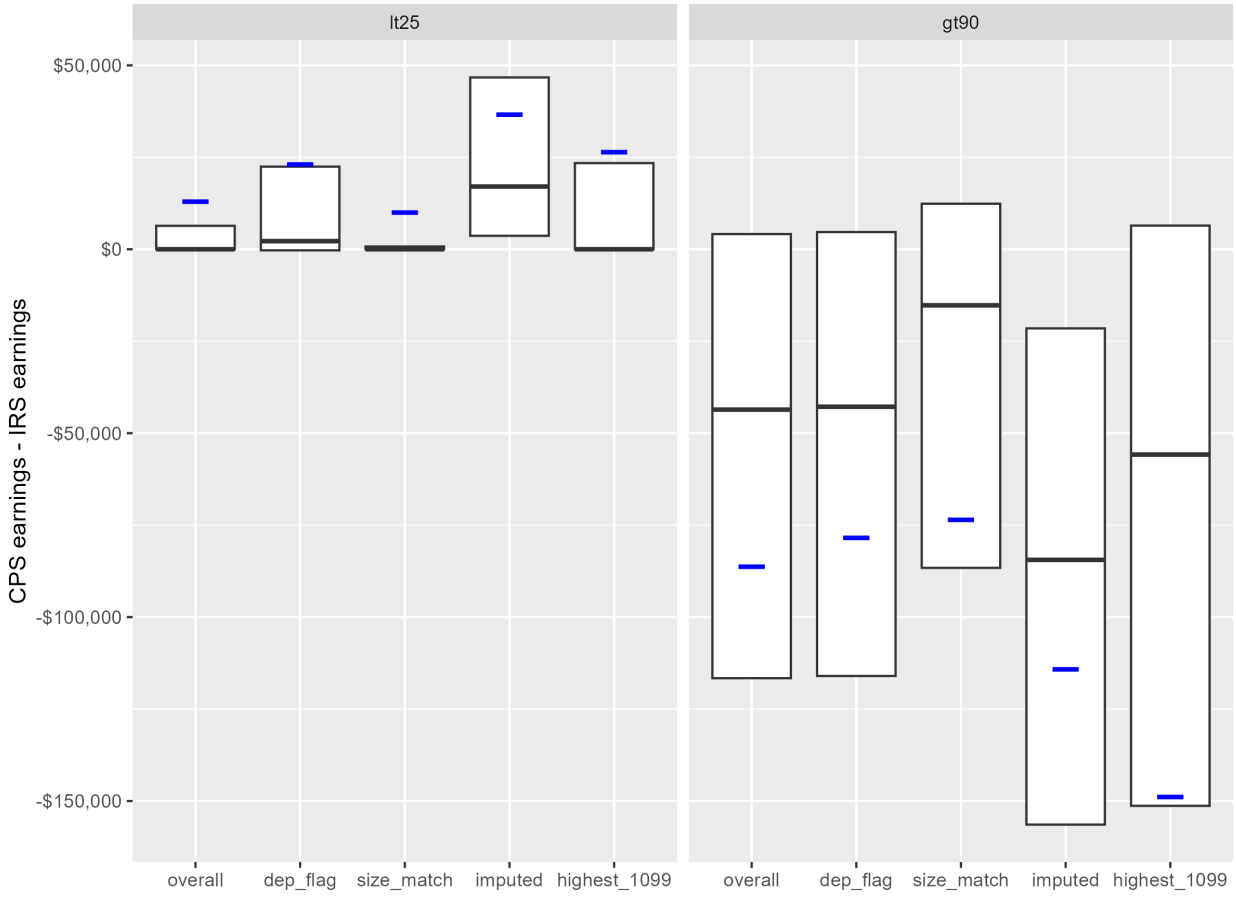

Figure B.6: Household earnings differences by household characteristics

Source: IDDA and Current Population Survey.
Note: Boxplots show the 25th, 50th, and 75th percentiles of earnings differences between household wage and salary income reported in the CPS and IDDA 1040 data. These are reported for respondents in the bottom 25 percent of household wage and salary income (left) or the top 10 percent (right), defined using the 1040 household WSI measure in the linked sample. On the horizontal axis at the bottom of the figure, "dep_flag" indicates respondents whose IDDA household includes at least one dependent; "size_match" indicates an exact match between household size constructed in the CPS and IDDA; "imputed" flags whether the highest wage/salary earner in the CPS household had any imputed income values; and "highest_1099" indicates whether the highest total earner in the CPS household had a 1099-MISC information return. The blue bars show mean earnings differences for each group. Release authorization CBDRB-FY23-0373.

