

## Methodology

The Congressional Budget Office uses a multistep methodology to produce its estimates of the distribution of income and taxes. The Current Population Survey (CPS) and the Statistics of Income (SOI) are the primary sources of data for CBO's estimates of population and household income. CBO estimates federal taxes for each household on the basis of income, demographic characteristics, and existing laws in the relevant year. CBO then groups the households into quintiles on the basis of their income and tabulates the income, taxes, and effective tax rates (the amount of tax liability divided by income) for each quintile.

### Sources of Data

CBO's analysis of effective tax rates draws information on income from two primary sources. The SOI, produced by the Internal Revenue Service, reports much of the information that taxpayers provide on their individual income tax returns. The March supplement to the Bureau of the Census's CPS contains survey data on both the demographic characteristics and income of a large sample of households.

CBO statistically matches each SOI record to a corresponding CPS record on the basis of demographic characteristics and income. The matching process begins by dividing all records in both the CPS and the SOI into demographic subgroups on the basis of household composition, including the marital status of the head of the household, the number of children, and the number of elderly members. Because income measures in the SOI and CPS are not directly comparable, CBO uses a regression equation for each demographic subgroup (which is estimated using data from the SOI) to calculate a predicted income for each record for each source of data. All CPS and SOI records are ranked by predicted income within demographic subgroups, and the two files are then matched, starting at the top of the income distribution: the SOI record with the highest predicted income is matched with the CPS record with the highest predicted income—after their sample weights are taken into account.<sup>1</sup> The SOI record with the next highest predicted income is then matched with the corresponding CPS record, and the process is repeated until all SOI records in the demographic subgroup have been paired with CPS records.

Each pairing results in a new record that takes on the demographic characteristics of the CPS record and the income reported in the SOI. Some types of income, such as most transfer payments and in-kind benefits, appear only in the CPS; values for those items are drawn directly from that survey. Because not all households have to file tax returns, some households do not appear in the SOI; thus, the CPS reflects more households. After all SOI records have been matched to CPS records, the remaining survey records are

---

<sup>1</sup> Because both the SOI and the CPS are stratified samples, each record is assigned a sample weight indicating its relative size in making up the entire population represented by the database. The matching process takes account of those weights by effectively considering each record to be replicated by the number of times indicated by its weight.

recorded as nonfilers, and the income values are taken directly from the CPS. CBO then estimates the tax liability for each matched record.

## **Who Pays Taxes?**

CBO's analysis of effective tax rates assumes that households bear the burden of the taxes that they pay directly, such as individual income taxes (including taxes on interest, dividends and capital gains) and employees' share of payroll taxes. The analysis assumes—as do most economists—that employers' share of payroll taxes is passed on to employees in the form of lower wages than would otherwise be paid. Therefore, the amount of those taxes is included in employees' income, and the taxes are counted as part of employees' tax burden. CBO estimates payroll taxes and individual income taxes, including refundable tax credits, with a tax "calculator" that applies the tax law for the relevant year to the tax return data from the SOI.

Excise taxes are assumed to fall on households according to their consumption of taxed goods (such as tobacco and alcohol). Excise taxes that affect intermediate goods, which are paid by businesses, are attributed to households in proportion to their overall consumption. CBO assumes that each household spends the same on taxed goods as similar household with comparable income in the Consumer Expenditure Survey.

Far less consensus exists about how to attribute corporate income taxes (and taxes on capital income generally). In this analysis, CBO assumes that corporate income taxes are borne by owners of capital in proportion to their income from interest, dividends, capital gains, and rents. Over the long term, however, some models suggest that at least part of the burden falls on labor income.<sup>2</sup>

## **Measuring Income**

This analysis focuses on households' adjusted pretax comprehensive income. That measure includes all cash income (both taxable and tax-exempt), taxes paid by businesses (which are imputed to individuals, as noted above), employees' contributions to 401(k) retirement plans, and the value of income received in-kind from various sources (such as employer-paid health insurance premiums, Medicare and Medicaid benefits, and food stamps). The calculations use the Census Bureau's fungible value measure to determine the cash equivalent of in-kind government transfer payments.

## **Income Quintiles**

In a series of steps, CBO combines the income and taxes of households to create tables showing distributions of income and taxes among income groups and types of households. First, CBO adjusts for household size by dividing household income by the square root of household size, to take account of the differing needs of larger and smaller

---

<sup>2</sup> See, for example, William C. Randolph, International Burdens of the Corporate Income Tax, CBO Working Paper 2006-09 (August 2006).

households.<sup>3</sup> In the next step, CBO ranks households by their (adjusted) income and groups them in quintiles (fifths of the distribution). The quintiles contain equal numbers of people, but because households vary in size, quintiles generally contain unequal numbers of households. CBO then tabulates overall income and taxes for each quintile as well as for smaller groupings at the top of the distribution.

### **Types of Households**

Separate tables in the appendix show effective tax rates and income for three different types of households: those with members under age 18 (households with children), those headed by a person age 65 or older and with no member under age 18 (elderly childless households), and all others (nonelderly childless households). Those tables group households into quintiles based on their position in the population-wide income distribution, not the sub-population distribution, so households within a sub-population need not be evenly spread across the income quintiles.

---

<sup>3</sup> CBO used adjusted income only to rank households in the income distribution; the values for average income are unadjusted for household size.