Income Taxation of U.S. Households: Facts and Parametric Estimates

Nezih Guner, Remzi Kaygusuz and Gustavo Ventura*

July 2013

Abstract

We use micro data from the U.S. Internal Revenue Service to document how Federal Income tax liabilities vary with income, marital status and the number of dependents. We report facts on the distributions of average taxes, properties of the joint distributions of taxes paid and income, and discuss how taxes are affected by marital status and the number of children. We also provide multiple parametric estimates of tax functions for use in applied work in macroeconomics and public finance.

JEL Classifications: E62, H24, H31
Key Words: Taxation, Tax Progressivity, Households.

*Guner, ICREA-MOVE, Universitat Autonoma de Barcelona and Barcelona GSE (email: nezih.guner@movebarcelona.eu); Kaygusuz, Faculty of Arts and Social Sciences, Sabanci University, Turkey (email: rkaygusuz@sabanciuniv.edu); Ventura, Department of Economics, Arizona State University, USA (email: gustavo.ventura@asu.edu). Guner acknowledges support from European Research Council (ERC) Grant 263600. We thank the editor, Gianluca Violante, an anonymous Associate Editor and two anonymous referees who provided excellent comments. We also thank Ezgi Kaya for excellent research assistance. Earlier versions of this paper circulated under the name "Income Taxation of U.S. Households: Basic Facts". The usual disclaimer applies.

1 Introduction

This paper has two goals. First, we aim to systematically describe how income taxes paid by a cross-section of U.S. households vary according to their income, marital status and number of dependent children. Second, we provide estimates of *effective tax functions* that capture the observed heterogeneity in the data that can be readily used in applied work.

These goals are motivated by the large and growing body of literature that utilizes dynamic macroeconomic models with heterogeneous households; see Heathcote, Storesletten and Violante (2009) for a recent survey. This literature has studied how existing models can account for properties of actual earnings, income and wealth distributions, and has used such models to address a host of policy questions.[1] As an input in this process, a large body of work, old and new and from related fields, documented the empirical properties of such distributions. However, the distribution of taxes effectively paid by households and the marginal tax rates that they face have received much less attention. This paper fills this void, by systematically documenting basic properties of the structure of income taxation for a cross section of U.S. households.

The model economies studied in the above mentioned literature require, in accordance with data, a mapping of household's income to taxes paid conceivably depending on the household's marital status, the presence of children and retirement status. This naturally matters when asking how well models with heterogeneous households match distributional properties of data, as well as for the fruitful use of these frameworks to address policy questions. A *ready-to-use*, systematic representation of this mapping is not currently available for different types of households, and we offer it here. Therefore, we provide different parametric estimates of effective taxes as a function of household's income for different types of households; all, married, unmarried, with and without dependent children. We also provide estimates for special cases; with and without positive social security income, with exclusively labor income and with imputations for state and local taxes.

We aim at providing estimates of taxes *effectively* paid by households according to their

[1] There is a large literature that uses dynamic macroecomic models with heterogenous agents to study tax reforms. See Ventura (1999), Altig, Auerbach, Kotlikoff, Smetters and Walliser (2001), Castañeda, Díaz-Jiménez and Ríos-Rull (2003), Díaz-Jiménez and Pijoan-Mas (2005), Nishiyama and Smetters (2005), Conesa and Krueger (2006), Erosa and Koreshkova (2007), Conesa, Kitao and Krueger (2009) and Guner, Kaygusuz and Ventura (2012 a-b), among others. Guvenen, Kuruscu and Ozkan (2009) study the effect of taxes on human capital accumulation and inequality.

income, family status and number of dependents. We use a large cross-sectional data set from U.S. Internal Revenue Service ('Public Use Tax File'), that is ideal for these purposes as it is representative of the entire set of U.S. taxpayers. For a notion of effective average tax rates out of Federal Income taxes, we find a substantial degree of heterogeneity generated by the U.S. income tax system and the underlying distribution of income. As we document, average rates increase non trivially with income, and this is reflected in the distribution of average tax rates paid. For instance, if we rank the married households by average tax rates that they face, average taxes for married (unmarried) households at top 1% are in excess of 27.7% (23.0%), while the median tax rate is about 8.5% (6.1%). These facts, in conjunction with the substantial income dispersion that we document in this data, implies that households at the top of the income distribution account for the bulk of taxes paid; the top 5% accounts for nearly 55.2% of all taxes paid, whereas the top 1% accounts for about 35.8%.

Using this data, we estimate four functional forms for effective tax rates. In each case, we report estimates for all households, as well as for married and unmarried households with different numbers of dependent children. We first estimate a two-parameter specification, which we refer to as the *log* specification (used by Guner, Kaygusuz and Ventura 2012-a, 2012-b). Our second set of estimates, another two-parameter specification, pertains to the functional form in Heathcote, Storesletten and Violante (2011), which we refer to as the *HSV* specification. Our third case is a three-parameter specification, which we refer to as the *Power* specification (a version of power function is used by Guvenen, Kuruscu and Ozkan 2009). Finally, we estimate the same functional form used in Gouveia and Strauss (1994), who provided estimates of tax functions for all taxpayers using the U.S. tax structure prevailing in 1989.[2] We find that while all specifications provide tax schedules for average rates that are similar and that track raw data quite well, the Gouveia and Strauss specification generates a better statistical fit than other specifications. We also compare the marginal taxes implied by these parametric specifications with the data. From the data, we

[2]Several papers estimated effective tax rates for the use in representative-agent models. See, for instance, Joines (1981), Seater (1982), Barro and Sahasakul (1983) for papers that used IRS data, and Mendoza, Razin and Tesar (1994) who estimated effective taxes for a large set of countries using national accounts and revenue statistics. Barro and Redlick (2010) use tax imputations from the TAXSIM program. Differently from these papers, Gouveia and Strauss (1994) used IRS data to estimate *tax functions*–average tax rates as a function of household income. Huggett and Parra (2010) estimate tax functions for retired and non-retired households from tabulated IRS data.