4.1 Address-Based Households in IDDA

Table B.2 shows how the address identifier ("MAFID") used to construct household incomes in IDDA maps onto households in the Current Population Survey, using the tax year 2012 as an example. The first column compares the rate of overlap between MAFIDs and households in the CPS ASEC: A CPS household is considered a "match" to its IDDA counterpart if all the household members in the CPS have the same value of MAFID in the IDDA sample and if all the members of the IDDA MAFID that show up in the CPS data are part of the same CPS household. This rate is high, at least 90 percent for all demographic groups. However, it does not rule out situations where some members of a CPS household are not represented in the IDDA samples.

Table B.2: Households and addresses in the CPS and IDDA

Group	MAFID overlap rate with CPS households	Mean CPS household size	Mean IDDA household size	Exact MAFID household size match
All	94%	2.87	3.12	63%
Hispanic	90%	3.5	3.9	47%
AIAN	91%	3.1	3.4	52%
Asian	94%	3.3	3.6	58%
Black	90%	3	3.3	49%
NHOPI	91%	3.7	4	52%
White	95%	2.7	3	68%
Foreign-born	93%	3.3	3.7	54%
U.S.-born	94%	2.8	3.1	64%

Note: Table shows characteristics of households in the linked CPS-IDDA 1040 sample, using tax year 2012 as an example. In IDDA, household size is defined as the total number of primary and secondary filers and dependents reported on all 1040 forms filed at a common address (MAFID). The household "overlap rate" gives the probability that, within the linked sample, CPS and IDDA households match 1:1 (all observed CPS household members are linked to the same MAFID, and all observed members of the MAFID are part of the same CPS household). Release authorization CBDRB-FY23-0373.

The second and third columns show that, on average, household size is slightly larger in IDDA than the CPS, possibly because some tax filers claim dependents who do not physically reside with them (for example, students or elderly family members). However, the relative differences in mean household size across demographic groups is preserved between both sources. Finally, column 4 suggests that the CPS and IRS household concepts are less tightly aligned for individuals who are non-White or Foreign born: even though the difference in mean household size between the CPS and IRS data is not particularly large for these groups, they are less likely to reside in a CPS household that has the exact number of members associated with their value of MAFID in the IDDA samples.

5 Conclusion

Studies using linked survey and administrative data have explored how income reporting patterns in surveys might vary across groups or along the income distribution. We apply these insights to IDDA to better understand the sources of divergence from public-use data documented in the paper. Overall, wage and salary-derived income measures align well between the CPS and IDDA at both the individual and household level. Consistent with prior literature, we find some "trouble in the tails" (Bollinger et al., 2019).

At high percentiles, relative to IDDA, the CPS tends to understate earnings levels and measures of earnings gaps across race/ethnicity groups. This pattern is expected and matched by respondent-level differences in earnings reported to the CPS and in the tax data. However, we add that it is driven primarily by the top of the White earnings distribution in our data setting. At low percentiles, we find that differences between the CPS and IDDA are sensitive to the fraction of non-zero W-2 earners with zero CPS earnings, especially for Black and Hispanic respondents.

Finally, we suggest that differences in the types of non-wage income captured in the CPS and in IDDA are meaningful for understanding discrepancies in total household income measures, such as those reported in Table 6 of the main paper.

