HOUSEHOLD AND INDIVIDUAL INCOME DATA FROM TAX RETURNS

Peter Sailer and Michael Weber, Internal Revenue Service
Peter Sailer, SOI Division CP:R:S:I:S, PO Box 2608, Washington, DC 20013-2608

Key Words: Tax Return, Filing Unit, Household

Traditionally, the Internal Revenue Service has carried out its mandate to measure "the operations of the internal revenue law" by showing the "number of returns" with various characteristics and amounts of income, tax, and deductions. The limitation of the tax return as a unit of measurement is that it does not necessarily represent an economic unit. Some married couples may file separate returns; some parents may turn some of their wealth over to their children, who then report the income on their own returns. Some dependent children may have earnings of their own--again reported on their own returns. Some aged parents may be living in the same household with their children, or be supported by them in a nursing home, and their combined incomes may place them in a completely different income class from the one in which a return-by-return distribution would put them.

Aggregating and Disaggregating Return Data

For several years now, the Statistics of Income Division (SOI) at IRS has been working on a new data base which will permit the aggregation of data from tax returns into household units, in which all married couples and their dependents are shown as single entities (see Hostetter et al., 1990; Hostetter and O'Conor, 1991). The opportunity to do so first arose as a result of the Tax Reform Act of 1986, which required taxpayers to report social security numbers (SSNs) for their dependents. (For Tax Year 1987, this rule applied to dependents age 5 and over, although a grace period was allowed for taxpayers who wrote "applied for" in the "dependent SSN" box; for 1993, the age limit was one year, and for 1996, it was one month.)

It is these dependent exemptions, along with the requirement of name and SSN information for the spouses of those who choose to file as "married filing separately," which make it possible for us to assemble "tax families" (see Czajka and Shirm, 1993). As the authors explained in a previous paper (Sailer and Weber, 1996), these units bear some resemblance to various categories of households for which the Census Bureau publishes data. Indeed, the data base contains the information necessary to recreate some Census household types rather accurately--for example, since the data base contains age information for all taxpayers and dependents, as well as address information for the taxpayers, it is possible to create the category "married couples, living together, with dependent children under age 18." For the purpose of this paper, however, we will not make such fine distinctions. Any taxpayers joined to a "parental" return because their SSNs appeared there (either as that of a spouse or a dependent) will be considered part of the same household, without regard to age or place of residence.

In addition to the aggregation just described, it is also possible to disaggregate some of the data which are commingled by husbands and wives who file joint returns. The way this is accomplished is by matching the sample of tax returns to a file containing information documents--Forms W-2 prepared by employers and Forms 1099 prepared by banks, stock brokers, pension funds, etc. Such a match was performed for the 1993 Statistics of Income sample.

The result of these efforts is a massive data base, including all the data normally edited to produce the annual Statistics of Income report (Internal Revenue Service, 1996), as well as information from all the associated information documents, and links tying parental and dependent returns together. Analysis of this data base begins with this paper, which will, needless to say, barely scratch the surface. All that we are attempting here is to demonstrate some of the consequences of regrouping the traditional Statistics of Income file--either by aggregating the data by household, or by dis-aggregating them into data for individuals--and to hint at some of the analysis these regroupings will allow.

Income Distributions and Tax Rates for Aggregated Data

How do the data aggregated by household differ from those presented in the annual report *Statistics of Income--Individual Income Tax Returns*? The top line in Figure 1 shows an income distribution for income tax returns, classified by size of adjusted gross income. The lower line shows data for the same returns, aggregated by household. As expected, the shape of the income distribution changes considerably, especially at the lower end. Nearly half of what looked like a large class of the "truly needy" (people with incomes under $5,000) turned out to be dependents or spouses of other taxpayers. Similarly, the $5,000 under $10,000 class decreased by 13 percent. And while it does not show up too well on this graph, all size classes above $55,000 increased as a result of the addition of these lower-income returns to the higher-income tax households.