compute both effective marginal rates, i.e. changes in tax liabilities associated with changes in household income, and statutory marginal tax rates. We find that effective marginal tax rates are below the statutory ones, especially for high income levels. We also find that the implied schedules of marginal tax rates, computed from the parametric tax functions, track the effective marginal tax rates from data reasonably well. Similarly, they always below the marginal rates computed from statutory data. In addition, the schedules become essentially flat after relatively low levels of income under the Gouveia and Strauss specification. The other specifications, in particular *Power* and *log* specifications, in contrast, generate marginal rates that are closer to the rates implied by data at high incomes.

The paper is organized as follows. Section 2 presents the basic data that we use in our calculations afterwards. Section 3 describes how effective average tax rates by households vary in cross section according to income, marital status and the number of dependent children. Section 4 reports facts on the distribution of tax rates, taxes paid and after-tax income distribution. Section 5 offers the parametric estimates for tax functions. Section 6 documents marginal tax rates from the data and from the parametric estimates. Section 7 concludes.

2 Data

We use data from the Internal Revenue Service *2000 Public Use Tax File.* With 145,663 records, it is a representative subsample of the universe of U.S. taxpayers who filed taxes in the year 2000. Since this data effectively contains no restrictions on income, either at the bottom or at the top, it allows for a representation of income and tax liabilities across all income levels.[3]

The notion of *income* that we use is standard in cross-sectional studies and encompasses all income flows accruing to households; labor income, asset income from different sources and transfers. It corresponds closely to the notion employed by Gouveia and Strauss (1994). We define *income* to include

- Salaries and wages;
- Interest income (taxable and not taxable);

[3] For details on variable definitions, weights used and other technical details, see the *Individual Tax File Sample Description* booklet that accompanies the data.

- Dividends, interest income and royalties;
- Realized capital gains;
- Business or professional income;
- Total pensions and annuities received plus taxable IRA distributions;
- Unemployment compensation;
- Social Security benefits;
- State income tax refunds and alimony received.

It is worth emphasizing that the notion of income that we use is different from the legal notions *Adjusted Gross Income* and *Taxable Income.* In the 2000 tax forms, Adjusted Gross Income was computed by subtracting from all reported income flows IRA and Keough accounts contributions, moving expenses, student loans interest, alimony paid, contributions to medical income savings accounts, among other items. Taxable Income is obtained by subtracting personal exemptions and deductions from Adjusted Gross Income.[4]

Our notion of Federal Income taxes is comprehensive as well. It corresponds to total income taxes owed after Credits (including the Earned-Income Tax credit).[5] From this notion of tax liabilities, we calculate for our purposes *effective* average tax rates. Marginal tax rates that we report in the next section correspond to the *statutory* marginal rates for each household given their taxable income in the data.

Sample Restrictions Households are included in the sample if (i) their income is strictly positive; (ii) their average tax rate is less than 40%. The second restriction eliminates those with reported taxes higher than the top *statutory marginal* tax rate, 39.5%. The resulting average level of income is US$ 53,063.[6]

[4] In terms of deductions, households can choose between a lump-sum *standard deduction* or multiple *itemized* deductions, the most common of which corresponds to mortgage interest paid.

[5] More specifically, we use the variable `TOTAL INCOME TAX` (E06500) in the 2000 Public Use Tax File.

[6] The corresponding average level of household income in the commonly used data from the Current Population Survey (CPS) data is somewhat higher: US$ 57,121.

2.1 Data Limitations

Despite the wealth of information in the IRS data and its advantages, there are important limitations in this data that the reader should be aware of. Several issues are worth noting.

First, there is no detailed information on Social Security taxes paid, neither by households or individuals themselves, nor by employers.[7] Furthermore, social security taxes paid by an individual are not a linear function of earnings due to the cap on taxable earnings. As a result, since the unit of observation is the *household*, not an individual, and social security taxes are based on individual labor income, it is not possible to impute social security taxes for a majority of individuals in the sample (i.e. married ones). Altogether, knowing the labor income of a married household is not sufficient to infer completely the social security taxes of individual spouses. All this implies that from the IRS data we cannot construct a broad notion of Federal tax liabilities in cross section (income plus social insurance taxes). For this reason, we focus only on the relationship between household income and Federal Income taxes.

Second, as it is the case in commonly available data sets, there is no information on employer's contributions to health insurance and pension plans. This is arguably a substantial component of an individual's labor income. Since these contributions are not subject to taxation, they do not appear in individual tax forms and thus, they are absent in the data.

Finally, there is no information on the individual contributions to health insurance and pension plans within firms. Labor income reported in the data set is net of these contributions.[8] Hence, this point and the previous one imply that total labor income accruing to households is underestimated in the IRS data as it is in other commonly used data sets.

2.2 Statutory Federal Income Taxes in 2000

Before presenting and discussing results on taxes paid by income, household structure and number of children, it is worth reporting the structure of statutory income taxes in 2000. Table 1 summarizes this information for three relevant categories: *married filing jointly, single and head of household*. Tax brackets are presented as defined in the law, according to

[7] There is only information on social security taxes on tip income, and amounts of Social Security and Medicare (FICA) taxes paid in excess of those required.