6 Additional Figures

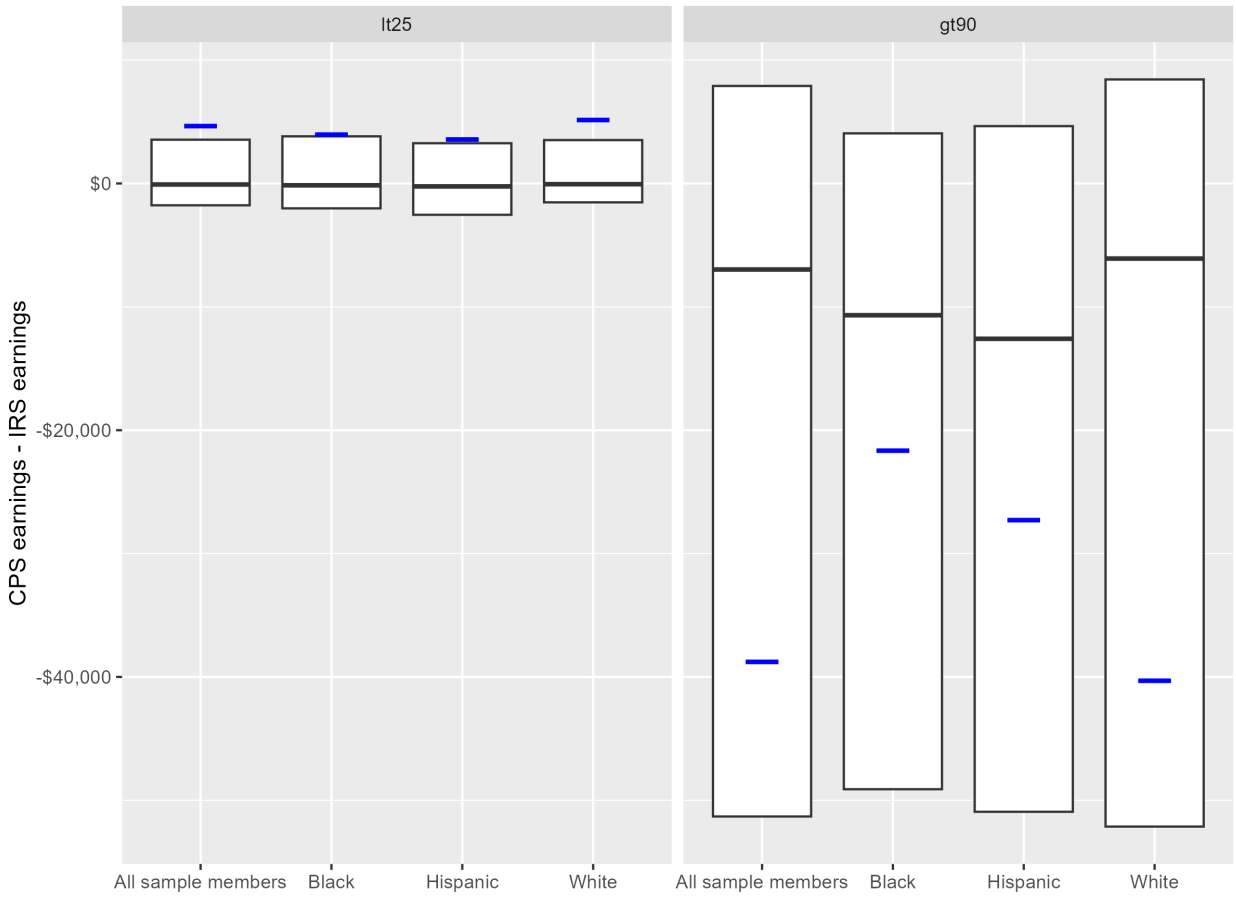

Figure B.7: Respondent-level earnings differences by race and ethnicity

Source: IDDA and Current Population Survey.
Note: Boxplots show the 25th, 50th, and 75th percentiles of earnings differences between individual wage earnings reported in the CPS and IDDA W-2 data. These are reported for Hispanic, non-Hispanic White, and non-Hispanic Black respondents in the bottom 25 percent of individual earnings (left) or the top 10 percent (right), defined using W-2 wage compensation in the linked sample. Release authorization CBDRB-FY23-0373.