A few more observations should be made about this new distribution. The lowest decile ends at about $5,811, the top decile starts at $69,527; the median income in this distribution is $23,796. This may be of some help to tax analysts who are trying to describe the "middle class," a group whose need for tax relief has been much discussed in the political arena in recent months.

Now that our data base has been re-grouped into household units, it is possible to compare data for one-return households with data for households which spread their incomes over more than one return. Of course, in many cases this is not a matter of choice. If a dependent child has a summer job, that child will have to file his or her own return. On the other hand, some taxpayers may be choosing to put a portion of their wealth in the names of their children in order to get an extra standard deduction to offset some of the income. Similarly, a taxpayer may be filing a separate return simply because his wife refuses to sign a joint return; on the other hand, he may be doing so to gain additional amounts in medical or miscellaneous deductions, both of which might be eliminated by the subtraction of the applicable percentage of the combined adjusted gross income of the married couple.

In Figure 2, the darker line represents the effective tax rates paid by all taxpayers filing joint returns, classified by size of adjusted gross income. Effective tax rate is described, for purposes of this paper, as total income tax (the regular tax, after credits, plus the alternative minimum tax) as a percent of adjusted gross income. The lighter line represents the same information for households filing multiple returns. As can be seen from this chart, at all but the very lowest income classes, multi-return households, as a group, pay taxes at a slightly lower rate than prevails for all joint returns. This is true in spite of the many safeguards built into the law that require children's unearned incomes in excess of $1,200 to be taxed at the same rates as their parents' income. It will be interesting, in future analyses, to determine the exact characteristics of households which gain from filing multiple returns and those which do not.

Income Distributions and Tax Rates for Disaggregated Data

As mentioned earlier, our new data base allows us not only to aggregate data by households, but, to some extent, also to disaggregate data into units representing individuals, even if those individuals file joint returns. The caveat "to some extent" is needed because some types of income tend to be commingled beyond anybody's ability to disaggregate. For example, in IRS files, all interest statements issued by banks (Forms 1099-INT) are identified by the SSN of whichever owner is listed first on the bank's records for the account. But the IRS does not know whether the account belongs to one or two (or three or more) individuals. We allow the taxpayers to sort that out for themselves, as long as the amount of interest is accounted for. But, of course, they sort it out on tax returns--which include joint returns-- which means that we have lost our ability to disaggregate this income.

Earned income, such as salaries and wages, is, of course, another matter. Here, the information documents clearly identify the owner of the income. For the purpose of this paper, we have simply used Forms W-2 to identify returns which represent two wage-earners. Both total adjusted gross income and total taxes on these two-wage-earner returns were allocated to the two taxpayers in the same proportion as their salaries. By doing so, we are dividing up the couple's unearned income based on their earning power in the current year, which, at least on average, should provide defensible results. All joint returns that did not have two wage-earners are left as belonging to a single individual. Not unexpectedly, the income distribution gained from this rearrangement nearly triples the occupants of the "Under $5,000" class, when compared to the arrangement by family (see Figure 3).

This second rearrangement of our data, even though not completely satisfactory, will be especially useful when the file has been sex-coded, since it will allow analysis of income and tax by gender of taxpayer. Another possible use is to study the taxation of two-earner couples. One of the arguments made against the tax law since enactment of the Tax Reform Act of 1986 is that it discriminates against two-earner married couples by abolishing the two-earner deduction and, thereby, re-introduces the "marriage tax." What is meant by "marriage tax" is that the two-earner couple is paying more in tax than it would if the partners had remained single.

Figure 4 shows the effective tax rates paid, in the aggregate, by single people, which rises from nearly 0 percent in the "Under $5,000" income class to around 20 percent at the "$95,000 under $100,000" income level. Not surprisingly, primary taxpayers on joint returns at each income level except the very lowest tend to pay considerably less in tax than their unmarried counterparts. It should be noted that the designation as primary or secondary taxpayer is strictly the filers' choice, but, in most cases, one would expect the primary taxpayer to be the one with the higher income--the few exceptions being the ones showing up in the very low income classes in Figure 4. Obviously, in the majority of cases, the additional wages of the secondary taxpayer do not push the primary taxpayer into as high a tax bracket as would have applied had he (or she) been filing singly and using the "single taxpayers" tax rate schedule.

