TAX POLICY CENTER
URBAN INSTITUTE & BROOKINGS INSTITUTION

TPC'S 2021 TAX MODEL TECHNICAL UPDATES

Benjamin Page

December 2021

The Urban-Brookings Tax Policy Center (TPC) large-scale microsimulation model produces estimates of how current and proposed tax policies will affect federal revenues and the distribution of tax burdens by income. The model is similar to those used by the Congressional Budget Office (CBO), the Joint Committee on Taxation (JCT), and the Treasury's Office of Tax Analysis (OTA). This year, TPC updated and revised its tax model in a variety of ways. In addition to incorporating new economic and budgetary projections published by the Congressional Budget Office (CBO) in January 2021, we implemented a new version of an algorithm to enable a streamlined update based on substantially revised CBO projections published in July 2021. We also incorporated major legislation enacted in response to the COVID-19 pandemic, including the American Rescue Plan (ARP) Act of 2021 and the Coronavirus Aid, Relief, and Economic Security (CARES) Act of 2020. Those updates enabled our distributional analyses of tax proposals in President Biden's fiscal year 2022 budget and various versions of the Build Back Better package, as well as a number of narrower proposals.

TPC's microsimulation model is the preeminent tax model outside of government. Its primary data source is the 2006 public-use file (PUF) produced by the Statistics of Income (SOI) Division of the Internal Revenue Service (IRS). We supplement these data with information on other demographic characteristics and sources of income that are not reported on tax returns from the US Census Bureau's March 2012 Current Population Survey (CPS). We perform various imputation and matching procedures to obtain a file that is nationally and state representative of the tax filing population for the 2011 tax year. We also use the CPS to obtain a sample of individuals who do not file individual income tax returns ("non-filers") but who would be affected by proposals to expand the tax rolls or to use the federal tax system to distribute social benefits.

THE ANNUAL ADJUSTMENT ALGORITHM

As noted above, the starting database for the tax model is a representative national sample of the population for calendar year 2011. In order to carry out revenue and distribution analysis for future years, we "extrapolate" or "age" the 2011 data. For that purpose, we use the actual 2012 through 2018 tax data available at the time we developed the database (mid-2021) as well as projections from a number of sources.

For the years from 2012 to 2031, we age the data based on:

- CBO forecasts and projections for growth of various types of income;
- CBO and JCT baseline revenue projections;
- IRS estimates of future growth in the number of tax returns;
- CBO and JCT estimates of the distribution of tax units by income;
- Census data on the size and age-composition of the population;
- Department of Education projections for growth in tuition and the number of post-secondary students;
- and CBO projections for growth in health care costs and personal consumption expenditures.

A two-step process produces a representative sample of the filing and non-filing populations in years beyond 2011. We first inflate the dollar amounts of income, adjustments, deductions, and credits on each record by their appropriate forecasted per capita growth rates. We use CBO's forecast for per capita growth of each major income source, such as wages, capital gains, and other forms of non-wage income (interest and dividends, business income, taxable pensions, Social Security benefits, and others).[1] We assume that most other items grow at CBO's projected growth rate for per capita personal income.

In the second stage of the extrapolation, we use a linear programming algorithm to adjust the weights on each record so that the major income items, adjustments, and deductions match aggregate targets. We also attempt to adjust the overall distribution of adjusted gross income (AGI) to match published information from the Statistics of Income (SOI) division of the Internal Revenue Service (IRS) for 2012 through 2018, and projections from CBO for years from 2019 through 2031.

We use a similar two-stage technique in the long-run module to age the data for each ten-year increment between 2040 and 2090. For 2040 and beyond, we rely primarily on projections from CBO and from DYNASIM3.

In the first stage of the long-run aging process, we use CBO's long-run inflation assumptions together with DYNASIM3 projections for the real growth in major income items such as wages and salaries, business income, capital income, pension income, and Social Security benefits, to grow the dollar amounts on the records in the tax model database.

In the second stage of the long-run extrapolation, we use our linear programming algorithm to adjust the weights on each record so that the major income amounts and certain other items match aggregate targets derived from the

1 In February 2021, CBO published detailed projections of income sources, personal exemptions, the standard deduction, itemized deductions, adjusted gross income (AGI), taxable income, and total income tax for years through the end of the budget window. See table 3 in https://www.cbo.gov/system/files/2021-02/51138-2021-02-11-revenueprojections.xlsx. We attempt to match these aggregate projections as closely as possible. In this same document, CBO provides its forecast for the distribution of AGI, which we use to inform our targets in the second stage of our extrapolation process. We incorporated the more recent CBO projections released in July 2021 when we updated the model using the interim adjustment algorithm.