[8] The entry for `Wages and Salaries` in the data corresponds to what households report in the first entry in their 1040 forms. The number entered is net of individual contributions to health and pension plans.

the legal notion of *Taxable Income.*

As the table shows, marginal tax rates range from 15% to 39.6%. The standard deduction for married people is not twice the standard deduction for singles. A similar remark applies to the width of the tax brackets. Very importantly, there is a wide range of income for which statutory marginal tax rates are unchanged; for instance, from $ 43,850 to $ 161,450 for married households, marginal rates change by only three percentage points (from 28 to 31 percent). Afterwards, statutory marginal rates increase non-trivially for high- income earners; to 36 and 39.6 percent, respectively. To calculate their taxable income, households have the option of choosing the standard deduction, or a host of *itemized* deductions (e.g. mortgage-interest deduction) that naturally become more attractive at high income levels. Altogether, these features contribute to generate the substantial differences in average tax rates across income levels that we document later, the rapid rise of average rates with income at relatively low income levels, as well as and the slow rise of average rates at relatively high income levels.

2.3 Descriptive Income Statistics

For a better understanding of the facts about tax liabilities in cross section, it is important to report on the properties of the distribution of income in the tax data. This is of interest, since there are no top-coding restrictions as in other commonly used data sets and as a result, the data is representative of the entire universe of U.S. taxpayers.

Table 2 summarizes the properties of the income distribution and highlights the substantial degree of concentration of income at the top. The richest 20% of households earns about 61.3% of total income, whereas the top 10%, 5% and 1% earn about 46.5%, 35.9% and 20.9%, respectively.[9] Table 2 also shows the differences between the notion of income that we consider and the legal notions of Adjusted Gross Income (AGI) and Taxable Income. The results show that the distribution of AGI is very close to the distribution of regular income. Nevertheless, as lump-sum deductions are concentrated at the bottom and have a large impact there, the distribution of taxable income is non-trivially more concentrated than the distribution of income. As a result, the Gini coefficient increases from 0.59 for regular income to about 0.63 for taxable income.

[9] From the IRS data one can also assess the importance of the *very rich*. For instance, the top 0.5% earned in 2000 about 16.2% of income.

It is important to relate these summary distributional statistics to standard summary measures of income inequality. For instance, CPS data indicates that each quantile earned in 2000 about 3.6, 8.9, 14.8, 23.0 and 49.8 percent of income respectively, whereas the top 5% earned about 22.1 percent with a Gini coefficient of about 0.462.[10] The tax data shows that each quintile earned about 2.0, 6.1, 11.3, 19.1 and 61.3 percent, respectively, whereas the top 5% earned about 35.9% with a Gini coefficient of about 0.59. Clearly, and as also emphasized by others (e.g. Piketty and Saez 2003), the degree of income concentration from tax data is substantially higher than the degree of income emerging from standard data sets. However, this degree of concentration is quite close to the one found in the Survey of Consumer Finances (SCF). This is not surprising as the SCF does not censor the income of richer households. For instance, Diaz-Gimenez, Glover and Rios-Rull (2011) using the SCF found that for 2007 each quintile earned about 2.8, 6.7, 11.3, 18.3 and 60.9, whereas the top 5% earned about 35.9% with a Gini coefficient of about 0.575.[11]

Table 3 shows the varying composition of income as income increases. The third and fifth columns display the fraction of income corresponding to capital income at different quintiles for two concepts of capital income. The first concept of capital income, includes all interest income, dividends, 1/3 of business income, capital gains, rental + royalties income and 1/3 of farm income.[12] The more comprehensive second one, adds to the previous one all pension and annuity payments. In both cases, and as expected, capital income as a share of total income rises rapidly as income goes up. Note that at the very top, more than 40% of income accrues from capital income under the first definition (about 41.4%), whereas under the second notion about 54.6% of income derives from capital income. This is obviously relevant for economic purposes; high income households face much higher marginal tax rates (see below)

[10] See `http://www.census.gov/hhes/www/income/data/historical/household/index.html`.

[11] Data sets such as the CPS and PSID underrepresent the top of the distribution of labor earnings and income. It is also known that there is underreporting in all income categories in the CPS, except in wage income. The data from the IRS is likely to exhibit underreporting of income in some categories (e.g. self-employed income). For instance, internal research from the IRS has found that the underreporting of individual income for tax purposes is the largest contributor to the difference between estimated tax liabilities and taxes effectively paid (tax gap). For instance, in 2006, underreporting of income accounted for 52% of the tax gap of *all* Federal taxes. See Black, Bloomquist, Emblom, Johns, Plumley and Stuk (2012) for details. According to Johns and Slemrod (2010), business income, as opposed to wages or investment income, accounts for about two-thirds of the understated individual income. Furthermore, misreporting as a percentage of adjusted gross income is increasing in adjusted gross income (reaching about 19% in 0.99 to 0.995 percentiles of the AGI distribution).

[12] We allocate 1/3 of business and farm income according to standard estimates of the share of capital income in total income.

and capital income is concentrated *there*. Transfers consists of unemployment insurance and social security payments, and labor income constitute the remaining component of household income. For most households, labor income constitutes the major source of income, but at higher levels of income, income from capital becomes increasingly important. Transfers constitute a small fraction of household incomes and their contribution is hump-shaped, with households around the middle of the distribution of household income receiving the largest share from transfers.

3 Income Taxes in Cross Section

In this section we report basic facts on how average and statutory marginal tax rates vary according to our broad notion of income, marital status and the number of children. For different levels of household income (quantiles as well as bottom and top percentiles), we calculate the *effective* average rate, defined as the average ratio of taxes paid to household's income in the corresponding income category. The marginal rate that we report corresponds to the one encountered by households in their actual tax filing, averaged across all households in the income category. Thus, marginal tax rates reported in this section correspond to mean *statutory* marginal rates.[13]

3.1 Married and Unmarried Households

Table 4 shows the findings for all households, as well as for married and unmarried households as a group. Married households correspond to those filing as *married filing jointly*, whereas unmarried households encompass all those filing as *single* and as *head of household*. We explicitly include head of households in our unmarried group, as this category is designed for households headed with unmarried individuals with dependents. Tables 5 and 6 show in detail the married and unmarried groups, according to the number of dependent children present in the household.

