Tax Policy Center
Urban Institute and Brookings Institution

A New Approach for Estimating the Impact of Tax Policies by Race and Ethnicity

Surachai Khitatrakun, Gordon Mermin, Benjamin Page, and Jeffrey Rohaly

February 10, 2023

ABSTRACT

This working paper describes the method the Tax Policy Center (TPC) has used to enhance its tax microsimulation model to enable analysis of the distributional effects of tax policies by race and ethnicity. It also presents preliminary results using the enhanced model. Like microsimulation models used in the federal government and other research organizations, TPC's model relies on a sample of individual tax returns as its primary data source. These data lack information on race and ethnicity because the IRS does not collect it from tax filers. TPC therefore developed a method to impute race and ethnicity in its tax data. The method replicates the tax units in the model database, with one copy for each race and ethnicity included in the analysis. Using an algorithm, those copies are then weighted to match aggregate statistics calculated from survey data for variables related to tax liability. TPC's initial work matches statistics calculated from the Current Population Survey and the Survey of Consumer Finances. TPC's method has the advantage that it can incorporate data from multiple surveys or other sources, and can be readily adapted to include updated or additional data. Preliminary results using the enhanced model indicate that, across all income categories, itemized deductions benefit tax units classified as White more than those classified as Black or Hispanic. Within the same income categories, tax units classified as Hispanic benefit less than those classified as Black or White.

The US tax code can affect racial income and wealth disparities because of longstanding discrimination in areas such as housing, education, and employment.[1] Taxes may also have disparate racial impacts due to differences in administration and enforcement reflecting biases embedded in the social fabric of the US (Dean 2021). In recognition of the US tax system's powerful potential role in advancing racial equity, President Biden's day-one executive order creating a new high-level Equitable Data Working Group explicitly named the US Treasury Department's Assistant Secretary for Tax Policy as a member.[2]

Measuring the effect of the tax system on racial equity is challenging, however, because the most complete information on tax liability comes from tax return data, which do not include the race or ethnicity of the taxpayer. Investigating the racial implications of tax policy therefore requires supplemental information, such as survey data that identify the race and ethnicity of the respondent and, in some cases, family members. Most research to date examining race and the federal tax system relies on data from such surveys. However, no single survey encompasses all the information required to analyze the effects of the full range of policies included in the tax code. In addition, survey data on incomes may be less reliable than tax return data.

The Treasury Department recently released results from its efforts to impute race and ethnicity to its model assessing the revenue and distributional effects of current and proposed tax policies.[3] Multiple other federal agencies are engaged in similar projects. Concurrently, thanks to generous seed funding from the Bill and Melinda Gates Foundation, the Urban-Brookings Tax Policy Center (TPC) has partnered with the Urban Institute's Racial Equity Analytics Lab to explore various approaches to adding race and ethnicity to TPC's tax microsimulation model.

This brief describes results from an initial implementation of TPC's preferred method and next steps to refine it so that we will be able to analyze the racial equity implications of all major tax policies. Our approach differs from others being pursued in government and those we know of in other public policy research organizations. It involves replicating each tax unit in the TPC tax model, with one copy for each racial and ethnic category included in the estimation. Each of the resulting units is assigned a race or ethnicity. Those units are then weighted, with the weights calculated so that the aggregate values for various statistics by race and ethnicity match values derived from household surveys while preserving the original model's weighted totals for the overall population.

The reweighted model can be used to estimate tax policy impacts by race and ethnicity. Because the original tax unit weights are effectively split among different races and ethnicities, we refer to this as the weight-splitting strategy. One could view the share of the original weight assigned to a replicated tax unit of a given race or ethnicity as analogous to the probability that the original unit is of that race or ethnicity, given its values for target variables (such as income and family structure) and the correlations between those variables and race or ethnicity in the survey data.

This brief details our initial effort to gauge the practicality and usefulness of the weight-splitting strategy by using data from two surveys from which we already derive important supplemental information for the TPC tax model—the Annual Social and Economic Supplement of the Current Population Survey (CPS), often called the March CPS, and the Survey of Consumer Finances (SCF)—to impute race and ethnicity identification for calendar year 2019, the year of the most recent SCF.

We begin with an overview of the TPC tax model, the CPS, and the SCF. We then discuss how we used information from the CPS and SCF to construct the aggregate targets by race and ethnicity that we sought to match in the TPC tax model (as adjusted with replicated tax units). Next, we describe how we imputed race and ethnicity weights to the TPC tax model database in order to match those targets. To illustrate the capabilities of the resulting enhanced model, we estimate the effect on after-tax incomes of all itemized deductions taken together by race and ethnicity. Finally, we discuss potential refinements to our imputation method.

BACKGROUND: THE TAX POLICY CENTER MICROSIMULATION TAX MODEL

TPC has a long history of using its microsimulation model to analyze the effects of tax policy proposals on not only the entire population but also certain groups. We routinely produce distributional estimates that show the impact of tax policies on households by marital status and presence of children as well as on taxpayers who are 65 or older. Expanding the capability of our existing model to analyze the impact of tax policy proposals on households of different races and ethnicities is a natural extension of this work.

The version of the TPC tax model used in this exercise (TM21) produces revenue and distributional estimates for each year from 2011 to 2032 (covering the 10-year budget window starting in 2022).[4] In addition, TPC's long-run module provides estimates at 10-year intervals from 2040 through 2090. Another module calculates how federal tax policy changes affect taxpayers in each state for the current year and a few selected years within the 10-year budget window.

The model's primary data source is the 2006 public-use file produced by the Statistics of Income Division of the Internal Revenue Service. The 2006 public-use file contains 145,858 records with detailed information from federal individual income tax returns for tax years 2003 through 2006.

We add information on other demographic characteristics and sources of income that are not reported on tax returns through a constrained statistical match with data from the US Census Bureau's March CPS. That match also generates a sample of individuals who do not file individual income tax returns (nonfilers).

We then augment the TPC tax model by turning to other data sources to develop imputations for supplemental variables (including wealth, education, consumption, health insurance, retirement savings, and other variables) that are then applied to each record in the matched public-use–CPS files. Those imputations allow us to analyze a wide variety of policy proposals. Finally, to extend the database to more recent and future years, we "age" the data by using information from published tax data as well as projections from various sources.[5]

TPC'S METHOD FOR MODELING RACE AND ETHNICITY

For the TPC tax model to reliably estimate the impact of tax policy by race and ethnicity, the key determinants of tax liability must vary across race and ethnicity in the same way as they do in the population. We considered the following approaches in the early stages of this work.

At one extreme, TPC analyses would rely on the existing race and ethnicity identifiers in the TPC tax model database, which we developed as part of the statistical match with the CPS. A limitation of this approach is that we imputed those identifiers using only a handful of variables, such as gross income and basic family structure. As a result, this method would likely provide reliable estimates in the case of policies for which the impact on tax liability depends primarily on that limited set of variables. It would not, however, fully capture the impact of differences across race and ethnicity of many other determinants of tax liability that we do not control for in the statistical match, such as home mortgage interest payments or capital gains realizations.

At another extreme, we would reconstruct the TPC tax model database by imputing race and ethnicity identifiers to the public-use file before matching the resulting data with CPS and subsequently imputing wealth, education, consumption, health, retirement, and other variables. However, this reconstruction could only be done through a full overhaul of the TPC tax model, which would be time consuming and costly. It would also make it difficult to incorporate additional information on racial differences from new data sources over time.

A middle ground would be to revise the race and ethnicity identifiers in the TPC tax model database based on a comprehensive set of variables. With these revised identifiers, the model could be used to estimate impacts across races and ethnicities for a wide range of tax policies. The weight-splitting approach falls within this set of alternatives.

After considering these alternatives, we chose the strategy presented here as the best option for the initial modeling efforts because the weight-splitting approach offers the following advantages:

- It does not require changes to other aspects of the existing TPC tax model.
- It incorporates information on relationships between key determinants of tax liabilities and races and ethnicities from more than one survey in a straightforward way.
- It allows information from additional surveys and other information sources to be added as needed.

This strategy takes the existing TPC tax model as given but replicates every tax unit into several units, with each of the copies representing a race or ethnicity but otherwise remaining identical to the original unit. The SCF data we analyzed include identifiers for only four categories of race and ethnicity: White Non-Hispanic, Black Non-Hispanic, Hispanic, and Others (which includes people identifying themselves with more than one race). Due to that data limitation, we could include only those categories in our analysis. For simplicity, we refer to the categories as White, Black, Hispanic, and Other in the remainder of the brief.

Weights were then estimated for the resulting tax units such that the weighted totals of selected demographic, income, tax deduction, and wealth variables in the TPC tax model by race and ethnicity closely matched comparable totals derived from relevant surveys, and the weight across races and ethnicities for each replicated tax unit equaled the TPC tax model weight for the original tax unit.[6] The enhanced model can be used to estimate tax policy impacts by race and ethnicity.

The first step in implementing the weight-splitting strategy was to construct statistics by race and ethnicity from survey data to use as targets in estimating weights. Constructing the target statistics was challenging because the two household surveys used in this project—the CPS and SCF—and the TPC tax model database differ in both the unit of analysis and in distributions of relevant variables, such as demographics and income.

HOUSEHOLD SURVEY DATA USED TO DERIVE TARGETS

Various household surveys potentially offer information on both race and tax variables. Surveys vary in both whose race is identified and what races and ethnicities are included (table 1). For this initial stage of our research, we chose to create targets using data from two of those household surveys: the March CPS and the SCF. Both surveys are used extensively by tax policy analysts and are, in many ways, complementary. The CPS contains detailed information on household composition and the income and transfers received by low- and middle-income households. The SCF oversamples high-income households and contains information on capital income and asset holdings that is missing from the CPS.

The CPS is administered by the US Census Bureau. Approximately 60,000 households are interviewed every March for the CPS's Annual Social and Economic Supplement. The household respondent is asked extensive questions on demographic characteristics and the income of each member of the household. The survey also includes detailed questions on race and ethnicity. Respondents can identify themselves and each member of their household as American Indian or Alaska Native, Asian, Black, Hawaiian or other Pacific Islander, or white. They may also choose to state more than one race. In a separate question, respondents are asked whether they and members of their household identify as Hispanic or Latino. In other questions, Asian and Hispanic respondents can provide more detailed information on their ethnicity (e.g., whether they identify as Chinese or Japanese). We used the 2020 March CPS because it included questions about income in 2019, the year of analysis for this brief.

Table 1
Within the Survey Unit, Whose Race or Ethnicity Is Identified?

Survey	Surveyed unit	Who Is Identified: Respondent	Spouse (if married)	Others in family	Others in residence but not in family
Current Population Survey	3 tiers: person, family, household	✓	✓	✓	✓
American Community Survey	3 tiers: person, family, household	✓	✓	✓	✓
Survey of Income and Program Participation	3 tiers: person, family, household	✓	✓	✓	✓
Survey of Consumer Finances	Primary economic unit	✓			
Consumer Expenditure Survey	Consumer unit	✓	✓	✓	✓
Panel Survey of Income Dynamics	Family unit	✓	✓		
National Survey of Early Care and Education	Household with at least one child under 13	✓		✓	
National Postsecondary Student Aid Study	Student	✓			
Health and Retirement Study	Household and respondent tiers; some family data	✓			
American Housing Survey	Household	✓			
Home Mortgage Disclosure Act	Loan applicant	✓			

The SCF is conducted every three years by the Federal Reserve Board, with the survey administered by NORC at the University of Chicago. The most recent survey was fielded in 2019 and included about 5,800 respondents. Given this small sample size, we choose to combine data from the three most recent surveys, conducted in 2013, 2016, and 2019.[7]

The survey data include information on families' balance sheets, pensions, income, and demographic characteristics. The SCF oversamples high-income households and therefore is a good source of information about certain tax-relevant variables that are concentrated among those households, such as income from capital gains. The SCF collects data on race and ethnicity only for the reference person in the survey. In the main data sample, these individuals are classified as Black, White, Hispanic, or Other. Given that restriction, TPC's weighting scheme includes only those four categories. For consistency, we classified tax units constructed from the CPS using the same four categories despite the fact that the CPS contains a more extensive list of racial classifications.[8]

CONVERTING THE UNIT OF ANALYSIS IN THE SCF AND CPS TO TAX UNITS

To create targets from the CPS and SCF for use with TPC's tax model, we first constructed consistent units for analysis across the datasets. For analyzing tax policy, the best unit of analysis is the tax unit. A tax unit is an individual or a married couple who file a tax return—or would file a tax return if they were required to do so (e.g., their income was above a specified threshold)—along with all dependents of that individual or married couple. Because the TPC tax model is based largely on tax return data, its unit of observation is the tax unit. However, we needed to create tax units using the available information for respondents in the CPS and SCF. Creating those tax units involved a series of decisions about how to allocate individuals from households in the surveys into tax units. A possible topic for further research is to explore the sensitivity of our results to alternative allocations.