DYNASIM3 and CBO forecasts. For example, we determine long-run targets for health insurance coverage and the number of post-secondary students by applying the demographic trends from DYNASIM3 to the health insurance status and student counts generated by the tax model for 2031. Similarly, we derive long-run targets for retirement coverage and contributions from a combination of CBO projections and the baseline imputations in the tax model.

We also use the second-stage reweighting algorithm to match DYNASIM3 targets for the age distribution of the population and other demographic characteristics, including the number of married and single tax units.

Finally, we use the reweighting process to target the distribution of tax units by income as projected by DYNASIM3, adjusted to match CBO's projected individual income tax revenue through 2040. For years after 2040, we rely exclusively on DYNASIM3's projection of changes in the income distribution.

The annual update also incorporates new imputations for wealth, education, consumption, health, retirement, and other factors, based on the latest available data.

WEALTH IMPUTATIONS

Because the income tax data in our model contain no direct information about wealth holdings, we rely on information from the 2016 Survey of Consumer Finances (SCF) to develop imputations for 18 categories of assets and debt.[2]

For each of these wealth components, we run probit and ordinary least squares (OLS) regressions against explanatory variables that exist in both the SCF and the tax model database. We use predicted probabilities from the probit regressions to estimate whether if a record will report a positive value for that item. For records with a positive projected value, we use the OLS regression coefficients to assign an amount.

We then calibrate the imputed number of tax units with a positive amount of each item, as well as the aggregate amount and distribution by income of each item, to match targets we create. For most variables, we base the targets on SCF data, but in certain cases, such as farm assets and debt, we rely on alternative data sources such as the Department of Agriculture.

Finally, we employ an adjustment algorithm and iterate to ensure that the overall distribution of net worth by the size of net worth holdings closely matches the distribution in the SCF, while ensuring that the distribution of each wealth component by income also remains consistent with the SCF data.

Because the SCF, by design, excludes the 400 wealthiest individuals – the "Forbes 400" – we impute their wealth using a different technique. Using published information for the 2016 Forbes 400, we categorize each member by age, marital status, and primary source of wealth. We then match these individuals to tax model records with the same marital status and age and with reported income consistent with the wealth profile reported by Forbes.

This process generates 400 tax model records flagged as tax units with Forbes individuals and with these individuals' associated net worth as reported by Forbes imputed to the tax units. We then assign asset and debt items consistent with that net worth by matching to similar records in the SCF with very high net worth. Finally, we apply the shares for each asset and debt item from the matched SCF record to the tax model record to impute detailed assets and debts.

[2] We also impute asset balances in defined benefit and defined contribution retirement accounts using a separate process discussed below.

EDUCATION IMPUTATIONS

To build the capacity to model tax incentives for education and their interaction with Pell Grants and other federal assistance to post-secondary students, we impute student characteristics to the tax model. First, we use data from the 2011-2012 National Postsecondary Student Aid Study (NPSAS)[3] combined with an indicator from the LAPUF as to whether a particular tax unit reported education tax incentives (such as the Lifetime Learning Credit or the above-the-line deduction for education expenses) to impute the presence of post-secondary students to each record in the database.

We then use the NPSAS to impute student characteristics –such as enrollment status, class year, and institution type – as well as education expenses – including tuition and fees, books, room and board, and transportation. We use these imputed characteristics to calculate potential education tax incentives and Pell Grants and assign take-up rates in order to match actual tabulations by income from SOI and the Department of Education.

CONSUMPTION IMPUTATIONS

To have the capacity to model the distributional impacts of federal excise taxes and other indirect taxes, including broad-based consumption taxes (e.g., a value-added tax) and environmental taxes, we impute consumption spending to each record in the tax model database. We use data from the Consumer Expenditure Survey (CEX) to produce estimates of consumption expenditures across 16 categories of goods and services for each tax unit in the tax model.

We also use the Urban Institute's Dynamic Simulation of Income Model (DYNASIM3) to estimate future consumption financed out of current wealth. DYNASIM3 is a dynamic microsimulation model that is designed specifically to project the population and analyze the long-run distributional effects of retirement and other aging issues.[4] This allows us to analyze transitional issues that arise in moving from an income- to a consumption-based tax system as proposed under some comprehensive reform options.

HEALTH IMPUTATIONS

To be able to analyze tax subsidies for health insurance and medical expenses – such as tax expenditures for employer-provided health benefits, tax credits for non-group insurance purchased through Affordable Care Act (ACA) marketplaces, and the ACA penalty on employers offering insufficient health insurance coverage –we impute health insurance status and employer-provided health benefits to each record in our database.