A central finding from tables 4, 5 and 6 is that effective average rates increase substantially as income increases. Increasing household income from the central quintile (40% to 60%) to the top quantile, increases the mean, effectively-paid average tax rate, from about 3.9% to 14.0% for married households, and from 8.2% to 16.7% for unmarried households. In terms of

[13] In section 5, we infer effective marginal tax rates from data.

statutory marginal rates, the increase goes from 13.5% to 27.7% for married households, and from 16.2% to 28.6% for unmarried ones. For households at the top 1%, average (marginal) rates are 23.1% (36.3%) for married households and 21.7% (34.6%) for unmarried ones. Hence, from these findings it is clear that there is a substantial degree of tax progressivity built into the tax system. This is the natural result of the observed degrees of income concentration, in conjunction with a tax schedule with increasing marginal tax rates as a function of income. Tables 4, 5 and 6 also report the standard deviation of average taxes paid for different income levels. The standard deviations relative to average tax rates are quite high at the bottom of the income distribution as many households pay zero taxes. They remain relatively constant and increase again slightly at the top of the income distribution. For all households, for example, standard deviations of average taxes for the richest 1% of households is about 10%, about half of their average tax rate (22.9%).

From Tables 4, 5 and 6 it also transpires that there are substantial differences in average rates between married and unmarried households. At low levels of income, effective rates are substantially higher for unmarried households, while these rates subsequently converge as income increases, and eventually become higher for married households at top levels of income. Figure 1 illustrates these differences in tax rates.[14] These differences are due to a host of factors; differences in the levels of income concentration between married and unmarried households, differences in standard deductions and differences in the width of tax brackets and children. These latter factors are arguably more important in reducing average rates at lower levels of income. For instance, children are concentrated in married households and they lead to higher personal exemptions and tax credits, thereby reducing average rates for these households.

Finally, Tables 4, 5 and 6 also report standard deviation of average tax rates across households at different income levels. As Table 4 documents there is significant variation across household in taxes paid, especially at low levels of income, possibly reflecting the fact that a large number of these households do not pay any tax. For the middle quintile (40-60%), the standard deviation of average tax rates (4.3%) is more than half of average tax rate (7%). Not surprisingly, there is less variation at higher levels of income. For households who are at top 1% of income distribution, the standard deviation of taxes (9.7%) is less than

[14] Income brackets for all, married, unmarried categories are the same and calculated using income data for all households.

half of their average tax rate (22.9%).

We now try to illustrate the effects of the differential tax treatment of married and unmarried households in the United States. To isolate these effects, we use data that is *not* affected by the presence of children – first panel in tables 5 and 6. Consider for instance an average married household at the top quintile of the distribution. If both wife and husband have the same income, their tax liabilities would likely be higher when married; they would pay as a married household an effective average rate of about 14% whereas as two single individuals, they would pay roughly 8% each. At another extreme, if only one of them earns all the household income, the average rate would be in excess of 17% if each filed as single, whereas it would be about 14% if they filed as a married couple. Other combinations can be constructed from these tables, reflecting the fact that two partners of similar incomes face a tax *penalty* if they marry, whereas those with sufficiently different incomes face a tax *bonus* or subsidy.[15]

Overall, the discussion above is driven by the fact that in the United States, the unit subject to taxation is the *household*, not the individual. The economic implications of this fact go beyond relative payments when married or not. Consider again a married household with no children with an income level at the top quintile. Table 5 indicates that this household faces a statutory marginal tax of about 27.8%. If all income is earned by one household member, the marginal tax on the first dollar of income earned by the other household member is also taxed at the same rate, 27.8%. This naturally creates large disincentives for labor supply of secondary earners. In contrast, if her/his income were treated as an individual income, the marginal tax rate would be substantially lower. If children are present, the same logic applies.[16]

[15] See McCaffery (1997) for a detailed account of the US tax system's treatment of married and single households. On the optimal taxation of couples, see Boskin and Sheshinski (1983), Apps and Rees (2010), Alessina, Ichino, and Karabarbounis (2010), Kleven, Kreiner, and Saez (2009) and Guner, Kaygusuz and Ventura (2012-a).

[16] In Guner et al (2012-b), we show that these features of the U.S. tax law have large effects on labor supply of married females. Kaygusuz (2010) studies how much changes in taxes contributed to the growth of married female labor supply in the US since 1970s. Prescott (2004) studies how cross-country differences in taxes affect cross-country differences in hours per worker. Bick and Fuchs-Schundeln (2012) and Chakraborty, Holter and Stepanchuk (2012) look at the relation between taxes and household labor supply across countries.

3.2 The Role of Children

Tables 5 and 6 illustrate the quantitative importance of children in affecting average effective rates for households. As we mentioned earlier, for unmarried households, we use information from the *single filing* category for those without children, whereas for those with children we use information from the *head of household* category.

For married households, children reduce effective rates although the overall effect varies across income levels. Households with income at the top 20%, face an effective average rate of about 15.0% when no children are present, a rate of about 13.2% when two children are present, and a rate of 11.7% when more than two children are present. Therefore, for these households, at extremes, the reduction in effective rates driven by the presence of children is in the ballpark three percentage points. At very high levels of income, the corresponding reductions in average rates is smaller. Meanwhile, for poorer married households the reduction is naturally much higher than at the top; nearly five percentage points at the central quantile. This is not at all surprising: children disproportionately affect tax liabilities of poorer households via lump-sum personal exemptions and tax credits.