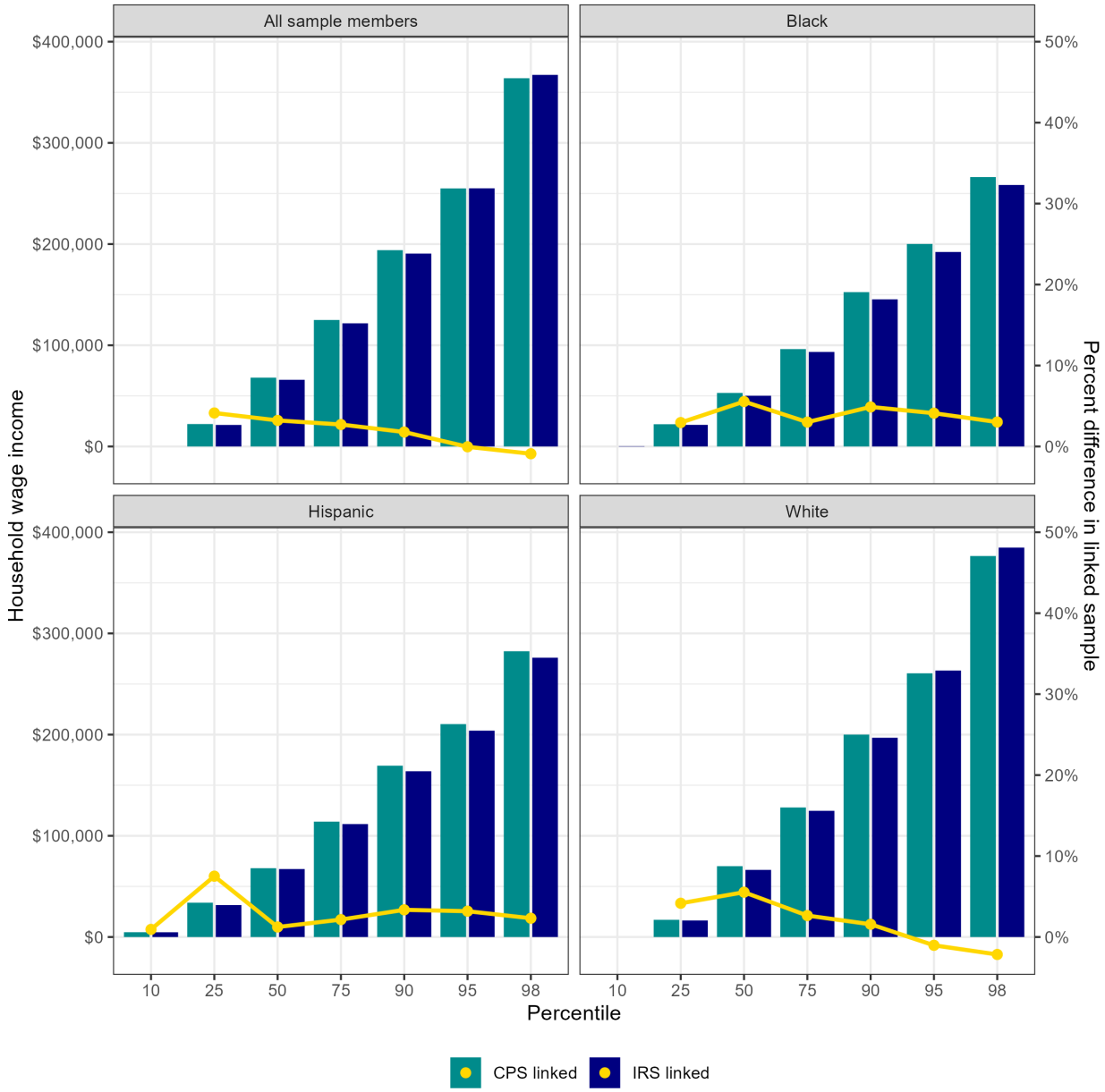

Figure B.8: Distribution of household wage and salary income in matched CPS-IDDA sample, 2019, by race and ethnicity

Source: IDDA and Current Population Survey.
Note: Figure charts selected percentiles of the distribution of household wage and salary income by race and ethnicity in 2019. The linked sample includes CPS ASEC respondents aged 16 or over who can be matched to the IDDA 1040 sample via the PIK, which restricts it to individuals listed as a primary or secondary filer on a Form 1040 filed for tax year 2019. Release authorization CBDRB-FY23-0373.