One way of looking at secondary taxpayers is to say that they start paying taxes at the rate where the primary taxpayer's income left off. This means that a secondary taxpayer with a small income will, if married to a taxpayer with a large income, pay more in tax than an equivalent single person. However, as is shown in Figure 5, in the range between $30,000 and $80,000, even secondary taxpayers, as a group, pay slightly lower taxes than their single counterparts.

When data for primary and secondary taxpayers are added together, it is apparent that, as a group, their tax rates are lower than those of single people, except at the "Under $10,000" adjusted gross income level, as is shown in Figure 6. (These lower income classes obviously include a lot of part-time workers married to full-time workers with larger incomes.) So it does not appear that members of two-earner couples, as a group, are paying taxes at a higher rate than are their single counterparts. It should, however, be noted that the data on members of two-earner couples represent an average of two groups: the "losers," who are generally members of married couples where the earnings of the two partners are very close, and whose combined income taxed at "joint" rates puts them into a higher tax bracket than the "single" tax rate schedule would have imposed on each of them filing on their own; and the "winners," who are generally couples with great differences in income levels, where use of the "single" tax rate schedule would have placed the higher earnings of one member into a much higher tax bracket.

Finally, it should be noted that effective tax rates paid members of two-earner couples are affected by more than just the applicable tax rate schedules. The fact that they appear to be paying taxes that are slightly lower than single taxpayers may simply be a reflection of the fact that they have more exemptions than their single counterparts. Again, this analysis is only meant to scratch the surface, and demonstrate the type of statistics that can now be produced with our expanded data base.

Future Work

This new data set will permit us to look at the characteristics of those couples who pay more and those who pay less than unmarried individuals with equivalent incomes. Aside from salaries and wages, future analyses could divide income from unincorporated businesses, pensions and annuities, social security income, and unemployment compensation between primary and secondary taxpayers on joint returns.

We are still in the midst of adding enhancements to the 1993 data base. These include gender coding and occupation coding. When we are done, our plan is to release at least two articles in the *Statistics of Income Bulletin*: one will show the typical SOI data on income, deductions, and tax items, but rearranged by family; the other will present salaries and wages, and perhaps other sources of income, by individual, with the data classified by sex of taxpayer. We also intend to examine information on individuals who do not file tax returns, but for whom we were able to gather income data from information documents.

It is the authors' hope that the data base described in this paper, and similar ones for future years, will provide opportunities for new break-throughs in the measurement of the operations of the Internal Revenue law.

Acknowledgments

The authors would like to extend their thanks to Marianne Cooley and Lorne Woodworth of the IRS Computing Center in Detroit, Michigan, for help with computer processing. Thanks also to Wendy Alvey of the Statistics of Income Division for her many helpful comments.

References

Czajka, John L. and Schirm, Allen L. (1993), "The Family That Pays Together: Introducing the Tax Family Concept, with Preliminary Findings," *Proceedings, Section on Survey Research Methods*, American Statistical Association.

Hostetter, Susan, Czajka, John L., Schirm, Allen L. and O'Conor, Karen (1990), "Choosing the Appropriate Income Classifier for Economic Tax Modeling," *Proceedings, Section on Survey Research Methods*, American Statistical Association.

Hostetter, Susan and O'Conor, Karen (1991), "Satisfying the Need of Income Policy Modelers While Preserving the Reliability of Descriptive Statistics," *Proceedings, Section on Survey Research Methods*, American Statistical Association.

Internal Revenue Service, *Statistics of Income--1993, Individual Income Tax Returns*, Publication 1304 (Rev. 3-96).

Sailer, Peter and Weber, Michael (1996), "Creating Household Data from Individual Income Tax Returns," *Proceedings, Section on Government Statistics*, American Statistical Association.