Current Population Survey

The CPS is a survey of households and provides variables for analysis at the household, family, and individual levels. Often, however, those categories do not coincide with the groupings of household members who would file taxes together. Fortunately, the CPS collects information on household members, including income and family/household relationships, which allows for detailed construction of tax units. For simple households, such as married parents with children under age 18, grouping individuals into tax units is simple as all household members are in the same tax unit. For more complex households, such as multigenerational families, families living with extended relatives, or unmarried parents, determining who would file a tax return together can be complicated.[9] For each household we proceeded as follows (for further information on the methodology see Rohaly, Carasso, and Saleem 2005):

- We created a tax unit for the household respondent. If there was a spouse in the household, we added the spouse to the tax unit and assigned them as married filing jointly for their filing status. Any dependent children or unmarried and childless relatives with limited incomes were added to the tax unit as dependents.[10]
- We examined the rest of the members of the household in turn, and for each one not already assigned to a tax unit, we created a new tax unit with that household member treated as the respondent for that tax unit.[11] We then combined that new respondent with any spouse or dependent children.
- After cycling through all household members, we had an initial set of constructed tax units for the household. To allow for the possibility that some of the initial tax units may have had incomes low enough that they might be dependents of another tax unit, we first identified the highest-income tax unit in the household. We then assigned to that unit as dependents any unmarried tax units with incomes below $4,200 who were either related to the highest-income tax unit or had no relatives in the household.[12]
- Finally, we assigned head of household filing status to most single tax units with dependents.[13]

Our next step was to assign race and ethnicity and age to each tax unit based on the characteristics of the tax unit respondent. Although the CPS has information about the race of others in the tax unit, the race of the respondent was used to classify the entire unit to conform with data from the SCF, which has information on the race and ethnicity of only the respondent. We calculated tax unit income by summing individual-level income variables across members of the tax unit.[14] We assigned tax unit weight as the average of individual-level CPS weights across members of the tax unit. The tax unit algorithm mapped individuals representing 128 million households into 166 million tax units. Compared to the tax model database, there are about 5 percent fewer constructed CPS tax units (table 2). The CPS tax units are 6 percentage points more likely to be married than tax model units. Average tax unit age and average number of dependents are very similar across the datasets.

Survey of Consumer Finances

The unit of analysis in the SCF is the primary economic unit (PEU), which consists of the "economically dominant single individual or couple (married or living as partners) in a household and all other individuals in the household who are financially interdependent with that individual or couple."[15] Tax units and SCF households can diverge for several reasons. First, unmarried couples classified in the SCF as living as partners would constitute a single PEU, but the partners would have to file their own tax returns. In addition, members of an SCF household with independent finances are not included in the PEU. These individuals would, however, file their own tax returns and therefore constitute one or more additional tax units. In both cases, one SCF household, and therefore one PEU, is associated with multiple tax units.

Table 2
Number of Tax Units and Demographic Characteristics in CPS and Tax Model Database

	2020 CPS	2019 Tax Model Database
Number of tax units (million)	166	174
Married (%)	41.75	36.36
Tax units with dependents (%)	30.26	30.67
Elderly (%)	23.35	24.33
Number of dependents (mean)	0.56	0.55
Age (mean)	48.96	49.25

Sources: 2020 March CPS and Urban-Brookings Tax Policy Center Microsimulation Model (version 0721-2).

Notes: CPS = Current Population Survey. Tax units are classified by characteristics of tax unit respondent. We used the 2020 CPS because it collected information on income in 2019.

Table 3
Number of Tax Units in the TPC Tax Model Database and Primary Economic Units in the SCF (thousands)

Year	Single[a]	Married	Living as Partners	All
2019				
TMDB[b]	110,775	63,294	n/a	174,069
SCF	70,584	58,059	12,829	128,642
2016				
TMDB[b]	105,568	62,020	n/a	167,588
SCF	66,893	59,089	11,983	125,982
2013				
TMDB[b]	101,081	60,718	n/a	161,799
SCF	64,730	57,800	11,211	122,530

Sources: Urban-Brookings Tax Policy Center Microsimulation Model (version 0721-2) and authors' calculations based on the 2013, 2016, and 2019 SCF.

Notes: TPC = Tax Policy Center; SCF = Survey of Consumer Finances; TMDB = tax model database.
n/a = not applicable.
(a) Includes singles, heads of household, and married individuals filing a separate return.
(b) Excludes dependent filers and records representing the Forbes 400.

The upper two rows of table 3 compare the counts of nondependent tax units in the 2019 TPC tax model database with the number of PEUs in the 2019 SCF. The table shows that the tax model database contains about 45 million more nondependent tax units than the SCF PEUs for 2019. We focused on tax units who are not claimed as a dependent on another tax return for two reasons. First, TPC's distribution tables exclude dependents who file a tax return, and we therefore did not assign race or ethnicity categories to dependent filers in the tax model database.[16] Second, it is not possible to properly identify dependent filers in the SCF. Members of the PEU who are financially dependent on the reference person (and partner, if applicable) could work and have earnings that would require them to file a tax return. But the nature of the SCF questions about earnings prevents the identification of what fraction, if any, of the earnings of the PEU could be attributed to members other than the reference person or partner.

To create tax units from the SCF household data, we began with a program developed by Kevin Moore, an economist at the Federal Reserve.[17] Moore's program creates tax units in the SCF in order to calculate individual income taxes using TAXSIM.[18] We modified Moore's program by adding categories of assets and debts and certain other items that Moore does not model, such as student loan interest, to the resulting database (Gale et al. 2022).

As suggested above, we made two major changes to the SCF data to better align it with the tax unit data in our model. First, we split the PEUs consisting of an unmarried couple living together as partners into two tax units.[19] We did the same for married couples who indicated they had been married for less than one year and filed separate returns for the previous year.[20] For variables such as earnings and business income, about which the SCF asks questions that allowed us to split income between the partners, we did so. We assumed other income items were split equally. We then calculated an overall income split between the partners and applied that to divide the amount for each asset and debt. For itemized deductions, we followed Moore and assigned itemized deductions to the partner who would benefit most from claiming those deductions.

Second, we created tax units from the non-PEU (NPEU) members of the SCF household. We used the relationship and marital status information provided for these individuals to form both single and married tax units. The SCF asks a limited number of questions about the income and net wealth of these NPEU household members. The survey provides an aggregate amount of wage income for all NPEU members and an aggregate amount for all other income. It also provides an indicator for the types of income included in that other income category for each household. Again, we followed Moore by assigning the wage income equally to all individuals under the age of 70. We then divided the aggregate amount of other income equally across the categories and across individuals. The one exception is that we assigned Social Security income only to those individuals age 62 or older. For wealth, we conducted a similar exercise.

The SCF does not collect the same amount of detailed data on asset holdings and debt of NPEUs as is asked of the PEU respondent. First, the holdings and debt for all NPEUs are combined. Second, details are provided only for select assets: vehicles, cash, and ownership share of the reference person's home. The remaining assets and debt are aggregated into one category. We split these amounts equally across categories when appropriate and then across each NPEU individual.

One important limitation of the SCF is that it asks for information about the race and ethnicity of the reference person only.[21] Because that is the only information available, we took the simplified approach of assigning that same response to the respondent's spouse or partner and their dependents (even though those individuals may identify as other races or ethnicities). In addition, we assigned the race or ethnicity of the reference person to each of the new tax units we created by splitting unmarried couples and separating out the NPEU members of the household.[22]

Table 4 shows the tax unit counts in the SCF and the TPC tax model database after the tax units were created. As the overall number of both single and married tax units still fell short of those in the tax model database, we increased the weights on the tax units created from the NPEU members in order to match tax unit counts from the tax model database.[23]

Table 4
Number of Tax Units in the TPC Tax Model Database and SCF (thousands)

Year	Single[a]	Married	All
2019			
TMDB[b]	110,775	63,294	174,069
SCF	107,014	58,359	165,373
2016			
TMDB[b]	105,568	62,020	167,588
SCF	103,449	58,454	161,903
2013			
TMDB[b]	101,081	60,718	161,799
SCF	95,846	58,023	153,869

Sources: Urban-Brookings Tax Policy Center Microsimulation Model (version 0721-2) and author's calculations based on the 2013, 2016, and 2019 SCF.
Notes: TPC = Tax Policy Center; SCF = Survey of Consumer Finances; TMDB = tax model database;
(a) Includes singles, heads of household, and married individuals filing a separate return.
(b) Excludes dependent filers and records representing the Forbes 400.

ALIGNING RACE AND ETHNICITY IN THE SCF AND CPS

After organizing the SCF and CPS into tax units, we compared the racial and ethnic composition across income classes and marital status in the two resulting datasets. Table 5 reveals some significant differences, particularly at the top of the income scale. For example, the CPS data include over twice as many Black and Hispanic households in the top 5 percent of the income distribution as does the SCF. Because a core goal of the CPS is to represent the demographic composition of the population, and because of its much larger sample size, we conformed our data to the racial composition in the CPS. Therefore, we reweighted the SCF tax units to match the distribution of tax units by race and ethnicity, marital status, and income in the CPS. Table 6 shows the adjustment factors we applied to the weights in the SCF.

Other Adjustments

The tax model database includes 400 observations meant to represent the Forbes 400, the 400 wealthiest individuals in the US. By design, the SCF specifically excludes the Forbes 400 from its survey. In addition, it is highly unlikely that the CPS captures these individuals. We therefore identified races and ethnicities for the Forbes 400 based on a public records search of the individuals named by Forbes for 2016.[24]

Finally, we excluded potential target variables when the CPS or the SCF had obvious limitations. For example, capital losses in the SCF differ quantitatively from capital losses in the tax model database. In addition, the SCF reports only an indicator of net business losses instead of an actual dollar value. As a result, we did not target dollar values of total losses.[25]

Creating Targets

We used the modified survey data from the CPS and SCF to calculate targets to be used in estimating race and ethnicity weights for the tax model database such that weighted totals would match the racial and ethnic distribution of the survey data. We calculated targets by marital status for five income percentile groups (0–25th, 25th–50th, 50th–75th, 75th–95th, and above 95th).[26] To create a target, we first calculated shares for an overall category for each race or ethnicity. For example, we calculated the share of married tax units in the 75th–95th income percentile classified as Black and the share of total capital gains income among unmarried tax units in the 25th–50th percentile received by tax units classified as Hispanic. We then applied those survey-based shares to the appropriate TPC tax model totals (such as the total number of married tax units in the TPC tax model database that fell in the 75th–95th percentile) to produce initial target values. In all, we matched over 2,800 targets for 79 variables in combination with four races and ethnicities, two marital statuses, and five income levels.