For unmarried households, the patterns just described above are similar but more pronounced; households at the top 20%, face an effective average rate of about 17.3% when no children are present, a rate of about 13.5% when two children are present, and a rate of 12.2% when more than two children are present. For households at the central quintile, the reduction associated to the presence of children in the ballpark of eight percentage points.

3.3 The Role of Income Sources

How do tax liabilities vary according to income sources? We provide an answer to this question in Table A2 in the Appendix, by examining average tax rates at different quintiles of the joint distribution of labor and capital income. As both capital and labor income increase along the diagonal, tax rates naturally increase. The data reveals that as capital income increases across quintiles, for a given labor income quintile, there are essentially no changes in average tax rates. On the other hand, holding a capital income quintile constant and increasing labor income across quintiles, tax rates tend to increase. This finding is consistent with the importance of labor income as a source of income, as Table 3 shows.

3.4 State and Local Taxes

How do state and local taxes vary as income changes? Our data allows to provide a partial answer to this question, as the I.R.S. data on state and local taxes is available only for those households who take itemized deductions in their filing of Federal Income taxes.

Table A3 in the Appendix presents state and local taxes that households pay at different levels of household income. Since itemized deductions are rarely taken at low levels of income, there are essentially no observations of state and local taxes at the bottom income quintile as the table shows. On average, state and local taxes amount to about 4-5%. Poorer households face lower state and local taxes than richer households, but the differences are much smaller than the ones we observe for federal income taxes. The overall structure of state and local taxes is rather flat as a function of income, as Table A3 demonstrates.

4 Income Taxes and Inequality

In this section, we report on a series of facts related on the distribution of taxes paid, the relative contribution of different income groups, as well as the relationship between taxes and after-tax inequality. We aim at presenting a snapshot on the relationship between the structure of income taxes, and the underlying degree of income inequality.

Tax Rates and Taxes Paid Table A4 in the Appendix describes the basic features of the distribution of average tax rates across households. As the table illustrates, a substantial fraction of households has *no* tax liabilities: this occurs for about 14.5% of the married group and for about 31.8% of the unmarried one. Median and mean effective tax rates are on the low side for both groups, with a median rate for married households of about 8.5% and a mean rate for married households of about 8.8%. For unmarried households, the median rate is of about 6.1% whereas the mean rate amounts to 6.4%.

The bottom panel of Table A4 shows the tax rates defining the top percentiles. Households at the top of the distribution face significantly higher average rates than those around the middle: the ratio of tax rates defining the bottom 95% to the median is in excess of a factor of 2 for married households, and of a factor of nearly 3 for unmarried households.

How tax liabilities are distributed? Table 7 answers this question, by calculating the share of total taxes paid by different percentiles of the income distribution. The top 20% of households earns about 61.3% of total income and pays more than three quarters of total taxes (79.1%). Similarly, the top 1% earns about 20.9% of total income, yet it accounts for about 35.8% of total tax collections.[17]

After-Tax Income Distribution How much do the before and after-tax income distributions differ? Table 8 shows income-distribution statistics before and after taxes.

Despite the vast heterogeneity we documented earlier in terms of income and tax payments, the results show that the before-tax and the after-tax income distributions effectively differ, albeit in relatively moderate amounts. The before-tax shares of the 60-80% quintile and the top quintile are, respectively, 19.1% and 61.3%, whereas the corresponding shares of the after-tax distribution become 20.2% and 58.5%. The largest changes occur for the top 1%, where shares decline from 20.9% to 18.4%. Overall, the changes in shares lead to the Lorenz curve associated to the before-tax income distribution to lie below the Lorenz curve of the after-tax distribution. These changes are summarized in the decline of the Gini coefficient due to the presence of income taxes, from about 0.59 to about 0.56.

How Progressive is the U.S. Tax System? There are different answers to this question. One the one hand, a clear picture emerges from our findings. Effective tax rates on most households are relatively low (below 10%) and differ substantially from those at the top. For instance, married household around median income experience tax rates around 4%, while those at the top 1% face tax rates of around 23%. Furthermore, as we showed earlier, taxes paid are heavily concentrated at the top. In a nutshell, the provisions in the law, in conjunction with the observed dispersion in income lead to the finding that the bulk of tax payments are concentrated in upper income households and that a large fraction of U.S. households have effectively no tax liabilities. From this perspective, the answer to the question above is that there is substantial progressivity in the tax burden as measured by effective, average tax rates. Put differently, moving a hypothetical household along the income ladder implies substantial increases in average tax rates. These findings are reflected in the comparison between the distributions of income before and after taxes.

On the other hand, tax rates at the very top of the income distribution are essentially

[17] The facts on the distribution of individual income tax liabilities are in line with estimates from the Congressional Budget Office (2012) for the year 2000. They estimate a share of taxes paid by the highest quintile of about 81.2%, and a share of taxes paid by the top 1% of about 36.6%.

constant as income changes. Once high income levels are reached, effective tax rates do not change. We calculate that at *ten* times the level of household income in the sample (about \$ 530,630 in 2000 dollars), the average tax rate for a married household was around 24%. At *fifteen* times the level of mean income (about \$795,945) the rate was about 25%, while at twenty times mean household income (about \$1,061,260), the rate effectively unchanged at around 25%. In other words, income tax rates at the top are flat and do not approach the statutory marginal rate on income that these households faced (39.6%). This occurs due to a host of factors, and underlies debates and proposals on tax reform.