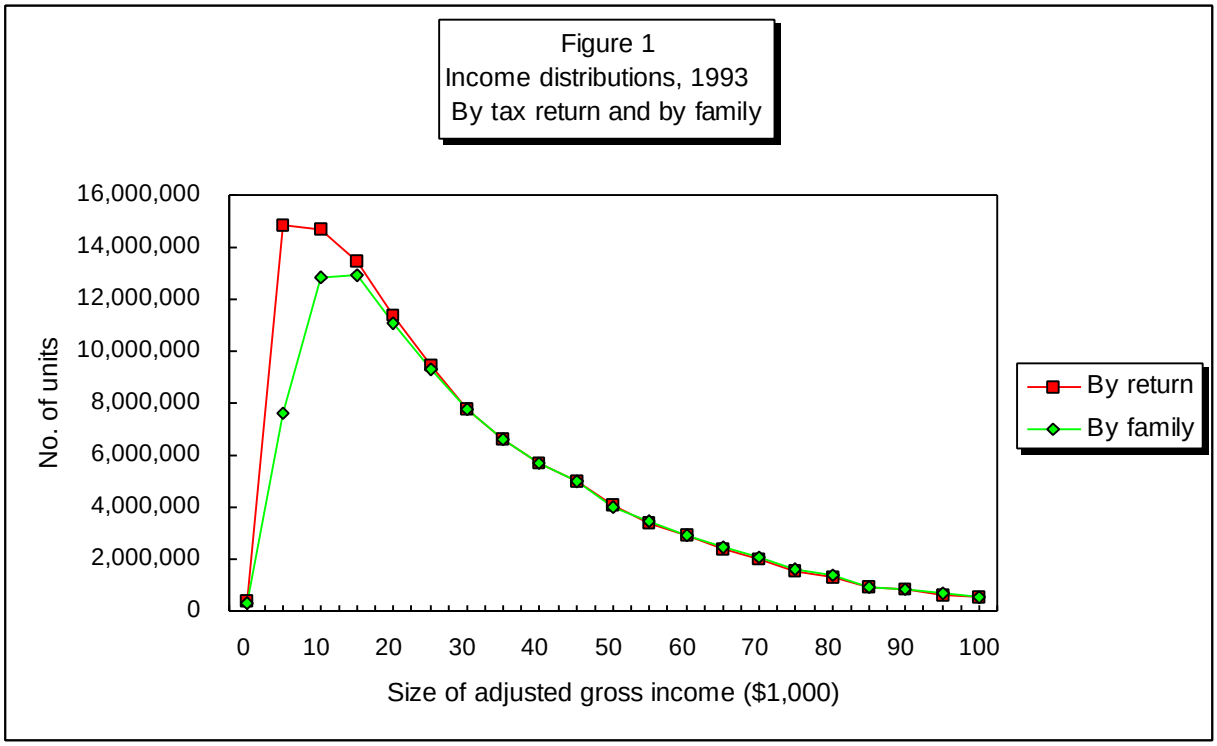
Figure 1
Income distributions, 1993
By tax return and by family
16,000,000
14,000,000
12,000,000
10,000,000
8,000,000
6,000,000
4,000,000
2,000,000
0
No. of units
0
10
20
30
40
50
60
70
80
90
100
Size of adjusted gross income ($1,000)
By return
By family