Table 5
Tax Unit Shares by Race and Income Percentile in the Current Population Survey and Survey of Consumer Finances

Married

Income Percentile[a]	CPS[b] White	CPS Black	CPS Hispanic	CPS Other	CPS All	SCF[c] White	SCF Black	SCF Hispanic	SCF Other	SCF All	Ratio of SCF to CPS White	Ratio Black	Ratio Hispanic	Ratio Other	Ratio All
0–25th	53.3	10.5	26.2	10.0	100.0	56.4	9.4	22.7	11.6	100.0	105.8	89.8	86.4	115.6	100.0
25th–50th	67.3	8.5	16.9	7.3	100.0	68.3	10.4	11.6	9.6	100.0	101.5	123.6	68.6	131.2	100.0
50th–75th	74.0	7.4	10.2	8.4	100.0	77.9	6.5	5.8	9.8	100.0	105.2	86.9	56.8	117.9	100.0
75th–95th	77.8	5.0	7.1	10.1	100.0	80.8	4.8	3.3	11.1	100.0	103.9	97.1	46.0	109.4	100.0
95th–100th	79.3	4.2	5.3	11.2	100.0	83.1	1.8	1.7	13.4	100.0	104.8	43.4	31.0	119.9	100.0
All	68.2	7.8	15.0	9.0	100.0	71.0	7.6	10.8	10.7	100.0	104.1	97.9	71.6	118.2	100.0

Single[d]

Income Percentile	CPS White	CPS Black	CPS Hispanic	CPS Other	CPS All	SCF White	SCF Black	SCF Hispanic	SCF Other	SCF All	Ratio of SCF to CPS White	Ratio Black	Ratio Hispanic	Ratio Other	Ratio All
0–25th	52.3	21.3	19.6	6.7	100.0	53.3	18.6	14.7	13.5	100.0	101.8	87.0	74.9	199.4	100.0
25th–50th	53.7	18.5	20.9	6.9	100.0	53.8	19.7	14.1	12.4	100.0	100.1	106.4	67.5	181.0	100.0
50th–75th	59.2	16.1	17.9	6.8	100.0	58.8	19.6	11.2	10.4	100.0	99.3	122.0	62.7	151.8	100.0
75th–95th	67.1	12.7	12.1	8.1	100.0	66.3	15.8	7.0	10.9	100.0	98.8	124.2	58.2	134.2	100.0
95th–100th	71.2	11.3	7.9	9.6	100.0	73.7	8.9	5.9	11.4	100.0	103.5	79.0	75.0	119.2	100.0
All	58.3	17.1	17.4	7.2	100.0	58.4	18.1	11.7	11.8	100.0	100.2	105.7	67.2	163.8	100.0

All Filing Statuses

Income Percentile	CPS White	CPS Black	CPS Hispanic	CPS Other	CPS All	SCF White	SCF Black	SCF Hispanic	SCF Other	SCF All	Ratio of SCF to CPS White	Ratio Black	Ratio Hispanic	Ratio Other	Ratio All
0–25th	52.7	16.8	22.4	8.1	100.0	52.4	18.3	15.7	13.6	100.0	99.4	109.1	70.1	167.4	100.0
25th–50th	59.3	14.3	19.3	7.1	100.0	57.3	17.5	14.1	11.1	100.0	96.5	122.5	73.1	157.0	100.0
50th–75th	65.4	12.4	14.7	7.5	100.0	64.9	14.0	10.7	10.3	100.0	99.2	112.9	72.9	138.3	100.0
75th–95th	71.6	9.5	10.0	9.0	100.0	76.0	8.1	5.8	10.1	100.0	106.2	86.0	57.7	112.7	100.0
95th–100th	74.6	8.3	6.8	10.3	100.0	83.7	2.2	1.8	12.4	100.0	112.2	26.1	25.8	120.9	100.0
All	62.4	13.2	16.4	8.0	100.0	63.0	14.2	11.4	11.4	100.0	101.0	107.7	69.2	143.0	100.0

Sources: Author's calculations based on the 2020 March supplement to the Current Population Survey and the 2013, 2016, and 2019 Surveys of Consumer Finances.
Notes: CPS = Current Population Survey; SCF = Survey of Consumer Finances.
(a) Income concept is the definition of income used in the CPS and therefore excludes realized capital gains.
(b) March 2020 CPS.
(c) Weighted average of the 2013 (20 percent), 2016 (30 percent), and 2019 (50 percent) SCF.
(d) Includes singles, heads of household, and married individuals filing a separate return.

Table 6

Weight Adjustment Factors to Align SCF with CPS Distribution of Race and Ethnicity by Income and Marital Status

	Married											
	2013				2016				2019			
Income Percentile[a]	**White**	**Black**	**Hispanic**	**Other**	**White**	**Black**	**Hispanic**	**Other**	**White**	**Black**	**Hispanic**	**Other**
0–25th	0.9859	0.9305	1.0625	1.0000	0.9784	1.1610	1.0997	0.7905	0.9121	1.1770	1.2413	0.8668
25th–50th	0.9441	0.9707	1.5228	0.8233	0.9766	0.9242	1.4459	0.7157	1.0074	0.7093	1.4400	0.7695
50th–75th	0.9262	1.4041	1.9730	0.8687	0.9649	1.0874	1.5983	0.8297	0.9515	1.1096	1.7908	0.8520
75th–95th	0.9461	1.0709	2.4723	0.9883	0.9487	0.9602	2.7900	0.9871	0.9782	1.0588	1.8403	0.8508
95th–100th	0.8674	9.7709	4.3346	1.6095	0.8974	2.7548	3.5045	1.2981	1.0342	1.6414	2.7987	0.5928

	Single[b]											
	2013				2016				2019			
Income Percentile	**White**	**Black**	**Hispanic**	**Other**	**White**	**Black**	**Hispanic**	**Other**	**White**	**Black**	**Hispanic**	**Other**
0–25th	1.0217	1.0722	1.4628	0.4356	1.0488	1.0816	1.1820	0.4887	0.9316	1.2310	1.3931	0.5427
25th–50th	0.9402	0.9873	1.4611	0.6996	1.0118	0.9135	1.4546	0.5590	1.0163	0.9382	1.5082	0.5063
50th–75th	0.9482	0.8356	1.8346	0.7953	1.0317	0.7854	1.4803	0.6781	1.0172	0.8353	1.5835	0.6064
75th–95th	0.9778	0.8353	1.5915	0.9455	1.0246	0.8157	1.6269	0.7042	1.0182	0.7881	1.8412	0.7099
95th–100th	0.8875	1.5288	2.2564	1.0782	1.0030	1.0313	1.1943	0.8385	0.9786	1.3573	1.2190	0.7709

Sources: Authors' calculations based on the 2020 March supplement to the Current Population Survey and the 2013, 2016, and 2019 Surveys of Consumer Finances.

Notes: SCF = Survey of Consumer Finances; CPS = Current Population Survey.

(a) Income concept is the definition of income used in the Current Population Survey and therefore excludes realized capital gains.

(b) Includes singles, heads of household, and married individuals filing a separate return.

Table 7 lists our targeted variables for each marital status and income subcategory. Targeted variables are either the number of returns with nonzero values for an item or the total amount of this item for tax year 2019. We targeted between 60 and 74 variables in each subgroup, including most of the items needed to calculate federal income taxes.[27] Targets included selected demographics, income, potential income tax deductions, and assets.

The sum of a given variable across all race and ethnicity categories in the TPC tax model generally differs from the corresponding total in a survey. However, to implement the weight-splitting strategy, the sum of targets across race and ethnicity for a given variable must match the TPC tax model total. To preserve the tax model totals, we created targets for each race and ethnicity in the TPC tax model by multiplying the shares for each race and ethnicity in the survey by the total in the tax model.

However, this method does not necessarily yield amounts within a racial or ethnic category that sum to the total for that race or ethnicity. To illustrate, suppose that in the TPC tax model database, 50 percent of all tax units in a subgroup have a pension plan. In addition, suppose that our basic demographic target from the CPS implies that tax units classified as Black should represent 5 percent of tax units in this subgroup. Denote *P* as the percentage of all tax units with a pension plan that is classified as Black in this subgroup in our SCF data, the source for our pension coverage targets. An inconsistency would arise if *P* were larger than 10 percent, because this would imply there were more tax units classified as Black with a pension plan than the total number of tax units classified as Black. To see this, suppose *P* is 15 percent. Applying this SCF share of 15 percent to the 50 percent of tax units with a pension plan in the TPC tax model would result in 7.5 percent of all tax units in this subgroup being classified as Black with a pension plan. But only 5 percent of tax units in this subgroup are classified as Black according to our demographic target. So simply applying shares in the way described above would mean that the target for the number of tax units classified as Black with a pension plan would exceed the total number of Black tax units—which clearly would be an impossible target to match.

To guarantee internal consistency across targets, we applied an iterative process to create race and ethnicity targets that summed to TPC tax model totals both within and across races and ethnicities. The algorithm recognizes that two constraints must be met simultaneously. First, for every race and ethnicity group, the number of tax units with and without, say, a pension plan, must equal the total number of all tax units. Second, the number of tax units with a pension plan across all races and ethnicities must sum to the number of tax units with a pension plan overall in the TPC tax model. (In a spreadsheet listing characteristics by race, this constraint amounts to ensuring that both rows and columns add up to the relevant totals.) These two constraints further imply that the number of tax units without a pension plan across all races and ethnicities must sum to the number of tax units without a pension plan overall in the TPC tax model.[28]

Given the resulting targets, we then calculated weights that, when applied to the tax units in the TPC tax model database, yielded a weighted sample that matched those targets by race and ethnicity.

Table 7
Targeted Items Used to Derive Race and Ethnicity Weights by Marital Status and Income Subgroup

Targeted Item	Target Type	Data Source	Adjusted Gross Income Percentile									
			Single with Income					Married with Income				
			0–25	25–50	50–75	75–95	95+	0–25	25–50	50–75	75–95	95+
Number of tax units												
All	Number of tax units	CPS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
With single or married filing separately tax filing status	Number of tax units	CPS	✓	✓	✓	✓	✓					
With head's age 26 under 35	Number of tax units	CPS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗
With head's age 35 under 45	Number of tax units	CPS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
With head's age 45 under 55	Number of tax units	CPS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
With head's age 55 under 65	Number of tax units	CPS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
With head's age 65 under 75	Number of tax units	CPS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
With head's age 75 and up	Number of tax units	CPS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
With two wage earners	Number of tax units	CPS						✓	✓	✓	✓	✓
With an undergraduate or graduate student	Number of tax units	CPS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
With either DB or DC pension from current job	Number of tax units	SCF	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
With DC pension from current job	Number of tax units	SCF	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
With a dependent	Number of tax units	CPS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Number of dependents												
Younger than 6	Number of People	CPS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Age 6 to 12	Number of People	CPS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Age 13 to 16	Number of People	CPS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Age 17	Number of People	CPS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Age 18 to 23	Number of People	CPS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
24 or older	Number of People	CPS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Number of taxpayers												
Age 26 to 44	Number of People	CPS						✓	✓	✓	✓	✓
Age 45 to 64	Number of People	CPS						✓	✓	✓	✓	✓
Age 65 or older	Number of People	CPS						✓	✓	✓	✓	✓
Income												
Wages and salaries	Number of tax units	CPS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Amount		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Taxable and tax-exempt interest	Number of tax units	CPS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Amount		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Dividend income	Number of tax units	CPS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Amount		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Schedule C income: gain	Number of tax units	SCF	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Amount		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Schedule C income: loss	Number of tax units	SCF	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Amount											
Net capital gain	Number of tax units	SCF	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Amount		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Net capital loss	Number of tax units	SCF	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Amount											
Total pensions and annuities	Number of tax units	CPS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Amount		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Schedule E: gain	Number of tax units	SCF	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Amount		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Schedule E: loss	Number of tax units	SCF	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Amount											
Unemployment compensation	Number of tax units	CPS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Amount		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gross Social Security benefits	Number of tax units	CPS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Amount		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cash income, narrow definition	Number of tax units	CPS										
	Amount		✓	✓	✓	✓	✗	✓	✓	✓	✓	✗