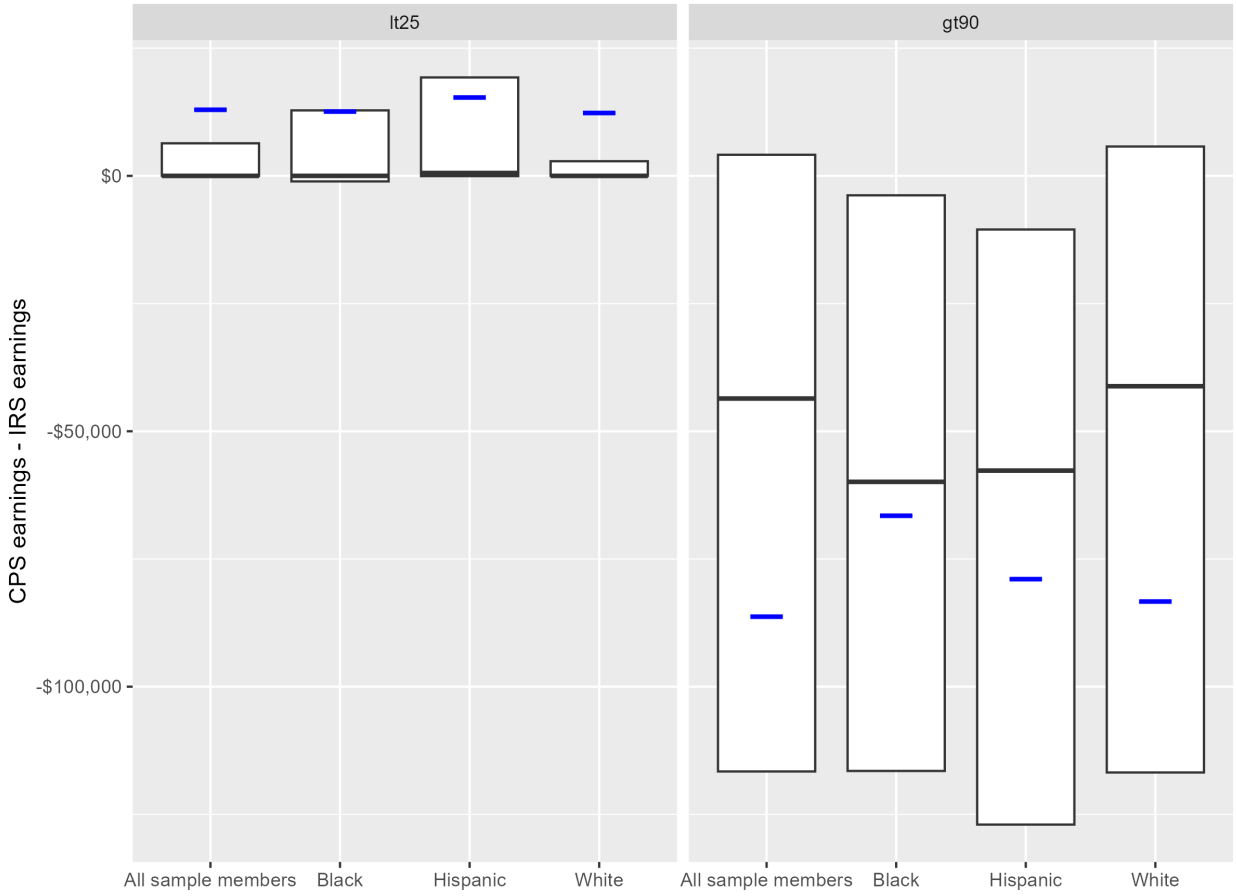

Figure B.9: Household earnings differences by race and ethnicity

Source: IDDA and Current Population Survey.
Note: Boxplots show the 25th, 50th, and 75th percentiles of earnings differences between household wage and salary income reported in the CPS and IDDA 1040 data. These are reported for Hispanic, non-Hispanic White, and non-Hispanic Black respondents in the bottom 25 percent of household wage and salary income (left) or the top 10 percent (right), defined using the 1040 household WSI measure in the linked sample. Release authorization CBDRB-FY23-0373.