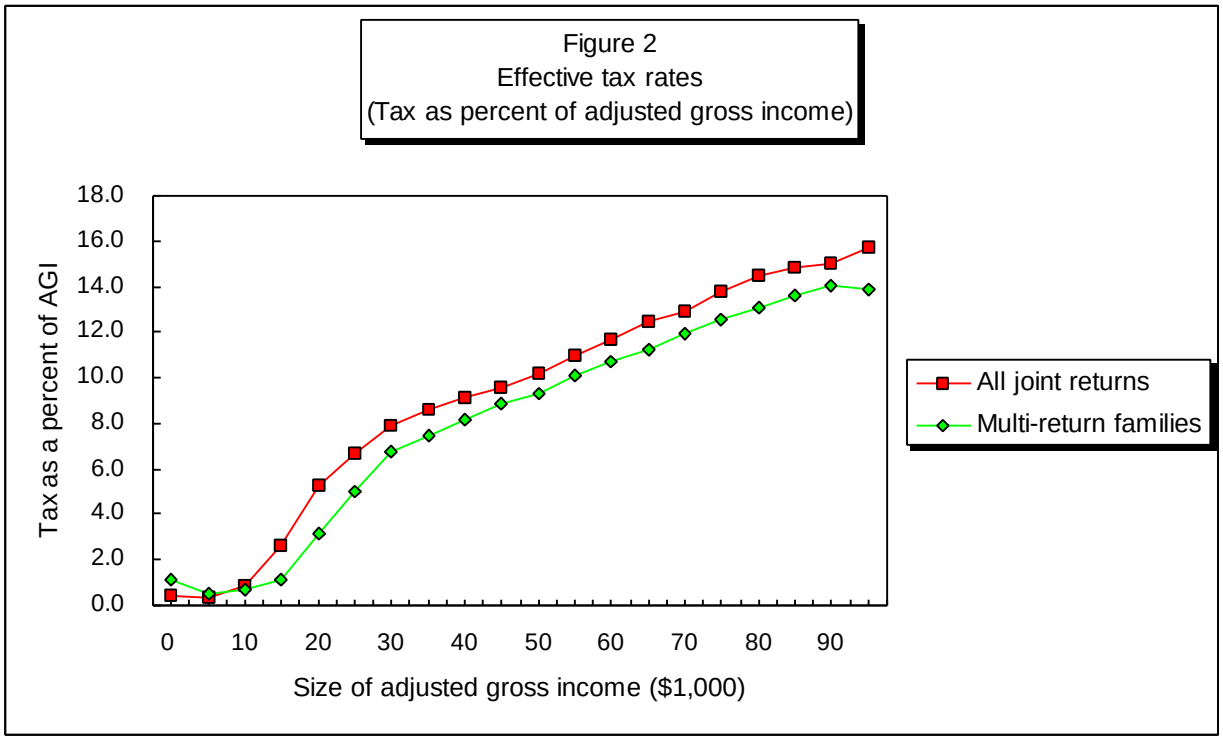
Figure 2
Effective tax rates
(Tax as percent of adjusted gross income)
18.0
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
Tax as a percent of AGI
0
10
20
30
40
50
60
70
80
90
Size of adjusted gross income ($1,000)
All joint returns
Multi-return families