Table 7, Continued

Targeted Item	Target Type	Data Source	Adjusted Gross Income Percentile: Single with Income 0–25	25–50	50–75	75–95	95+	Married with Income 0–25	25–50	50–75	75–95	95+
Supplemental Security Income	Number of tax units	CPS	✓	✓	✓	✓	✗	✓	✓	✓	✓	✗
	Amount		✓	✓	✓	✓	✗	✓	✓	✓	✓	✗
Temporary Assistance for Needy Families	Number of tax units	CPS	✓	✓	✓	✓	✗	✓	✓	✓	✓	✗
	Amount		✓	✓	✓	✓	✗	✓	✓	✓	✓	✗
Veterans' benefits	Number of tax units	CPS	✓	✓	✓	✓	✗	✓	✓	✓	✓	✗
	Amount		✓	✓	✓	✓	✗	✓	✓	✓	✓	✗
Supplemental Nutrition Assistance Program	Number of tax units	CPS	✓	✓	✓	✓	✗	✓	✓	✓	✗	✗
	Amount		✓	✓	✓	✓	✗	✓	✓	✓	✗	✗
Potential income tax deductions												
Home mortgage interest	Number of tax units	SCF	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Amount		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Property taxes	Number of tax units	SCF	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗
	Amount		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Charitable contributions	Number of tax units	SCF	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Amount		✓	✗	✓	✓	✓	✓	✓	✓	✓	✓
Student loan interest	Number of tax units	SCF	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Amount		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Assets and debts												
Cash	Number of tax units	SCF	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
	Amount		✗	✓	✓	✓	✓	✗	✗	✓	✓	✗
Stocks	Number of tax units	SCF	✗	✗	✗	✗	✗	✗	✗	✓	✓	✗
	Amount		✗	✓	✓	✓	✓	✗	✓	✓	✓	✓
Houses	Number of tax units	SCF	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Amount		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Real estate	Number of tax units	SCF	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗
	Amount		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Houses and real estate	Number of tax units	SCF	✓	✓	✓	✓	✗	✓	✓	✗	✓	✗
	Amount											
Nonpassive business	Number of tax units	SCF	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓
	Amount		✗	✓	✓	✓	✓	✗	✓	✗	✓	✓
DC pension and individual retirement account assets	Number of tax units	SCF	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Amount		✓	✓	✓	✓	✓	✓	✗	✓	✓	✓
Total financial assets	Number of tax units	SCF	✓	✓	✗	✗	✗	✓	✗	✗	✗	✗
	Amount		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Total nonfinancial assets	Number of tax units	SCF	✓	✓	✓	✓	✗	✓	✓	✓	✓	✗
	Amount		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Total debts	Number of tax units	SCF	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Amount		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Net worth	Number of tax units	SCF	✓	✓	✗	✗	✗	✓	✗	✗	✗	✗
	Amount											
Number of targeted variables			71	73	72	72	60	74	73	74	74	60
Tolerance level (%)			12	5	12	5	15	5	5	5	15	12

Source: Urban-Brookings Tax Policy Center Microsimulation Model (version 0721-2) with preliminarily imputed race weights.

Notes: DB = defined benefit; DC = defined contribution; CPS = Current Population Survey; SCF = Survey of Consumer Finances. ✓ and ✗ denote variables that could and could not be targeted, respectively.

A gray cell denotes that we did not target this variable for the relevant subgroup. The reason may be that this variable could be calculated as a function of other targets' variables (e.g., net worth is defined as the sum of total financial and nonfinancial assets less total debts); that virtually every tax unit had such an item (e.g., number of tax units with some cash income); or that the source data lack such details (e.g., Schedule C and E losses).

For the tax model database, "single" includes tax units whose income tax filing statuses were single, head of household, or married filing separately; "married" includes tax units whose income tax filing statuses were married filing jointly.

IMPUTING WEIGHTS TO REPRESENT RACES AND ETHNICITIES

METHODOLOGY

Our methodology uses the relationship between target variables and race and ethnicity in survey data to assign weights to the tax units in the TPC tax model database. Every tax unit is replicated into four copies classified as Black, Hispanic, White, and Other. The weight on each copy can be thought of as an estimate of the probability that a tax unit belongs to a particular race or ethnicity, given the tax unit's characteristics (such as income, assets, marital status, and number of children), multiplied by the TPC tax model weight of the original, unsplit, tax unit. The weights on the copies are derived such that the totals of chosen TPC tax model variables by race and ethnicity reproduce their targeted totals derived from data sources containing race and ethnicity information. To estimate those weights, we adapted a constrained, parametric regression methodology proposed by Schirm and Zaslavsky (1997).

The methodology imposes two constraints. First, the summation of each TPC tax model observation's four race and ethnicity weights must be identical to its original weight. Second, the weighted totals of the explanatory variables in the regression specification must match their targeted totals within a specified tolerance level. The regression's parametric nature is specified with an identifying assumption that the race and ethnicity weights as a proportion of the original weights must be identical for all observations with similar characteristics. These constraints and the parametric specification together ensure a unique solution to the race and ethnicity weights (Schirm and Zaslavsky 2001, page 9).

To be specific, the parametric specification is a Poisson regression. Define W^h as observation h's original weight and w_s^h as h's weight for race or ethnicity s (s = 1 to S, where S is the number of categorized races and ethnicities), so that $\sum_s w_s^h = W^h$, $\underline{x}^h = \left(x_1^h, x_2^h, \dots, x_k^h, \dots x_K^h\right)$ is defined as h's observed characteristics, where K is the number of characteristics accounted for in the estimation; X_{ks} is the race and ethnicity s's weighted total of x_k (i.e., $X_{ks} = \sum_h w_s^h x_k^h$), $\underline{X}_s = (X_{1s}, X_{2s}, \dots, X_{ks}, \dots X_{Ks})$; δ^h is h's idiosyncratic constant; and $\underline{\beta}_s = (\beta_{1s}, \beta_{2s}, \dots, \beta_{ks}, \dots \beta_{Ks})$ is the race and ethnicity–specific coefficient estimates for race or ethnicity s. The resulting regression specification is shown in equation 1:

$$w_s^h = \exp\left(\underline{\beta}_s' \underline{x}^h + \delta^h\right) \tag{1}$$

Notice that the race and ethnicity share w_s^h / W^h only depends on h's characteristics $\underline{x}^h$ and does not depend on h's idiosyncratic parameter δ^h. This holds because $\exp\left(\underline{\beta}_s' \underline{x}^h + \delta^h\right) = \exp\left(\underline{\beta}_s' \underline{x}^h\right) \times \exp\left(\delta^h\right)$, so that $\exp\left(\delta^h\right)$ is cancelled out of the numerator and denominator. This regression specification, with the two constraints specified above, results in the unique set of race and ethnicity weights represented in equations 2 and 3:[29]

$$\sum_s w_s^h = W^h \quad (2)$$

$$\sum_h w_s^h x_k^h = X_{ks} \quad (3)$$

We estimated this constrained model by using a maximum likelihood, iterative two-step approach. Denote $\delta_{(i)}^h$ as the household-specific constant and $\underline{\beta}_{s(i)}$ as the race and ethnicity–specific coefficient estimates derived from the i^{th} iteration.[30] In the first step of the i^{th} iteration, we calculated $\delta_{(i)}^h$ by substituting equation 1 into equation 2:

$$\delta_{(i)}^h = ln\left(\frac{W^h}{\sum_s \exp\left(\underline{\beta}'_{s(i-1)}\underline{x}^h\right)}\right)$$

In the second step of the i^{th} iteration, we derived $\underline{\beta}_{s(i)}$ using a Newton-Raphson method. That is, based on $\underline{\beta}_{s(i-1)}$ and the remaining distances between the targeted and derived totals, $\underline{d}_s \equiv \underline{X}_s - \sum_h w_s^h \underline{x}^h$ with $\underline{d}_s = (d_{1s}, d_{2s}, \ldots, d_{ks}, \ldots d_{Ks})$:

$$\underline{\beta}_{s(i)} = \underline{\beta}_{s(i-1)} + D_{s(i)}^{-1}\underline{d}_{s(i)}$$

with the first-order partial derivative matrix $D_{s(i)} \equiv \sum_h w_{s(i)}^h \underline{x}^h \underline{x}^{h'}$, and $w_{s(i)}^h = \exp\left(\underline{\beta}'_{s(i-1)}\underline{x}^h + \delta_{(i)}^h\right)$.

The iterations continued until every difference (d_{ks} for all k and s) was within a prespecified tolerance level.[31] Finally, we calculated race and ethnicity weights using equation 1 with the derived coefficient estimates of δ^h and $\underline{\beta}_s$.

IMPLEMENTATION

To estimate Schirm and Zaslavsky's constrained model, we first stratified observations in the TPC tax model database for tax year 2019 into two groups based on marital status.[32] We further stratified observations within each marital status group into five income percentile groups (0–25th, 25th–50th, 50th–75th, 75th–95th, and above 95th) based on a measure of cash income in the TPC tax model that roughly corresponds to measures available in the CPS and SCF data.[33] We then performed the estimation for each of the resulting 10 marital status and income subgroups separately. Not only does the estimation by subgroup reduce the computational burden through a substantial reduction in the number of observations involved, but it also greatly improves the ultimate quality of the race and ethnicity–weighted TPC tax model database.[34]

Using the derived race and ethnicity targets, we estimated Schirm and Zaslavsky's constrained model for each of the 10 subgroups separately and imputed race and ethnicity weights for each subgroup's observations based on the subgroup's derived coefficient estimates. We set the tolerance level between 5 and 15 percent as shown in the bottom row of table 7. In theory, a larger tolerance allows more targets to be accommodated, but potentially at a cost of less precision. In practice, table 8 shows that in most cases the race and ethnicity–weighted totals matched their respective targets more closely than the specified totals. Specifically, out of 2,812 targets, the weighted race and ethnicity totals deviated from their respective targeted totals by more than 2 percent for just 3.6 percent of all targets (101 targets in total).[35] In addition, this imprecision is at the subgroup level and does not result in similar imprecision for the combined targeted totals of all 10 marital status and income subgroups.[36] The next section assesses whether the TPC tax model produces reasonable estimates of tax burden across races and ethnicities when using the race and ethnicity–weighted TPC tax model database.