5 Parametric Estimates

In this section, we provide estimates of tax *functions* for applied use. Specifically, we posit parametric functional forms for effective average tax rates, and estimate the relevant parameters for all households, married and unmarried households, distinguishing by the number of dependent children. We also provide estimates for a number of special cases. For all parametric functional forms, we represent household income as multiples of mean household income in the economy. Hence, all the parameters that we estimate can be readily used in applied work. [18]

In our choice of functional forms for average tax rates, we are guided by the basic, concave shape of tax rates as a function of income that was evident in our earlier description of tax rates in section 3. Average rates start at near zero, and grow rapidly as income increases. The growth of average tax rates eventually stabilizes, and rates become nearly constant at high levels of income. For instance, for the case of married households, rates are essentially zero around a quarter of mean household income, about 8.3 percent around mean income, and grow to about 17 percent around three times mean income. Subsequently, rates become relatively *flat* as a function of income; for instance, they are about 20.5 percent around five times mean income, about 21.3 percent around seven times mean income, and around 24 percent around ten and fifteen times mean income. All the specifications we present and

[18] Our strategy is to estimate parametrically average (effective) tax rates for different types of households by their filing status and the number of children, capturing in this way the complex nature of the tax legislation in a parsimonious way. An alternative strategy would be to estimate from the data a relation between household income and household taxable income, and use the statutory tax rates to figure out taxes paid for each household. We briefly report in the Online Appendix how income and taxable income are related in the data, and corresponding implications for tax liabilities.

discuss below are consistent with these patterns.

Functional Forms We estimate four specifications for average tax rates. The first two specifications have two parameters while the last two require the estimation of three parameters. In the first case, we posit that

$$t(\tilde{y}) = \alpha + \beta \ \log(\tilde{y}), \tag{1}$$

where t is the average tax rate, and the variable $\tilde{y}$ stands for multiples of mean household income in the data. That is, a value of $\tilde{y}$ equal to 2.0 implies an average tax rate corresponding to an actual level of income that is twice the magnitude of mean household income in the data. This specification was used by Guner et al (2012-a, 2012-b). We refer to it as the *log* specification.

Notice that for this specification, marginal tax rates, m, are given by

$$m(\tilde{y}) = \alpha + \beta \ \log(\tilde{y}) + \beta = t(\tilde{y}) + \beta. \tag{2}$$

That is, marginal tax rates differ from average tax rates by the constant factor β. In macroeconomic terms, this specification is consistent with *balanced growth*: if all incomes increase by a given factor, average and marginal tax rates are unchanged, and total taxes paid increase by the same factor.

Our second and third specification are also consistent with balanced growth. The second one corresponds to the function used in Heathcote, Storesletten and Violante (2011). We refer to it as the HSV specification. It is given by

$$t(\tilde{y}) = 1 - \lambda \tilde{y}^{-\tau}, \tag{3}$$

with corresponding marginal tax rate

$$m(\tilde{y}) = 1 - \lambda(1-\tau)\tilde{y}^{-\tau}. \tag{4}$$

In this specification, the parameter λ controls the level of the tax rate, whereas the parameter τ controls the curvature, or degree progressivity in the tax schedule. If $\tau = 0$, average and marginal tax rates are constant as income changes (flat-rate tax), whereas $\tau > 0$ implies a progressive tax.

Our third specification is given by

$$t(\tilde{y}) = \delta + \gamma\ \tilde{y}^{\varepsilon}, \tag{5}$$

and

$$m(\tilde{y}) = \delta + (1+\varepsilon)\ \gamma\ \tilde{y}^{\varepsilon}. \tag{6}$$

We refer to this specification as the *power* specification. A version of this power function is used by Guvenen et al (2009).

Finally, we also estimate the same functional form used by Gouveia and Strauss (1994):

$$t(y) = b[1 - (sy^{p} + 1)^{-1/p}]. \tag{7}$$

In this case, the variable y stands for the level of household income in the data set. We refer to this as the *GS* specification. The corresponding marginal tax function is

$$m(y) = b[1 - (sy^{p} + 1)^{-1/p-1}] \tag{8}$$

Some comments about the specifications are in order. First, the *HSV* specification is a special case of the *Power* one. The two coincide when $\delta = 1$, $\gamma = -\lambda$ and $\epsilon = -\tau$. Second, The *log*, the *HSV* and the *GS* specification imply that the ratio of marginal to average rates approaches 1 from below. These properties suggest that if average tax rates become relatively flat at high levels of income, then effective marginal tax rates will become close to average rates. Instead, the *power* specification implies that as income grows, the ratio of marginal to average rates approaches $1+\epsilon$.[19] Thus, if estimates for the *power* specification dictate a high value of ϵ, this specification may have problems in reproducing the levels of marginal tax rates at high income levels. We return to these issues later, with a discussion of effective marginal tax rates implied by the different parametric estimates and their raw data counterparts.

Parameter Estimates Tables 9 and 10 show the parameter estimates for all households and for married and unmarried households (with and without children present in the

[19] This requires that ϵ be positive. If instead ϵ is negative, then the ratio converges to one. Our estimates discussed later are always such that $\epsilon > 0$.

household), for all the specifications that we consider. As the tables demonstrate, in all cases parameters are estimated quite precisely.[20]

For illustration purposes, Figure 2 plots the resulting average tax rates under all specifications for the universe of married households, alongside with data averages at each bin.[21] The figure shows that the resulting shape of average tax rates are similar under all cases; all track the shape of average rates at most income levels. The data shows that a married household around mean income (three times mean income) faces an average rate of about 7.7% (16.5%). The *log* specification implies that a married household around mean income (three times mean income) faces an average rate of about 8.5% (14.9%). The corresponding values under the *HSV* specification are 8.7% (14.5%), under the *power* specification are 8.3% (15.1%), and under the *GS* specification are 7.7% (17.0%), respectively. Overall, the fit of all tax functions is very good. Indeed, it is good even at high levels of income.

