CENTER ON BUDGET AND POLICY PRIORITIES

EMBARGOED FOR RELEASE:
6:00 p.m. Eastern Time, September 4, 2000

AN ANALYSIS OF NEW IRS INCOME DATA
by Isaac Shapiro, Robert Greenstein, and Wendell Primus

APPENDIX:
A Brief Comparison of the Strengths and Weaknesses of IRS and Census Data

There are numerous differences between the income data reported to IRS on tax returns and the data reported to the Census Bureau through a survey of U.S. households. In combination, these differences make the IRS data much more accurate in describing income levels and trends for the top part of the income distribution, particularly the top one percent or five percent of the population. The largest comparative advantage of the Census data is for poor and near-poor households. The Census data on the income levels of these households are substantially superior to the IRS data.

As this paper explains, the principal divergence in what the IRS and Census data show concerning income trends over time is found in the trends for those at the top of the income spectrum. The Census data for that group are incomplete, and the IRS data much more comprehensive.

Definition of Income. The IRS data measures adjusted gross income (AGI). It does not include income that is not part of adjusted gross income. The principal form of income that is not part of AGI consists of cash benefits provided through government programs, including means-tested cash assistance (such as Temporary Assistance to Needy Families benefits, General Assistance and Supplemental Security Income benefits) and most Social Security benefits. AGI also does not include child support payments. These income sources are relatively more important to lower-income families than those at higher-income levels. On the other hand, capital gains income, an important source of income to higher-income families, is not reported to the Census but is included in the IRS data. In addition, for several reasons primarily related to privacy and to gaining cooperation from those who are surveyed, the Census Bureau places a limit on the amount of income recorded on the Census forms. For example, a CEO of a major corporation or an athlete who receives a salary of several million dollars would have only $999,999 of that income recorded on the Census form. This limit is not increased each year and is adjusted only intermittently; it results in an understatement of the income of the wealthiest households and also can somewhat distort income trends among these households over time. No such artificial limits exist in the IRS data. Recognizing the incompleteness of its data on very high-income households, the Census Bureau does not publish data on the incomes of the top one percent of the population.

Family or Household Unit. In the IRS data, there is one record for each tax filing unit, including dependents who earn fairly small amounts of income and file separately. In the Census data, all

related individuals who reside in the same household are considered as one record. One example where a household could constitute different units in the Census and IRS data would be a three-generation household in which the adults in the two oldest generations work. These two generations would constitute separate tax filing units in the IRS data but would be represented as one household in the Census data. Consequently, there are more tax filers than households. For higher-income families, this difference is relatively unimportant.

Population Coverage. The IRS data cover the bulk of the population — those who file federal income tax returns. Individuals and families with very low incomes are not required to file federal tax returns, however, although many do file in order to receive the Earned Income Tax Credit. Those who do not file are not represented in the IRS data. The Census data covers the entire non-institutionalized population and is weighted accordingly to represent this population.

Sampling methodology. The IRS data are based on a sample of about 125,000 filers, while the Census data are based on a sample of 50,000 households. There are substantial penalties for failing to complete an income tax form, particularly for higher-income individuals. There is no such penalty for not participating in the Census sample, and many household surveys have difficulty obtaining the cooperation of high-income individuals. Typically, surveys are more likely to be completed by other households. In addition, the probability of being included in the IRS sample increases with income; this is not the case in the Census sample. In other words, high-income families are overrepresented in the IRS sample, and the measurement of income thus is much better for higher-income families.

Survey Versus Administrative Data. The Census data are collected through a household survey that normally takes about 30 minutes to complete. Some types of income data collected on household surveys, such as dividends and interest, are subject to considerable response error (i.e., respondents do not always recall these data correctly). By contrast, the IRS data are based upon a tax form that is prepared over some number of hours, with the tax filer having been reminded of most major income sources through forms from employers or financial institutions. Substantial penalties exist for misreporting income to the IRS; there is no penalty for misreporting income data to the Census Bureau.

Implications

The primary implication of these differences between the Census and IRS data is that the average level of income for the highest-income families is understated significantly in the Census data, while the average level of income for low-income families is understated significantly in the IRS data (and some other low-income families are not included in the data). In recognition of these difficulties, the Census Bureau does not report data on the top one percent of households, and the IRS study examined here reports information on the bottom 50 percent of filers but does not break out income data for any subgroups in that category.

Because the methodologies for collecting these data are largely the same across time — that is, the IRS applies the same methodology each year, and the Census Bureau methodology is largely consistent over time as well — some of these gaps in data do not necessarily affect the measurement of income trends. For example, if income not included in the IRS data is growing at about the same rate as the income that is included, then the trend for the bottom portion of the population reflected in the IRS data may not differ much from the trend for the bottom of the population as reported in the Census data. Indeed, as noted in this analysis, the IRS data show that average income for the bottom 90 percent of filers grew 3.6 percent between 1995 and 1997, and the Census data show an identical rate of income growth over this period — 3.6 percent — among the bottom 80 percent of households. If the omitted income is growing at a different rate, however, the data on income trends can be distorted. For example, if capital gains income (omitted in the Census data) is growing considerably faster than wage income — as it has been — the income trends for the top five percent will be considerably less accurate in the Census data than in the IRS data.

The Congressional Budget Office combines the Census and IRS data and uses the strengths of both data sets to produce the best set of consistent income data across time. The only difficulty is that these CBO data are produced with a considerable time lag. The most recent set of actual data available from CBO is for 1995. The income trend for higher-income families shown in these CBO data — as well as the trend of widening income disparities — is similar to that shown in the IRS data for the same years. Thus, there is strong reason to believe that the basic conclusions reported here will soon be embodied in forthcoming CBO studies. Although the precise figures in the CBO data will not be an exact match with the IRS data, it is safe to predict that the CBO data for the 1995-1997 period will show the same trend of rapid income growth among the top one percent of the income spectrum that far outpaces income growth for the rest of the population.