Table 8
Number of Variables Whose Race and Ethnicity–Weighted Totals Deviate from Their Targeted Totals by More Than 2 Percent

Targeted Item	Target Type	Adjusted Gross Income Percentile										All
		Single with Income					Married with Income					
		0–25	25–50	50–75	75–95	95+	0–25	25–50	50–75	75–95	95+	
Number of tax units												
All	Number of tax units	0	0	0	0	0	0	0	0	0	0	0
With single or married filing separately tax filing status	Number of tax units	0	0	0	0	0	n/a	n/a	n/a	n/a	n/a	0
With head's age 26 under 35	Number of tax units	0	0	0	0	0	0	0	0	0	n/a	0
With head's age 35 under 45	Number of tax units	0	0	0	0	0	0	0	0	0	0	0
With head's age 45 under 55	Number of tax units	0	0	0	0	0	0	0	0	0	0	0
With head's age 55 under 65	Number of tax units	0	0	0	0	0	0	0	0	0	0	0
With head's age 65 under 75	Number of tax units	0	0	0	0	0	0	0	0	0	1	1
With head's age 75 and up	Number of tax units	0	0	0	0	0	0	0	0	1	0	1
With two wage earners	Number of tax units	n/a	n/a	n/a	n/a	n/a	0	0	0	0	0	0
With an undergraduate or graduate student	Number of tax units	0	0	0	0	0	0	0	0	0	0	0
With either DB or DC pension from current job	Number of tax units	0	0	0	0	0	0	0	0	0	0	0
With DC pension from current job	Number of tax units	0	0	0	0	0	0	0	0	0	0	0
With a dependent	Number of tax units	0	0	0	0	0	0	0	0	0	0	0
Number of dependents												
Younger than 6	Number of People	0	0	0	0	0	0	0	0	0	0	0
Age 6 to 12	Number of People	0	0	0	0	0	0	0	0	0	0	0
Age 13 to 16	Number of People	0	0	0	0	0	0	0	0	0	0	0
Age 17	Number of People	0	0	0	0	0	0	0	0	0	0	0
Age 18 to 23	Number of People	0	0	0	0	0	0	0	0	0	0	0
24 or older	Number of People	0	0	0	0	0	0	0	0	0	0	0
Number of taxpayers												
Age 26 to 44	Number of People	n/a	n/a	n/a	n/a	n/a	0	0	0	0	0	0
Age 45 to 64	Number of People	n/a	n/a	n/a	n/a	n/a	0	0	0	0	0	0
Age 65 or older	Number of People	n/a	n/a	n/a	n/a	n/a	0	0	0	1	1	2
Income												
Wages and salaries	Number of tax units	0	0	0	0	0	0	0	0	0	0	0
	Amount	0	0	0	0	0	0	0	0	0	0	0
Taxable and tax-exempt interest	Number of tax units	0	0	0	0	0	0	0	0	0	0	0
	Amount	0	0	2	0	2	0	0	0	1	0	5
Dividend income	Number of tax units	0	0	0	0	0	1	0	0	0	0	1
	Amount	0	0	1	0	2	1	1	0	1	1	7
Schedule C income: gain	Number of tax units	0	0	0	0	0	0	0	0	1	0	1
	Amount	0	0	0	0	0	0	0	0	1	0	1
Schedule C income: loss	Number of tax units	0	0	0	0	0	0	0	0	0	0	0
	Amount	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0
Net capital gain	Number of tax units	0	0	2	0	0	0	0	0	1	0	3
	Amount	1	0	1	0	2	0	0	1	1	1	7
Net capital loss	Number of tax units	0	0	2	0	0	1	0	0	0	0	3
	Amount	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0
Total pensions and annuities	Number of tax units	0	0	0	0	0	0	0	0	0	0	0
	Amount	2	0	1	0	0	0	0	1	0	1	5
Schedule E: gain	Number of tax units	0	1	1	0	0	0	1	0	1	0	4
	Amount	3	0	1	1	0	1	1	0	1	0	8
Schedule E: loss	Number of tax units	0	0	0	0	1	0	0	0	0	1	2
	Amount	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0
Unemployment compensation	Number of tax units	0	0	0	0	0	0	0	0	0	0	0
	Amount	0	0	0	0	0	0	0	0	0	0	0
Gross Social Security benefits	Number of tax units	0	0	0	0	0	0	0	0	1	0	1
	Amount	0	0	0	0	0	0	0	0	1	1	2
Cash income, narrow definition	Number of tax units	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0
	Amount	0	0	0	0	n/a	0	0	0	0	n/a	0

Table 8, continued

Targeted Item	Target Type	Adjusted Gross Income Percentile: Single with Income 0–25	25–50	50–75	75–95	95+	Married with Income 0–25	25–50	50–75	75–95	95+	All
Nontaxable transfers												
Supplemental Security Income	Number of tax units	0	0	0	0	n/a	0	0	0	1	n/a	1
	Amount	0	0	0	0	n/a	1	0	0	1	n/a	2
Temporary Assistance for Needy Families	Number of tax units	0	0	0	0	n/a	0	0	0	0	n/a	0
	Amount	0	0	0	0	n/a	0	1	0	0	n/a	1
Veterans' benefits	Number of tax units	0	0	0	0	n/a	0	0	0	0	n/a	0
	Amount	0	0	0	0	n/a	0	0	0	0	n/a	0
Supplemental Nutrition Assistance Program	Number of tax units	0	0	0	0	n/a	0	0	0	n/a	n/a	0
	Amount	0	0	0	0	n/a	0	0	0	n/a	n/a	0
Potential income tax deductions												
Home mortgage interest	Number of tax units	0	0	0	0	0	0	0	0	0	0	0
	Amount	0	0	0	0	0	0	0	0	0	0	0
Property taxes	Number of tax units	0	0	0	0	0	0	0	0	0	n/a	0
	Amount	0	0	0	0	0	0	0	0	1	0	1
Charitable contributions	Number of tax units	0	0	0	0	0	0	0	0	0	0	0
	Amount	0	n/a	0	0	0	0	0	0	0	0	0
Student loan interest	Number of tax units	0	0	0	0	0	1	0	0	0	0	1
	Amount	0	0	0	0	0	1	0	0	0	0	1
Assets and debts												
Cash	Number of tax units	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0
	Amount	n/a	0	1	0	0	n/a	n/a	0	1	n/a	2
Stocks	Number of tax units	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0	0	n/a	0
	Amount	n/a	2	3	0	0	n/a	0	2	1	1	9
Houses	Number of tax units	0	0	0	0	0	0	0	0	0	0	0
	Amount	0	0	0	0	0	0	0	0	1	0	1
Real estate	Number of tax units	0	0	0	0	0	0	0	0	1	n/a	1
	Amount	3	2	2	0	0	0	2	0	1	0	10
Houses and real estate	Number of tax units	0	0	0	0	n/a	0	0	n/a	0	n/a	0
	Amount	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0
Nonpassive business	Number of tax units	0	0	0	0	n/a	0	0	0	0	0	0
	Amount	n/a	1	2	1	0	n/a	1	n/a	1	0	6
DC and IRA assets	Number of tax units	0	0	0	0	0	0	0	0	0	0	0
	Amount	0	0	0	0	0	0	n/a	0	0	0	0
Total financial assets	Number of tax units	0	0	n/a	n/a	n/a	0	n/a	n/a	n/a	n/a	0
	Amount	0	0	2	0	0	0	0	1	1	1	5
Total nonfinancial assets	Number of tax units	0	0	0	0	n/a	0	0	0	0	n/a	0
	Amount	2	0	1	0	0	0	0	0	1	0	4
Total debts	Number of tax units	0	0	0	0	0	0	0	0	0	0	0
	Amount	0	1	0	0	0	0	0	0	1	0	2
Net worth	Number of tax units	0	0	n/a	n/a	n/a	0	n/a	n/a	n/a	n/a	0
	Amount	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0
All targets		**11**	**7**	**22**	**2**	**7**	**7**	**7**	**5**	**24**	**9**	**101**

Source: Urban-Brookings Tax Policy Center Microsimulation Model (version 0721-2) with preliminarily imputed race weights.

Notes: DB = defined benefit; DC = defined contribution; IRA = individual retirement account; n/a = not applicable.

A gray cell denotes that we did not target this variable for the relevant subgroup. The reason may be that this variable could be calculated as a function of other targets' variables (e.g., net worth is defined as the sum of total financial and nonfinancial assets less total debts); that every tax unit viruallly had such item (e.g., number of tax units with some cash income); or that the source data lack such details (e.g., Schedules C and E losses).

For the tax model database, "single" include tax units whose income tax filing statuses were single, head of household, or married filing separately; "married" includes tax units whose income tax filing statuses were married filing jointly.

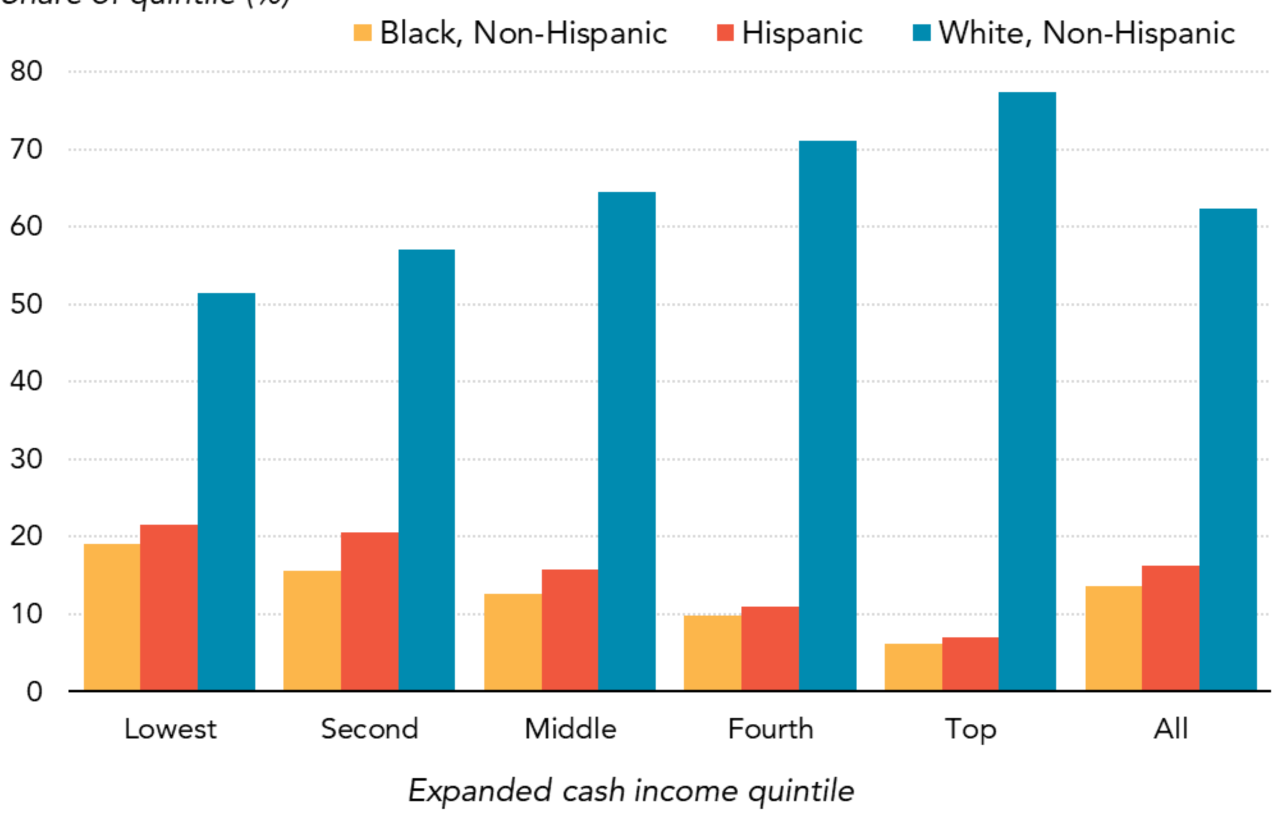
PRELIMINARY RESULTS

The targets we derived shed some light on average differences across race and ethnicity. For example, we estimate that overall, 14 percent of tax units are classified as Black, 16 percent Hispanic, and 62 percent White. Tax units classified as Black or Hispanic are more likely to be in the bottom of the income distribution and less likely to be in the top. For example, in the bottom income quintile, 19 percent are classified as Black, 22 percent Hispanic, and 51 percent White (figure 1). In the top quintile of the income distribution, in contrast, 6 percent of tax units are classified as Black, 7 percent Hispanic, and 77 percent White. These distributions mean that progressive tax policies—those that benefit low-income households more than high-income households—will also tend to benefit tax units classified as Black or Hispanic more than those classified as White. The reverse is also true; regressive tax policies—those that benefit high-income households disproportionately—will tend to benefit tax units classified as White more than those classified as Black or Hispanic. This is a well-known feature of the tax code, and our methodology can help to quantify the size of the relative advantages.