The role of children and marital status are straightforward; average tax rates tend to be lower for married households, and tend to decrease with the presence of children in the household. This is straightforward to see for the *log* and *HSV* specification. Note that when $\tilde{y}$ equal 1.0, household income equals mean income, and the average tax rate equals α in the *log* case and $1-\lambda$ in the *HSV* case.

Tables 9 and 10 also provide a more formal statistical comparison between different specifications. For each case, we report Akaike information criteria (AIC), originally proposed by Akaike (1973), and widely used to compare non-linear models. AIC basically favors the model with the maximum log-likelihood criteria with an adjustment for simplicity. A smaller value for AIC indicates a better fit.[22] The results in Tables 12 and 13 provide a consistent story. Across all specifications, three-parameter specifications perform better than two-parameter specifications and the *GS* specification performs best. Between two-parameter specifications *log* performs betters than the *HSV* case.[23]

[20] We estimate the *log* and the *HSV* specification using Ordinary Least Squares. We estimate the *Power* and the *GS* formulation using Nonlinear Least Squares.

[21] From the data we calculate average tax rates at 0.2, 0.6, 1.2, 1.6, etc. times the mean household income. The value for 0.2 corresponds to the average tax rate for households in interval of 0 to 0.4 times the mean income, the value for 0.6 corresponds to the average tax rate for households in interval of 0.4 to 0.8 times the mean income etc. The parametric estimates are evaluated at 0.2, 0.6, 1.2, 1.4, etc.

[22] Formally, $AIC = -2\ln L + 2q$, where L is likelihood value and q is the number of parameters.

[23] The basic picture remains the same if we used other commonly used statistics, such as the Bayesian information criteria (BIC), calculated as $BIC = -2\ln L + (\ln N)q$, where N is the number of observations or the squared correlation between data and the predicted values.

Special Cases We present in Table 11 the parameter estimates for a number of special cases, as they can be useful in different research applications. We consider the cases of (i) only households with positive social security income; (ii) households without social security income; (iii) households with only labor income, and (iv) and when state and local income taxes are included alongside federal income taxes. We present results for all households, as well as for all married and unmarried ones. To save space, we focus on one two-parameter specification (the *log* case) and one three-parameter specification (the *GS* case) with better AIC. We present the results for the *HSV* and *power* functions in Table A5 in the Appendix.

Figure 3-a shows tax functions for households with positive social security income as well as those without any social security income for the *log* case. The figure also shows the corresponding tax function in the general case. As the figure shows, considering households with only social security income implies a counter-clockwise shift in the tax function. Households that receive positive transfers from the social security system face lower taxes at lower levels of income. At very low levels of incomes this difference is significant. At 60% of mean household income, for instance, households with positive social security income face average tax rates that are about 7% percentage points *lower* than for the case of all households. This difference declines, however, at higher levels of income: it is about 4.5 percentage points at the mean level of household income and reverses around three times mean household income. Overall, these patterns are not surprising. At low levels of income, social security transfers constitute the bulk of income of these households and social security transfers receive preferential tax treatment; at higher levels of income, retired households, who constitute the bulk of social security recipients, have access to less deductions than more typical households and the contribution of social security transfers to household income declines. Not surprisingly, the picture for the households without any social security income is exactly the opposite. They pay higher (lower) taxes at lower (higher) levels of income.

In Figure 3-b, we show, as an illustration, how the *log* tax function for all households is affected if we include as tax liabilities the sum of federal, state and local taxes. Consistent with Table A2, state and local taxes imply an almost parallel, upward shift of about 5 percentage points after the mean income.

Finally, Figure 3-c shows how the *log* tax function is affected when households have only labor income. The figure displays a flatter tax function than in the benchmark case. For comparison purposes, Figure 3-c plots the function for households who do not have any social security income as this is a natural group to compare with those who only have labor income. Two functions appear similar, especially at low levels of income. Households who have only labor income face on average lower tax rate than all households, possibly reflecting different characteristics of these households in terms of household size, etc.[24]

Comparisons with Previous Estimates It is of interest to compare the estimated tax functions with the existing ones from Gouveia and Strauss (1994), who provided estimates for effective rates using data from 1980 and 1989 for all households. This comparison is displayed in Figure 4, where the corresponding average rates are plotted for these three years.[25]

The figure indicates that there are only minor differences in the resulting average tax functions between 1989 and 2000. Differences occur only at higher income levels and are in the ballpark of one percentage point. The results largely suggest that changes in taxes that took place in 1991 and 1994 did not affect effective average rates significantly. In contrast, as the figure demonstrates, the changes in the tax structure that took place in the 1980's, affected the shape of average rates significantly. For higher income households, the differences are quite substantial; for instance, at five time mean income levels, the differences between 2000 and 1980 is in excess of eleven percentage points.

5.1 Caveats

To close this section, it is important to mention two caveats regarding the estimated tax functions, and their use in applied work. First, as we discussed earlier in section 2.1, an arguably important portion of labor income is not considered in the I.R.S. data nor in other data sets; wage and salary income recorded is net of individual contributions to pension and health plans, and employers contributions to these are not recorded. Researchers using the estimated tax functions should be aware of this fact, and adjust their notion of income or taxable income appropriately if they deem it necessary.

Secondly, our benchmark parametric estimates pertain only to Federal Income Taxes, and for the reasons we explained earlier, they do not take into account state income taxes as well

[24] One caveat with these calculations is that given the nature of the tax data, observations on labor income refer to the labor income of households and *not* of individuals.

[25] For comparison purposes, nominal income has been adjusted, and the estimated parameter s has been adjusted for 1980 in order to make the comparison possible.

as Social Security and Medicare taxes. Of course, these other forms of taxation also distort behavior in economic models. Hence, researchers using our estimates should, if necessary and possible in the context of their analyses, impute these forms of taxation to get a more accurate notion of distorting taxes.