We impute health insurance status using a statistical match with the Urban Institute's Health Insurance Policy Simulation Model (HIPSM).[5] HIPSM is a detailed microsimulation model of the health care system designed to estimate the cost and coverage effects of proposed health care policy options.[6]

We impute employer-provided health benefits by statistically matching tax units with employer-sponsored health insurance to employers offering health coverage in the 2015, 2017, and 2018 Kaiser/HRET employer surveys.[7] We

---

[3] The NPSAS is produced by the National Center for Education Statistics.

4 For a detailed description of the projection methods employed by DYNASIM3, see Favreault, Smith, and Johnson (2015). "The Dynamic Simulation of Income Model (DYNASIM): An Overview."

5 For a description of the statistical matching process, see Mermin and Buettgens (2020). "Description of The Tax Policy Center Microsimulation Model's Revamped Health Module."

6 For a detailed description of HIPSM, see Buettgens and Banthin (2020). "The Health Insurance Policy Simulation Model for 2020."

7 The Kaiser/HRET annual survey of employer sponsored health benefits is sponsored by the Kaiser Family Foundation and Health Research & Educational Trust.

impute employer and employee contributions for health, dental and vision insurance; Health Savings Accounts; Health Reimbursement Arrangements; and Medical Flexible Spending Accounts.

RETIREMENT IMPUTATIONS

To be able to analyze the revenue and distributional implications of tax measures related to retirement savings, we impute a comprehensive set of pension and retirement savings variables to each tax unit in the tax model database. These variables include, when relevant, eligibility, contribution amounts, accrued benefits, and asset balances of each taxpayer's defined benefit pension, a defined contribution pension, and an Individual Retirement Arrangement (IRA). We rely on information from the SCF to impute pension characteristics as well as pension and IRA asset balances, and we use SOI data to impute IRA characteristics. We supplement and calibrate these imputations based on information from a number of official economic reports and projections.[8]

OTHER IMPUTATIONS

To complete the tax model database, we perform a number of other imputations. First, we use tabulations from the TRIM3 microsimulation model to adjust the reported values of certain non-taxable transfer payments.[9] The reported values obtained through our statistical match with the CPS generally undercount both the number of recipients and total dollar amounts for food stamps (SNAP), Temporary Assistance for Needy Families (TANF), and Supplement Security Income (SSI). We therefore adjust our counts and amounts to match the TRIM3 reported values more closely.

We also impute mortgage interest on second homes and deductible interest on home equity loans. In addition, the model contains imputations for all itemizable deductions—including charitable contributions, medical expenses, and home mortgage interest—for "non-itemizers," people who claim the standard deduction on their tax return. These imputations allow us to model the distribution and revenue implications of proposals to replace certain deductions with credits that would be available to all taxpayers regardless of itemization status.

To be able to estimate proposals that would tax unrealized gains on assets held at death, we use data from the 2019 SCF to impute unrealized capital gains for records in the tax model. We use a statistical match between the tax model database and the SCF to impute the ratio of unrealized gains to asset values for the four types of unrealized gains reported in the SCF: primary residence; other real estate; stocks and mutual funds; and businesses. After the match, we employ an adjustment algorithm to ensure that the amounts and distribution of unrealized gains in the tax model match those in the source data.

STATE WEIGHTS

To perform state-level analyses, we create additional weights to make the tax model database representative for each of the 50 states, the District of Columbia, and other areas.[10] We impute these 52 state weights to the tax model

8 These sources of information include the Bureau of Economic Analysis, Bureau of Labor Statistics, Census Bureau, Congressional Budget Office, Department of Defense, Department of Treasury's Office of Tax Analysis, the Federal Reserve Board, the Internal Revenue Service's Statistics of Income program, Joint Committee on Taxation, Office of Personnel Management, and Thrift Savings Fund.

9 TRIM3 is maintained and developed by the Urban Institute, under primary funding from the Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation (HHS/ASPE). Information presented here is derived in part from the Transfer Income Model, Version 3 (TRIM3) and associated databases. TRIM3 requires users to input assumptions and/or interpretations about economic behavior and the rules governing federal programs. Therefore, the conclusions presented here are attributable only to the authors of this report.