FIGURE 1

Racial or Ethnic Classification of Tax Units by Income Quintile

Share of quintile (%)



Source: Urban-Brookings Tax Policy Center Microsimulation Model (version 0721-2).
Notes: Totals add to less than 100% as not every tax unit is in one of the three categories.

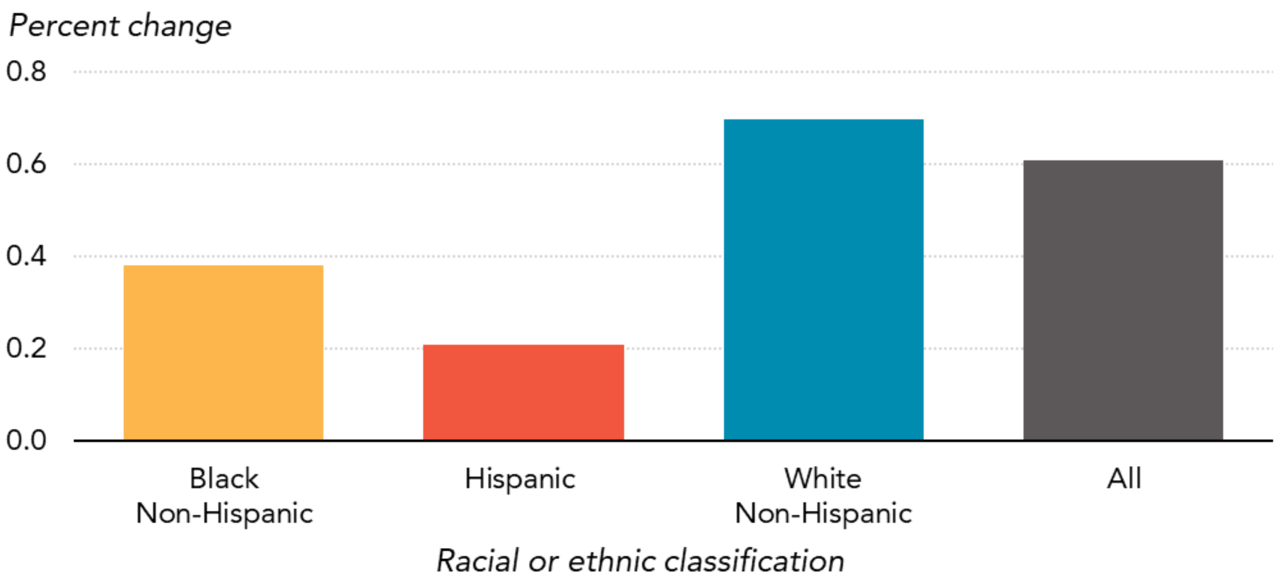
Although summary data from surveys make clear the general correlation between race and income, and thus between race and the effects of tax policies that depend on income, our method allows estimation of the dollar value of particular tax policies to tax units of different races and ethnicities.

Beyond the effects on the income distribution, demographics and other factors may affect the incidence, by race, of specific tax policies. For example, even within income quintiles, tax units classified as Black or White are more likely to report making charitable contributions than are those classified as Hispanic. Moreover, among those who contribute, tax units classified as Black or Hispanic report lower amounts of institutional giving than those classified as White on average.[37] Those factors suggest that the deduction for charitable contributions is likely to disproportionately benefit tax units classified as White, even within a given income class.

Estimates using the TPC tax model data, weighted by race and ethnicity, can reveal the combined effects of differences in income distribution and other factors and allow us to quantify the impacts of particular tax policies by race and ethnicity. To illustrate those capabilities, we produced preliminary estimates of the impact on after-tax incomes of itemized deductions. We estimated those policies at 2019 income levels assuming 2019 tax law.[38]

FIGURE 2

Change in After-Tax Income Attributable to Itemized Deductions



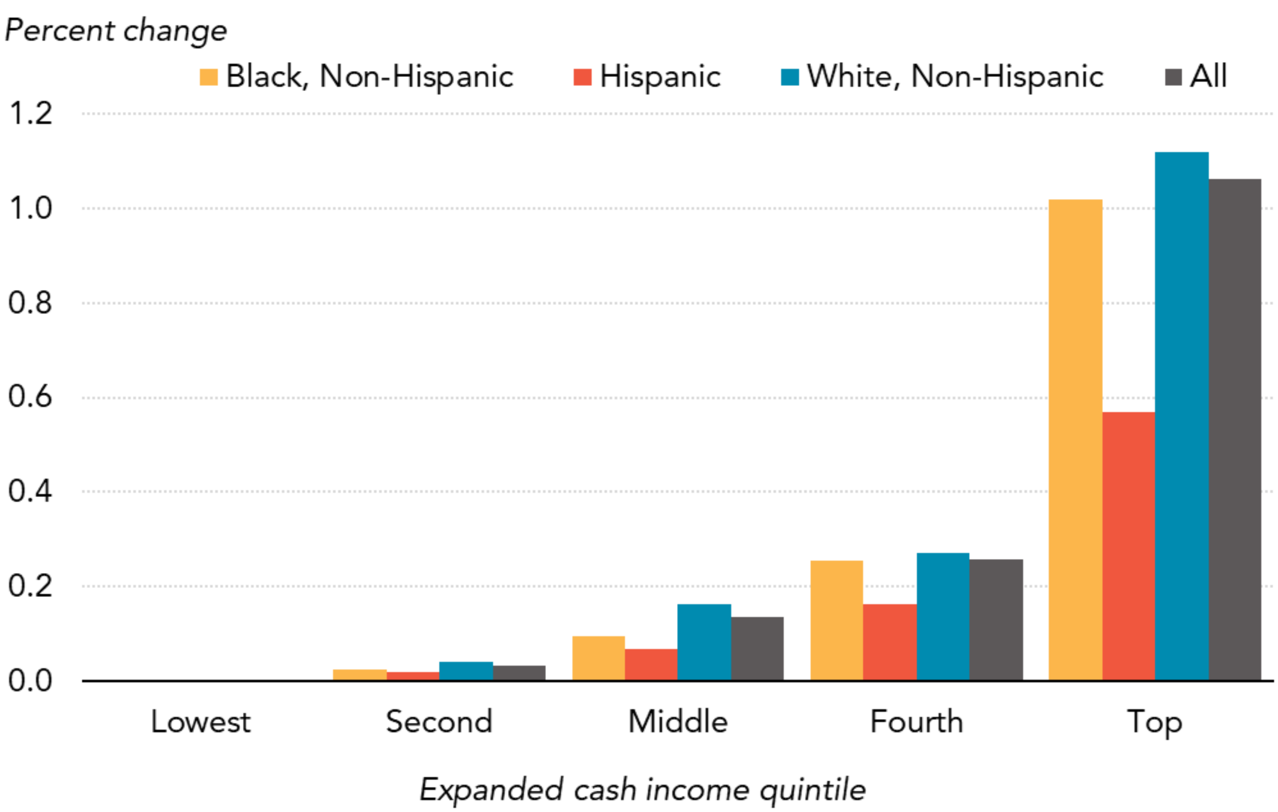
Source: Urban-Brookings Tax Policy Center Microsimulation Model (version 0721-2).
Note: In addition to Black, Hispanic, and White, the category All includes all other racial and ethnic categories.

To reduce their taxable income, taxpayers can choose to take a standard deduction or itemize deductions for qualifying expenses.[39] The itemized deductions that account for the largest total reductions in revenues are those for charitable contributions, mortgage interest payments on owner-occupied residences, and state and local taxes.[40] The total amount of itemized deductions must be relatively large to exceed the standard deduction. As a result, only 11 percent of tax units itemized deductions in 2019. Itemizers generally have relatively high incomes, and the value of itemized deductions increase as a taxpayer's income tax rate rises. Therefore, itemized deductions tend to provide the largest benefit to the highest-income households (Sammartino and Toder 2019).

We estimate that itemized deductions disproportionately raise the after-tax incomes of tax units classified as White (figure 2). Overall, itemized deductions boost the after-tax incomes of units classified as White by 0.7 percent, whereas those classified as Black gain an average of 0.4 percent, and those classified as Hispanic an average of 0.2 percent.

FIGURE 3

Change in After-Tax Income Attributable to Itemized Deductions



Source: Urban-Brookings Tax Policy Center Microsimulation Model (version 0721-2).
Note: In addition to Black, Hispanic, and White, the category All includes all other racial and ethnic categories.

White and Black households benefit relatively more from itemized deductions than Hispanic households, even within the same income categories. Among those in the top income quintile (which receives almost 90 percent of the benefits of the policy), itemized deductions raise the after-tax incomes of units classified as White by an average of 1.1 percent, compared to 1.0 percent for those classified as Black, and 0.6 percent for those classified as Hispanic (figure 3).

Most of this difference in the benefit from itemized deductions can be accounted for by lower deductible charitable contributions, on average, among tax units classified as Hispanic, as described above. Lower rates of homeownership (and therefore less benefit from the deduction for mortgage interest payments) also limit the benefit Hispanic households receive from itemized deductions.[41]

Next steps

Our strategy for modeling the effects of tax policy on different races and ethnicities has produced promising preliminary results. However, further refinements are necessary to produce estimates of the distributional impact of tax policy by race and ethnicity as a regular feature of TPC policy analysis.

Using data from additional surveys could enable us to enhance our estimates and extend our analysis of racial and ethnicity disparities to other provisions of the income tax as well as other types of taxes. For example, the Survey of Income and Program Participation provides detailed information on government benefit programs. The Consumer Expenditure Survey contains information on consumer purchases that we could use to account for differences in buying patterns across races and ethnicities in estimating the impact of excise taxes (such as cigarette or gasoline taxes) or a broader consumption tax (such as a value-added tax). The National Survey of Early Care and Education could provide additional information on the racial incidence of child care tax subsidies, and the National Postsecondary Student Aid Study could do the same for higher education subsidies. The American Housing Survey and Home Mortgage Disclosure Act data contain information that could improve estimates of the effect of housing subsidies, such as the home mortgage interest deduction or first-time homebuyer subsidies.

The CPS and SCF used in our current analysis provide some data in these areas, but more details could increase the reliability of our estimates. Our approach is well-suited to incorporating additional information, such as additional target variables or updated survey information. Targets for new variables could be added to the weight-splitting algorithm without altering existing targets. However, deriving those new or improved targets from process that would necessarily differ depending on the survey's structure.[42] We would then need to resolve any inconsistencies across surveys.

Our current analysis considers only three racial and ethnic categories—Black, Hispanic, and White—with all others relegated to a residual category, because that is the level of detail reported in the public-use SCF. (The 2022 SCF will address some of these limitations by increasing the sampling of families classified as Black, Hispanic, or Asian.[43]) Limiting the number of categories in this way allowed us to produce a comprehensive and consistent set of targets for each racial and ethnic category.

An alternative approach would be to expand the number of racial categories for the particular targets for which we have sufficient information. For example, the CPS includes the information on income, family structure, and other variables required to create some targets for a wider range of racial and ethnic categories. We could incorporate those additional targets (and possibly others from additional surveys) into our weight-splitting algorithm. That approach would allow us to confidently estimate for those additional racial categories the distributional impact of tax policies whose effects are determined only by variables included in that limited set of targets. However, estimation of a broader range of tax proposals for those additional racial categories—in particular, proposals dependent on variables such as wealth and different forms of capital income for which the most reliable source of information is the SCF—would need to be interpreted carefully given the lack of information at the same level of racial and ethnic detail.

A high priority for improving our estimates is to incorporate information about whether the members of a tax unit are eligible for claiming certain tax benefits, such as the earned income tax credit or the child tax credit. Taxpayers typically file a return using a Social Security number (SSN), but certain taxpayers ineligible for an SSN—for example, undocumented immigrants—can file using an individual tax identification number. However, taxpayers filing with an individual tax identification number are ineligible for the earned income tax credit. Under current law through 2025, those taxpayers can claim the $2,000 per child, partially refundable child tax credit, but only for for children with a valid SSN.[44]

The share of taxpayers filing without an SSN likely differs across racial and ethnic categories. For example, to the extent that they account for a disproportionate share of undocumented immigrants, Hispanics may be more likely to lack an SSN. Therefore, our current methodology might not accurately estimate receipt of the earned income tax credit or child tax credit for certain racial and ethnic categories unless we incorporate additional information on immigration status and possession of an SSN. Neither the CPS nor the SCF includes the necessary data to do so.