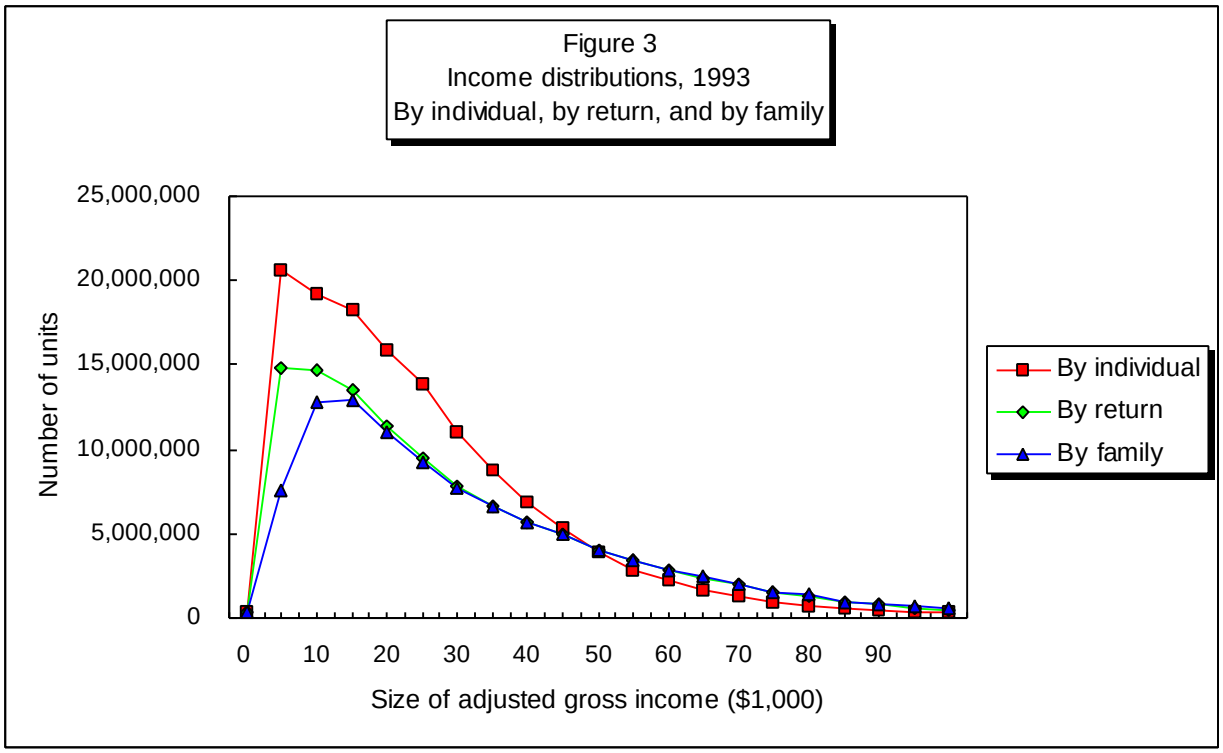
Figure 3
Income distributions, 1993
By individual, by return, and by family
Number of units
25,000,000
20,000,000
15,000,000
10,000,000
5,000,000
0
0
10
20
30
40
50
60
70
80
90
Size of adjusted gross income ($1,000)
By individual
By return
By family

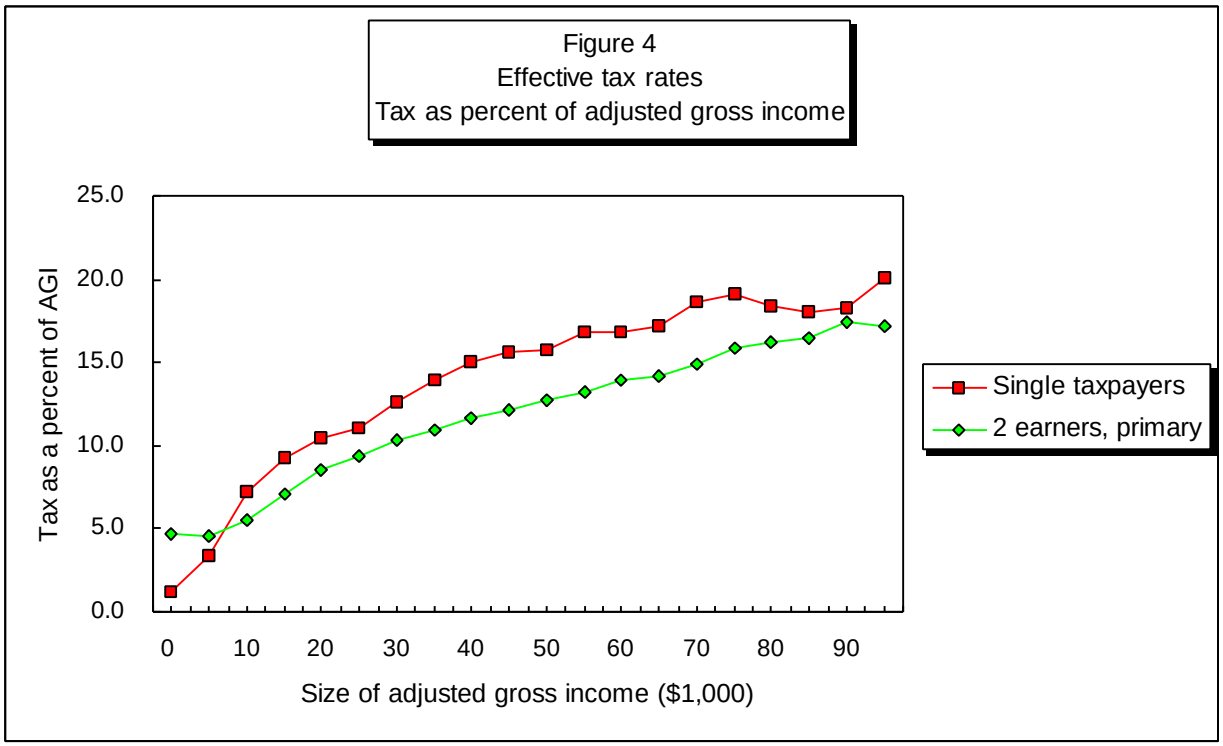
Figure 4
Effective tax rates
Tax as percent of adjusted gross income
Tax as a percent of AGI
25.0
20.0
15.0
10.0
5.0
0.0
0
10
20
30
40
50
60
70
80
90
Size of adjusted gross income ($1,000)
Single taxpayers
2 earners, primary