10 Those other areas include, for example, returns filed from Army Post Office and Fleet Post Office addresses by members of the armed forces stationed overseas; returns filed by other U.S. citizens abroad; and returns filed by residents of Puerto Rico with income from sources outside Puerto Rico or with income earned as U.S. government employees.

database using a method that guarantees each state's weighted totals of chosen observed characteristics will match state targets. For years in which published IRS tax return data are available by state, targets come from these publications. When IRS data are not available by state, targets are detailed projections based on available IRS data and macroeconomic assumptions.[11]

UPDATING TAX CALCULATORS TO INCORPORATE CHANGES IN TAX LAW

The tax model includes a set of detailed tax calculators that: (a) compute individual income tax liability for all tax units in the sample under current law and under alternative policy proposals; (b) compute the employee and employer shares of payroll taxes for Social Security and Medicare; (c) assign the burden of the corporate income tax and excise taxes to tax units; and (d) determine the expected value of estate tax liability for each tax unit in the sample using an estate tax calculator in combination with age-specific mortality rates. In order to enable accurate analysis, those tax calculators must be continually updated to incorporate any changes in tax law.

Based on the extrapolated data set, we can simulate policy options using a detailed tax calculator that captures most features of the federal individual income tax system, including the alternative minimum tax (AMT). The model's current law baseline reflects major income tax legislation enacted through October 2021, including the American Rescue Plan (ARP) Act of 2021, the Coronavirus Aid, Relief, and Economic Security (CARES) Act of 2020, and the 2017 Tax Act commonly referred to as the Tax Cuts and Jobs Act (TCJA).

AN INTERIM ADJUSTMENT ALGORITHM

Due to its complexity, we generally employ the annual adjustment algorithm just once a year, typically after CBO releases its Budget and Economic Outlook in late January or early February. However, when economic circumstances change dramatically (as they have during the ongoing pandemic), CBO may substantially revise its economic projections during the year.

To expeditiously and accurately incorporate these revisions, we have developed an interim adjustment algorithm to update the tax model baseline database and tax parameters of the relevant annually updated tax model. Specifically, using the original CBO projections incorporated in the annually updated tax model as a benchmark, we derive additional aging factors from the revised CBO economic projections and apply these factors to further inflate the dollar amounts and tax parameters in the annually updated tax model.

However, we implement only a simple linear programming algorithm to further adjust the weights on each record to match the revised numbers of people, workers and unemployed. This speedy, albeit less sophisticated, algorithm allows us to provide the most accurate possible estimates of a proposed tax policy's impacts in a timely manner.

ANALYSES USING UPDATED MODEL, AND NEXT STEPS

TPC's 2021 annual update enabled a comprehensive analysis of the distributional effects of the changes to tax law proposed in President Biden's fiscal year 2022 budget. That analysis was published in June 2021. However, in July the CBO published substantially revised economic and budgetary projections. TPC judged that accurate estimation of the distributional effects of tax proposals required incorporating the new projections using the interim adjustment algorithm described above.

11 For more information, see Khitatrakun, Mermin, and Francis (2016). "Incorporating State Analysis into The Tax Policy Center's Microsimulation Model: Documentation and Methodology."

The interim adjustment process began with a revision of TPC's interim adjustment algorithm to incorporate new projections more quickly and accurately. (The algorithm was first developed and used in the fall of 2020 to incorporate projections that were revised due to the economic and budgetary effects of the COVID-19 pandemic). The algorithm was then used to adjust the model to incorporate the new projections. As part of the updating process, TPC created over 100 tables presenting distributional aspects of current tax law estimated using the new baseline. Those tables are important both for use by researchers outside TPC (including for interpretation of TPC analysis), and for review of the accuracy of the baseline.

The interim update allowed analysis of the Ways and Means Committee reconciliation legislative recommendations (a precursor to the revenue provisions in the Build Back Better (BBB) proposal) in September 2021, and an analysis of the BBB as reported by the House Budget Committee in November. The updated model was also used to estimate various versions of particular provisions as BBB moved through the legislative process. Those included a number of different proposals for extending elements of the Child Tax Credit provisions of the ARP, as well as many different versions of proposals to raise or eliminate the cap on deductions for state and local taxes.

Going forward TPC plans to continue to pursue a more comprehensive update of the tax model, based on a more recent public use datafile from the IRS. That effort is more extensive that the annual or interim updates.

ACKNOWLEDGMENTS

This brief was funded by the Peter G. Peterson Foundation. We are grateful to the foundation and to all our funders, who make it possible for the Urban-Brookings Tax Policy Center to advance its mission.

The views expressed are those of the authors and should not be attributed to the Urban-Brookings Tax Policy Center, the Urban Institute, the Brookings Institution, their trustees, or their funders.