Fortunately, Urban Institute researchers have developed methods to estimate the immigration status of survey respondents.[45] We may be able to adapt this method to include possession of an SSN as a target in our weight-splitting strategy. Implementing that improvement would be challenging. We would have to impute possession of an SSN to tax units in the tax model database, matching the correlations with income and family makeup in the enhanced CPS data that include imputed immigration status.

Because that imputation would also require us to modify our modeling methodology for the child tax credit and earned income tax credit, we believe it would best be accomplished in conjunction with an annual update of the TPC tax model. The enhanced CPS data could also be used to create targets by race and income for holding an SSN, and those targets could be incorporated into our weight-splitting algorithm. These changes would allow us to improve our estimates of the distributional effects of some of the most important aspects of the code targeted toward low-income taxpayers.

This brief only examined estimates for 2019, the base year for our targets. Estimating policy impacts for future years requires an assumption about how racial and ethnic differences in the factors determining tax liability will evolve over time. Calculating new weights through the weight-splitting strategy for each future year would be both challenging (because it would require us to develop a method for projecting a set of targets by race and ethnicity for all future years) and time consuming (because weights must be split for each year separately).

As a simpler alternative, we plan to adapt our current methodology for projecting base-year tax model data to future years by incorporating Census Bureau population projections by race and ethnicity. If the adapted methodology is successful, the modified TPC tax model database would closely replicate both currently anchored economic targets (which are not classified by race and ethnicity) and demographic targets (which will be partly classified by race and ethnicity according to the Census projections).

Finally, a new TPC tax model module must be built to produce estimates of the distributional impact of tax policy by race and ethnicity as a regular feature of our tax model analysis. This change will require modifying the current set of TPC tax model fundamental programs to estimate and output distributional impacts by race and ethnicity and modifying the table-generating program to produce user-friendly tables that contain distributional estimates by race and ethnicity. The distributional estimates by race and ethnicity presented in this brief were produced using a customized program that crudely replicates this module capacity. Implementing this customized program involves many time-consuming steps that render it unsuitable for a regular production of distributional estimates by race and ethnicity.

We recognize the need to examine our results for accuracy in a number of additional ways. To the extent that other researchers and research organizations, such as the Treasury Department, release estimates of the racial impact of tax proposals, we will compare our results, and when possible review in depth the underlying behavioral and circumstantial factors that lead to differences in estimates of tax liability by racial category.

We also hope to obtain estimates by racial category for tax-related variables by using survey data that are directly linked to tax return data. The Census Bureau has access to a limited amount of tax return information that can, in many cases, be matched to CPS, and other survey, respondents. In other words, Census can link data from a tax unit's actual tax return to the responses the same household gave to the survey. Census could then tabulate that linked data to produce the same type of shares and targets used in our weighting process. Those tabulations could be useful in two ways. First, a comparison could validate the targets we derived from survey data. Second, they could provide alternative, or possibly additional, targets that we could use in our weighting strategy alongside our survey-generated targets.

With reliable estimates of tax liabilities by race and gender we will be able to investigate racial disparities in the effects of the tax code. Ultimately, we hope to expand the scope and reliability of our methodology to enable us to include analysis by race and ethnicity as a regular feature of our estimates of the distributional impact of tax policy. Incorporating race and ethnicity into our analyses would provide an additional lens through which to view the fairness and equity of tax policies. Public dissemination of this type of analysis would lead to a more informed debate and would promote the development of a more just system for all taxpayers.

ACKNOWLEDGMENTS

This brief is a product of the Urban Institute's Office of Race and Equity Research and Racial Equity Analytics Lab, which operates with the generous support of the Ballmer Group, the Bill and Melinda Gates Foundation, the Salesforce Foundation, and Urban's general support donors. Lead funding for this brief was provided by the Bill and Melinda Gates Foundation. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission.

REFERENCES

Dean, Steven A. 2021. Statement before the US House of Representatives, Committee on Ways and Means, Washington, DC, June 10. https://gop-waysandmeans.house.gov/wp-content/uploads/2021/06/Dean_Testimony.pdf.

Gale, William G., Swati Joshi, Christopher Pulliam, and John Sabelhaus. 2022. *Simulating Income Tax Liabilities in the Survey of Consumer Finances*. Washington, DC: Brookings Institution.

Rohaly, Jeffrey, Adam Carasso, and Mohammed Adeel Saleem. 2005. *The Urban-Brookings Tax Policy Center Microsimulation Model: Documentation and Methodology for Version 0304*. Washington, DC: Urban-Brookings Tax Policy Center.

Sammartino, Frank, and Eric Toder. 2019. "What Are the Largest Nonbusiness Tax Expenditures?" Washington, DC: Urban-Brookings Tax Policy Center.

Schirm, Allen L., Alan M. Zaslavsky. 1997. "Reweighting Households to Develop Microsimulation Estimates for States." *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.

Schirm, Allen L., Alan M. Zaslavsky. 1997. "An Evaluation of a Reweighting Method for Small Area Estimation." Washington, DC: Mathematica Policy Research, Inc., September 30, 2001.

APPENDIX

Table A1
Components Included in Income Measures

Income component	Dataset		
	TMDB	CPS	SCF
Compensation			
Wages and salaries	✓	✓	✓
Employee contributions to retirement plans	✓	✓	✓
Business and flow-through income (Schedules C, E, and F)			
Business/farm income or loss (Schedules C and F)	✓	✓	✓
Rents, royalties, and income from trusts	✓	✓	✓
Partnership income or loss	✓	✓	✓
S corporation income or loss	✓		✓
Investment income			
Taxable interest	✓	✓	✓
Tax-exempt interest	✓	✓	✓
Dividends	✓	✓	✓
Realized net capital gains	✓		✓
Retirement income			
Taxable individual retirement account distributions	✓	✓	✓
Taxable pension distributions	✓	✓	✓
Nontaxable pension distributions	✓	✓	✓
Other taxable income	✓		✓
Transfer payments			
Taxable Social Security benefits	✓	✓	✓
Nontaxable Social Security benefits	✓	✓	✓
Unemployment compensation	✓	✓	✓
Workers' compensation	✓	✓	✓
Supplemental Nutrition Assistance Program			✓
Supplemental Security Income	✓	✓	✓
Temporary Assistance for Needy Families	✓	✓	✓
Other cash transfer payments	✓	✓	✓

Sources: Author's calculations based on the Urban-Brookings Tax Policy Center Microsimulation Model (version 0721-2) database (TMBD), the 2020 March supplement to the Current Population Survey (CPS), and the 2019 Survey of Consumer Finances (SCF).

Note: The components included in the income measures used in calculating income categories (e.g., 0-25th percentile) for data are taken from the tax model database (TMDB), the Current Population Survey (CPS), and the Survey of Consumer Finances (SCF).

Table A2
Income Cutoffs by Marital Status (2019 dollars)

Income cutoff	Dataset		
	TMBD	CPS	SCF
Single			
25th percentile	15,000	18,000	15,300
50th percentile	29,700	32,000	29,500
75th percentile	52,200	55,300	50,900
95th percentile	115,700	121,500	113,300
Married			
25th percentile	46,900	50,400	50,000
50th percentile	88,400	91,000	91,600
75th percentile	148,000	151,700	152,700
95th percentile	343,800	316,300	430,700

Sources: Author's calculations based on the Urban-Brookings Tax Policy Center Microsimulation Model (version 0721-2) database (TMBD), the 2020 March supplement to the Current Population Survey (CPS), and the 2019 Survey of Consumer Finances (SCF).

Table A3
Share of Total Income by Marital Status and Income Group

Income Group	Dataset		
	TMDB (%)	CPS (%)	2019 SCF (%)
Single with income			
0 to 25th percentile	4.75	5.62	4.15
Above 25th but at most 50th percent	11.92	13.76	12.16
Above 50th but at most 75th percent	21.57	23.47	21.86
Above 75th but at most 95th percent	32.16	34.43	32.54
Above 95th percentile	29.60	22.72	29.30
All	100.00	100.00	100.00
Married with income			
0 to 25th percentile	4.61	6.14	5.23
Above 25th but at most 50th percent	11.83	14.21	11.37
Above 50th but at most 75th percent	20.19	23.94	19.16
Above 75th but at most 95th percent	29.31	33.79	30.12
Above 95th percentile	34.06	21.92	34.12
All	100.00	100.00	100.00

Sources: Author's calculations based on the Urban-Brookings Tax Policy Center Microsimulation Model (version 0721-2) database (TMBD), the 2020 March supplement to the Current Population Survey (CPS), and the 2019 Survey of Consumer Finances (SCF).

NOTES

[1] Aravind Boddupalli and Kim Rueben, "Racial Disparities and the Income Tax System," TPC Features, January 30, 2020, https://apps.urban.org/features/race-and-taxes/.

[2] "Executive Order on Advancing Racial Equity and Support for Underserved Communities through the Federal Government," The White House, January 20, 2021, https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/20/executive-order-advancing-racial-equity-and-support-for-underserved-communities-through-the-federal-government/.

[3] Julie-Anne Cronin, Portia DeFilippes, and Robin Fisher, "Tax Expenditures by Race and Hispanic Ethnicity: An Application of the U.S. Treasury Department's Race and Hispanic Ethnicity Imputation," Office of Tax Analysis Working Paper 122, January 2023, https://home.treasury.gov/system/files/131/WP-122.pdf; and Robin Fisher, "Estimation of Race and Ethnicity by Reweighting Tax Data," Office of Tax Analysis Technical Paper 11, January 2023, https://home.treasury.gov/system/files/131/TP-11.pdf.

[4] The TPC tax model is similar to those used by the Congressional Budget Office, the Joint Committee on Taxation, and the Treasury's Office of Tax Analysis. For more detail, see "Brief Description of the Tax Model," Tax Policy Center Resources, March 9, 2022, https://www.taxpolicycenter.org/resources/brief-description-tax-model.

[5] For more detail, see "Aging and Extrapolation Process" in "Brief Description of the Tax Model," Tax Policy Center Resources, March 9, 2022, https://www.taxpolicycenter.org/resources/brief-description-tax-model.

[6] For example, consider a household record with an original weight of 10. Given four racial categories, we can think of the process as initially splitting this household record into four records, each with a weight of 2.5. Given the targets from the survey data, the reweighting algorithm might settle on final weights of 6 for Black, 2 for White, 1 for Hispanic, and 1 for Other. Thus, the weights still sum to 10—and so the total across all racial categories would be unchanged—but the individual race and ethnicity weights reflect that the algorithm has decided that this tax unit is most consistent with the income, demographics, wealth, and tax deduction pattern exhibited by households classified as Black in the original survey data.

[7] We constructed targets separately for each of the three surveys and then constructed a combined weighted average target in which the weights are 50 percent for 2019 SCF, 30 percent for 2016 SCF, and 20 percent for 2013 SCF.

[8] In future work, we might consider the feasibility of creating targets by finer race and ethnicity classifications for certain variables for which a more detailed racial and ethnic breakdown is available.

[9] Black and Hispanic households are much more likely than white households to include relatives besides parents and children. See table H3, Households, by Race and Hispanic Origin of Household Reference Person and Detailed Type: 2022, at https://www.census.gov/data/tables/2022/demo/families/cps-2022.html.

[10] Following IRS rules, we assigned children as dependents if they were unmarried and under age 19 or students under age 24. We assigned other relatives, including adult children, as dependents if their incomes were below $4,300 and they passed the support test, meaning the respondent and spouse provided at least half of their resources.