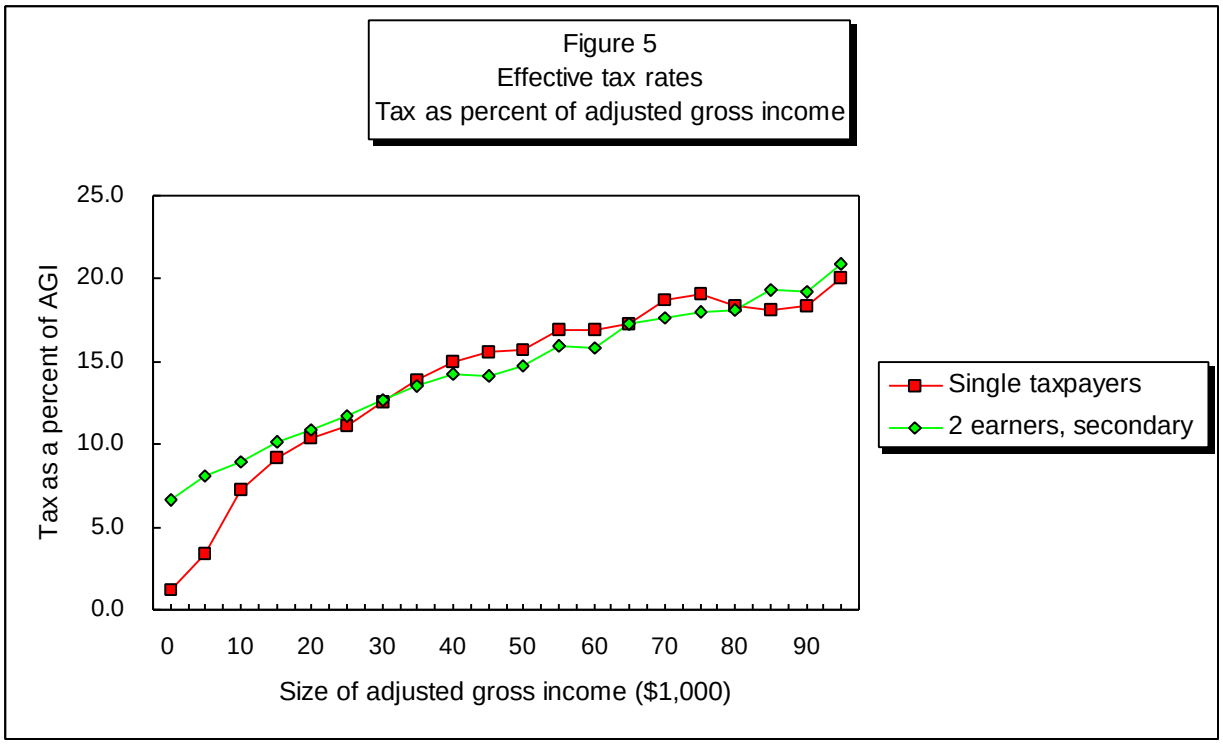
Figure 5
Effective tax rates
Tax as percent of adjusted gross income
Tax as a percent of AGI
25.0
20.0
15.0
10.0
5.0
0.0
0
10
20
30
40
50
60
70
80
90
Size of adjusted gross income ($1,000)
Single taxpayers
2 earners, secondary