References

Abraham, Katharine, John Haltiwanger, Claire Hou, Kristin Sandusky, and James Spletzer, "Reconciling Survey and Administrative Measures of Self-Employment," *Journal of Labor Economics*, 2021, *39* (4), 825–860.

_ , _ , Kristin Sandusky, and James Spletzer, "The Rise of the Gig Economy: Fact or Fiction?," *AEA Papers and Proceedings*, 2019, *109*, 357–361.

Bee, Adam and Joshua Mitchell, "Do Older Americans Have More Income Than We Think?," SEHSD Working Paper 2017-39, Social, Economic, and Housing Statistics Division, U.S. Census Bureau 2017.

Bollinger, Christopher R., Barry T. Hirsch, Charles M. Hokayem, and James P. Ziliak, "Trouble in the Tails? What We Know about Earnings Nonresponse 30 Years after Lillard, Smith, and Welch," *Journal of Political Economy*, 2019, *127* (5), 2143–2185.

Chenevert, Rebecca L., Mark A. Klee, and Kelly R. Wilkin, "Do Imputed Earnings Earn Their Keep? Evaluating SIPP Earnings and Nonresponse with Administrative Records," SIPP Working Paper No. 275, Social, Economic, and Housing Statistics Division, U.S. Census Bureau 2015.

Derenoncourt, Ellora, Chi Hyun Kim, Moritz Kuhn, and Moritz Schularick, "Changes in the Distribution of Black and White Wealth since the US Civil War," *Journal of Economic Perspectives*, 2023, *37* (4), 71–89.

Flood, Sarah, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren, and Michael Westberry, *Integrated Public Use Microdata Series, Current Population Survey: Version 11.0 [dataset]* Minneapolis, MN: IPUMS 2023. https://doi.org/10.18128/D030.V11.0.

Gideon, Michael, Misty L Heggeness, Marta Murray-Close, and Samuel L Myers, "The Returns to Education and the Determinants of Race and Gender Earnings Gaps Using Self-Reported versus Employer-Reported Earnings Data," *Journal of Income Distribution*, 2022.

Hokayem, Charles, Christopher Bollinger, and James P Ziliak, "The Role of CPS Nonresponse in the Measurement of Poverty," *Journal of the American Statistical Association*, 2015, *110* (511), 935–945.

Imboden, Christian, John Voorheis, and Caroline Weber, "Self-Employment Income Reporting on Surveys," CES Working Paper 23-19, Center for Economic Studies, US Census Bureau 2023.

Kim, ChangHwan and Christopher R Tamborini, "Do Survey Data Estimate Earnings Inequality Correctly? Measurement Errors among Black and White Male Workers," *Social Forces*, 2012, *90* (4), 1157–1181.

Meyer, Bruce and Nikolas Mittag, "Combining Administrative and Survey Data to Improve Income Measurement," NBER Working Paper No. 25738, National Bureau of Economic Research 2019.

Meyer, Bruce D, Wallace KC Mok, and James X Sullivan, "Household Surveys in Crisis," *Journal of Economic Perspectives*, 2015, *29* (4), 199–226.

Murray-Close, Marta and Misty L Heggeness, "Manning Up and Womaning Down: How Husbands and Wives Report Earnings When She Earns More," Institute Working Paper 28, Opportunity & Inclusive Growth Institute, Federal Reserve Bank of Minneapolis 2019.

Roemer, Marc I, "Assessing the Quality of the March Current Population Survey and the Survey of Income and Program Participation Income Estimates, 1990–1996," Working Paper, Housing and Household Economic Statistics Division, U.S. Census Bureau 2000.

Tamborini, Christopher R. and ChangHwan Kim, "Are Proxy Interviews Associated With Biased Earnings Reports? Marital Status and Gender Effects of Proxy," *Social Science Research*, 2013, *42* (2), 499–512.