[11] Tax unit respondents are generally the household respondent or subfamily members whom the CPS identifies as family respondents. The process described works because individuals within a household in the CPS are ordered by family with respondent first, followed by spouse, child, and other relatives.

[12] IRS rules required qualifying relative dependents have incomes below $4,200 in 2020.

[13] IRS rules require that head of household filers be unmarried, have dependents, and provide at least half the cost of "keeping up the home." The CPS does not have detailed information on expenses, so we approximated the cost test by comparing the tax unit's share of household income to a threshold. Using one-half as the threshold resulted in substantially fewer head of household filers in the CPS than in IRS data. To get closer to actual IRS totals, we set the share of income threshold at one-quarter for tax units with incomes above the tax filing threshold and dispensed with the cost test entirely for units with incomes below the tax filing threshold.

[14] For tabulation of targets for taxable income items such as wages or interest income, we used the combined income of the tax unit respondent and spouse. Dependents' income is generally not reported on the nondependent's tax return.

[15] See "Codebook for 2019 Survey of Consumer Finances" at https://www.federalreserve.gov/econres/files/codebk2019.txt.

[16] An example of a dependent filer is a teenage child who is claimed as a dependent on her parents' tax return but reports enough income that she is required to file a separate tax return.

[17] Moore's program computes tax rates and liabilities for the 1989–2019 SCFs; see http://users.nber.org/~taxsim/to-taxsim/scf27-32/code/frbscftax.sas.

[18] TAXSIM is a tax calculator housed at the National Bureau of Economic Research and maintained by Daniel Feenberg. See "TAXSIM Related Files at the NBER" at http://users.nber.org/~taxsim/.

[19] If the couple had an even number of dependents, they were split evenly between the partners; if there was an odd number, the extra dependent was assigned to the partner with higher income.

[20] The SCF asks about asset holdings at the time of the interview but about income from the previous full year. We inflated those income values to be consistent with the year of the survey.

[21] The SCF defines the male in an opposite-sex couple and the older person in a same-sex couple as the reference person, even if that individual is not the one responding to the survey.

[22] A possible refinement for future versions would use information on cohabiting partners of different races and ethnicities from another dataset, such as the CPS, to impute a potentially different race or ethnicity for each individual in an unmarried couple.

[23] We took this approach because we believe the SCF accurately measures the population captured by the PEU members of the household. We therefore chose to adjust the weights on the NPEU members to match our targets.

[24] For example, for a Forbes 400 individual we identified as White Non-Hispanic based on publicly available information, we set the White Non-Hispanic weight to the observation's original weight and the remaining three race and ethnicity weights to zero. We then excluded the Forbes 400 observations from our targets and the reweighting algorithm. We used the 2016 Forbes 400 because that is the year for which we imputed these individuals in our microsimulation tax model as part of a previous project to impute asset and debt holdings.

[25] In addition, because of its small sample size, an SCF observation with a large weight may overly influence the totals by race and ethnicity, creating potential data anomalies. For example, particular observations with surprising combinations of income and charitable contributions skewed the targets to the extent that they could not be matched by the weight-splitting algorithm. In those cases, the problematic targets were excluded.

[26] The income measures used to stratify observations in data from the CPS, the SCF, and the tax model database differ somewhat due to data constraints; see table A1 for details on the particular income components included in each measure. Despite the different definitions of income, however, the distribution of tax units by income appears to be quite similar across the three datasets, both in terms of the income cutoff by marital status for the percentile categories (table A2) and the share of total

income by percentile category (table A3). As a more specific example, one key difference in the income measures across the datasets is that, unlike the measures used for the tax model database and the SCF, the CPS measure of income does not include capital gains realizations. Using data from the tax model database, we verified that tax units rarely move across income groups when capital gains are excluded from the tax model database income measure. For example, 94 percent of the tax units in the top 5 percent of income including capital gains are also in the top 5 percent of income excluding capital gains; for the bottom quartile, the figure is 99 percent. For the tax model database data the income measure used in this analysis is a narrower definition than that used in TPC's usual distributional analysis. For a detailed discussion of that income measure, see "Income Measure Used in Distributional Analyses by the Tax Policy Center" at https://www.taxpolicycenter.org/resources/income-measure-used-distributional-analyses-tax-policy-center.

[27] The number of race and ethnicity targets for each subgroup varies because some targets are only relevant to some subgroups. For example, married tax units cannot have head of household tax filing status by construction, and there were few tax units in the topmost income group whose primary taxpayers were younger than 26 or received any transfer income.

[28] Using share of tax units with and without a pension as an example, in the first step of the first iteration, we multiply the SCF share of tax units with a pension plan for a race or ethnicity group *s* by the number of tax units with a pension plan in the tax model to derive a preliminary number of tax units of group *s* with a pension plan. Similarly, we multiply the SCF share of *s* without a pension plan by the number of tax units without a pension plan in the tax model to derive a preliminary number of tax units in group *s* without a pension plan. The summation of these preliminary numbers of tax units in group *s* with and without a pension plan will generally differ from this subgroup's number of tax units in group *s*. As a result, we adjust these preliminary numbers by using a common scaling factor such that the adjusted numbers sum to this subgroup's number of tax units in category *s*. We do this for every race and ethnicity category *s*.

After this adjustment, the total number of tax units with and without a pension plan across all races and ethnicities will generally differ from the number of tax units with and without a pension plan in the tax model. As a result, in the second step of the first iteration, we apply a common scaling factor across all races and ethnicities to adjust the number of tax units in each racial and ethnic category with a pension plan such that they sum to the number of tax units with a pension plan in the tax model. We similarly adjust those without a pension plan.

We repeat this iterative process until the derived number of tax units with and without a pension plan sums to the number of tax units by race and ethnicity, and simultaneously, the derived number of tax units with and without a pension plan across all races and ethnicities sums to the number of tax units with and without a pension in the tax model. For most targets, it required only a few iterations for the raking algorithm to achieve this outcome.

[29] Without a parametric specification, there would be an infinite number of solutions for race and ethnicity weights that satisfy the constraints specified by equations 2 and 3. Our specification reduces the number of unknowns to only ($\underline{\beta}$ and δ). With more observations than the number of parameters ($\underline{\beta}$ and δ) plus the number of constraints, a solution of ($\underline{\beta}$ and δ) is unique.

The following trivial example shows how the identifying assumption brings about a unique set of race and ethnicity weights. Suppose that we had two tax units, *A* and *B*, with identical characteristics. Their original weights are W^A and W^B, respectively. Suppose also that there were only two race and ethnicity categories 1 and 2, and the targeted race and ethnicity weights were W_1 and W_2, respectively. By construction, $W_1 + W_2 = W^A + W^B$. Without the identifying assumption, there would be an infinite number of solutions for $(w_1^A, w_1^B, w_2^A, w_2^B)$. In contrast, with our identifying assumption, the race and ethnicity category 1 and 2's shares of the original weights must be identical for both observations given their identical characteristics. As a result, we obtain the unique race and ethnicity weights $w_s^h = \left(\frac{W_s}{W_1+W_2}\right) W^h$ for s = 1 and 2 and h = A and B.

30 As a starting point, $\underline{\beta}_{s(0)}$ is set to zero, resulting in the original weights being equally split across all S races and ethnicities.

31 Because equation 3 cannot be satisfied exactly in general, in our calculations the condition is considered to be met if it holds within a specified tolerance level. That is, based on the derived coefficient estimates for any marital status and income subgroup, every race and ethnicity total of any targeted variable ($\sum_h w_s^h x_k^h$) must be within the specified tolerance level of the targeted race and ethnicity total (X_{ks}).

32 For the tax model, we defined married tax units as couples filing a joint return. For married filing separately tax units, we observed information for only one spouse. As a result, we grouped married filing separately tax units with single and head of household tax units.

33 See table A1 for details on the particular income components included in each measure.

34 In particular, the composition of income (specifically wages, retirement income, capital gains, and business income) by marital status and income group can vary wildly. Furthermore, some variables (such as transfer income for high-income taxpayers) are much less relevant for some subgroups and are likely to be captured with imprecision in public surveys. Targeting by marital status and income subgroup allowed us to capture variations across subgroups and focus on variables most relevant to each subgroup.

35 The deviations tend to be small due to the employed estimation technique. Recall that the estimation repeatedly searched for the best way to satisfy all constraints via a change in the coefficient estimates. In essence, it evaluated the differences between weighted and targeted totals and changed coefficient estimates to reduce such differences accordingly. For this evaluation to work, most differences must be sufficiently close to zero that a slight change in the coefficient estimates, which would cause all differences to change somewhat, should still result in these differences being relatively close to zero. Loosely speaking, it is as if this were a process of eliminating the largest difference observed in the previous iteration, and the process repeated until the last large difference was eliminated. That is, the iterations stopped when every weighted race and ethnicity total was within the specified tolerance of its targeted total, and so the tolerance should be binding for only one target.

36 We targeted 57 variables for all subgroups, corresponding to 228 targets by race and ethnicity (4 race and ethnicity targets for each of the 57 variables). The weighted race and ethnicity totals across all 10 subgroups deviate from their respective targeted totals by more than 2 percent for only 2, or 0.9 percent, of these 228 targets.

37 As estimated by the weighted tax model database, 85 percent of tax units in the top quintile reported making some charitable contributions, with $9,870 in contributions on average among tax units that contributed. The percentage contributing and average contributions, respectively, in the top quintile were 86 percent and $10,850 for non-Hispanic White tax units, 88 percent and $7,450 for non-Hispanic Black tax units, and 72 percent and $4,000 for Hispanic tax units.

38 We chose the year 2019 to match the year of the CPS and SCF survey data used in TPC's weighting strategy. Estimates based on years influenced by the COVID-19 pandemic could differ due both to the temporary tax measures enacted in response to the pandemic and to changes in household behavior during that time.

39 In 2019, the standard deduction was $24,400 for married couples filing a joint return, $18,350 for heads of household, and $12,200 for single filers and married individuals filing separate returns. These amounts are indexed annually for inflation. See "Standard Deduction: 1970 to 2021" at https://www.taxpolicycenter.org/statistics/standard-deduction.

[40] For the latest tax expenditure estimates, see https://www.jct.gov/publications/2020/jcx-23-20/. For the 2020 fiscal year, for example, the deduction for state and local government taxes provided a benefit of $21.1 billion. The comparable figure for the home mortgage interest deduction was $25.5 billion. The deduction for charitable contributions provided a benefit close to $50 billion.

[41] Tax units classified as Black also have lower rates of homeownership than those classified as White. However, they also are more likely to have mortgage interest payments, largely offsetting the impact of the lower rates of ownership.

[42] This process likely needs to be done on a case-by-case basis for each new survey, rather than by a fixed procedure, because each survey will have unique characteristics that will necessitate judgment calls.

[43] The SCF plans to more than double the sampling of Asian families in its area probability sample from about 200 to 450. This is still a relatively small sample, however, when looking at individual variables across income and marital status categories. See Kevin B. Moore and Karen M. Pence, "Improving the Measurement of Racial and Ethnic Disparities in the Survey of Consumer Finances," FEDS Notes, June 21, 2021, https://www.federalreserve.gov/econres/notes/feds-notes/improving-the-measurement-of-racial-and-ethnic-disparities-in-the-survey-of-consumer-finances-20210621.html.

[44] Children without a valid SSN are instead eligible for the $500 nonrefundable credit for other dependents.

[45] That approach follows methodology originally developed by Jeffrey Passel, currently at the Pew Research Center. For further details, see Jeffrey S. Passel and D'Vera Cohn, "Methodology," Pew Research Center, November 27, 2018, https://www.pewresearch.org/hispanic/2018/11/27/unauthorized-immigration-estimate-methodology/.