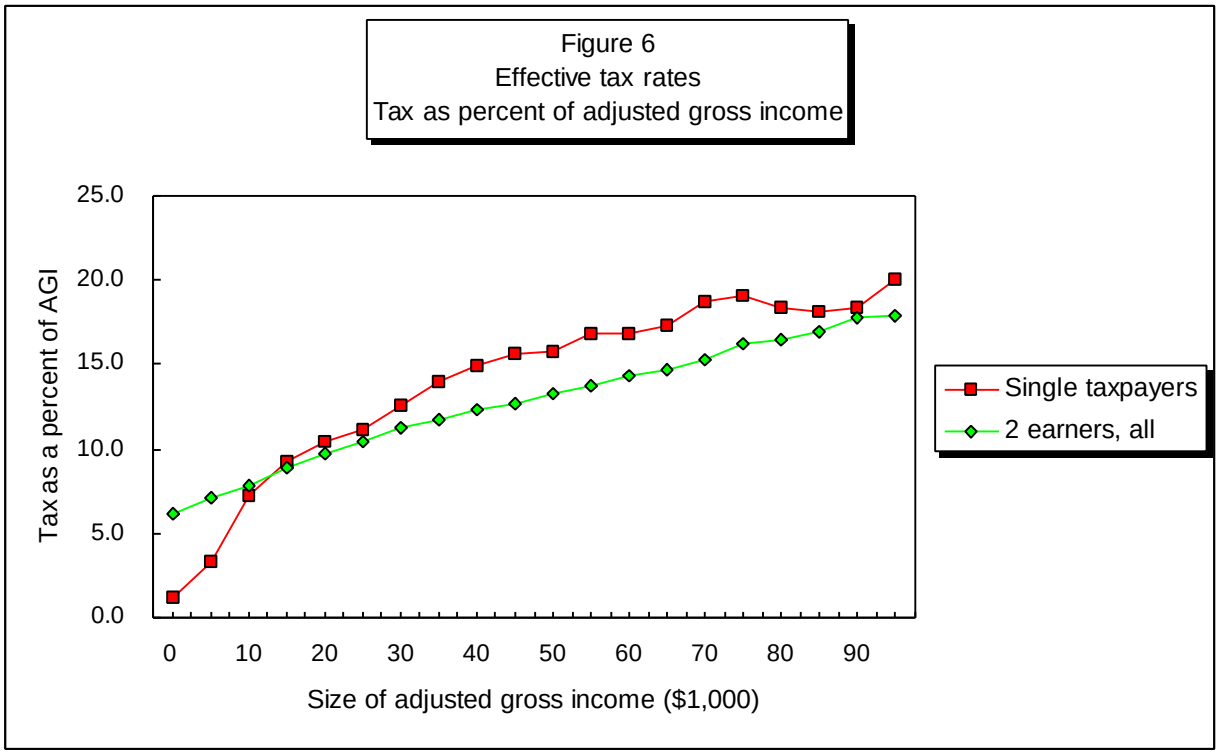
Figure 6
Effective tax rates
Tax as percent of adjusted gross income
Tax as a percent of AGI
25.0
20.0
15.0
10.0
5.0
0.0
0
10
20
30
40
50
60
70
80
90
Size of adjusted gross income ($1,000)
Single taxpayers
2 earners, all