6 Marginal Tax Rates

We now turn our attention to the marginal tax rates emerging from the data and our parametric estimates. We compare the *effective* marginal tax rates implied by our parametric estimates with measures of marginal tax rates from the data. For these purposes, we compute directly effective marginal tax rates from the data. We also compare these effective marginal rates with the statutory ones that households encounter in their tax filing.

There is debate on whether effective or statutory marginal rates are the relevant measures of distortions. Effective rates reflect the inframarginal exemptions, deductions, etc., that reduce average rates. Yet, it can be argued that for many economic decisions the relevant marginal rates are those from the actual tax schedule (statutory rates), as they are the operative ones for decisions on the margin; e.g. to work overtime or not, labor force participation decisions of secondary earners, buying or selling extra units of assets, make charitable contributions, etc. We do not take sides on this debate here. Instead, we report on the effective marginal rates emerging from our data and parametric estimates, and compare them with the actual statutory rates at different levels of income.

Computation of Effective Marginal Rates Our approximation to marginal tax rates in the data is as follows. For a given level of income, say y_0, we compute the variation in tax liabilities when income increases to $y_0+\Delta y$ and when income decreases to $y_0-\Delta y$. Let $m(y_0^+)$ be the marginal tax rate when income increases from y_0, and let $m(y_0^-)$ be marginal tax rate when income decreases from y_0. Let average tax rates be given by $t(y)$ as before. Hence $m(y_0^+)$ and $m(y_0^-)$ are given by

$$m(y_0^+) \equiv \frac{(y_0 + \Delta y)t(y_0 + \Delta y) - y_0 t(y_0)}{\Delta y},$$

and

$$m(y_0^-) \equiv \frac{(y_0 - \Delta y)t(y_0 - \Delta y) - y_0 t(y_0)}{-\Delta y}.$$

Simple algebra implies that

$$m(y_0^+) = [t(y_0 + \Delta y) - t(y_0)] \ \frac{y_0}{\Delta y} + t(y_0 + \Delta y), \tag{9}$$

and

$$m(y_0^-) = [t(y_0) - t(y_0 - \Delta y)] \ \frac{y_0}{\Delta y} + t(y_0 - \Delta y). \tag{10}$$

Not surprisingly, Equations (9) and (10) show that marginal tax rates exceed average rates by a factor that is proportional to the change in average rates as income changes. Given these marginal rates, we calculate our approximation to the marginal tax rate at income level y_0 by averaging out $m(y_0^+)$ and $m(y_0^-)$.

We operationalize the calculation of marginal tax rates as follows. We create a range of income levels, defined as multiples of mean household income, that correspond to the income levels y_0 above. We set the minimum income level to 0.25 times the mean household income and the maximum income to 10 times the mean income and the distance (mesh size) between income levels, Δ, is given by 0.25 times mean income. Given a particular y_0, we calculate $t(y_0)$ as the average tax rate in interval $[1.2y_0, 0.8y_0]$. Given values for y_0, $t(y_0)$ and Δ, we compute the marginal tax rates using equations (9) and (10), and average $m(y_0^+)$ and $m(y_0^-)$ to find the marginal tax rate at income level y_0.

Findings Our findings are summarized in Figures 5 and 6 for the case of all married households. Figure 5 shows the effective marginal tax rates computed as discussed above, alongside the corresponding statutory rates. For any particular income level, the statutory marginal tax rates simply reflect the average value of the statutory marginal tax rates around that income level (similar to Tables 4-6). From the figure, it is clear that statutory marginal rates can exceed effective marginal rates by substantial amounts, and that the gap grows with household income. For instance, at twice mean income, the statutory is above 25%, whereas it is around 22.5% in effective terms. At seven times mean income, the statutory rate is about 35%, whereas the effective marginal rates is 25-26%. Overall, Figure 5 illustrates the fact that as effective average tax rates flatten out at high levels of income, effective marginal rates follow, despite the fact that statutory marginal tax rates increase with income to a highest possible level of 39.6%. We conclude from these findings that the avenues for reducing households' tax liabilities in practice lead to much lower marginal tax rates in effective terms. For households at high levels of income, the gap relative to statutory rates is large and of about ten percentage points.

Figure 6 illustrates that the marginal tax rates emerging from our parametric specifications track the concave-shaped empirical estimates well. A few observations of the Figure 6 are in order. First, the marginal rates emerging from the GS specification become essentially constant at relatively low levels of income (about twice mean income). Second, the gap between the *log*, *HSV* and *Power* specification grows with income. Indeed, the three-parameter *Power* specification leads to the largest marginal rates at high levels of income (in excess of 30% at ten times mean income), and is the one closest to the statutory marginal tax rates.

7 Conclusion

We presented basic facts on the effective taxation of U.S. households in cross-section, distinguishing them by their marital status, the number of dependent children, and other characteristics. We have done so by exploiting the rich cross-sectional data from the U.S. Internal Revenue Service for the year 2000. This allowed us to document the substantial degree of heterogeneity observed in income and taxes paid across U.S. households.

A central contribution of our paper is the estimation of parametric estimates of effective tax *functions* that can be readily used in applied work. We estimated four specifications for different household categories (e.g. married households). All these specifications account for the patterns of average taxes as a function of income quite well.

We conclude the paper by mentioning one caveat in interpreting our results. The caveat is that they pertain to the structure of federal income taxation prevailing in the year 2000. Naturally, the temporary changes that occurred in 2001 an 2003 (*Economic Growth and Tax Relief Reconciliation Act of 2001* and the *Jobs and Growth Tax Relief Reconciliation Act of 2003*) are not captured in our analysis. Nonetheless, we view the snapshot presented of the relationship of taxes and income in cross section as a very good approximation of the nonlinearity (and potential distortions) present in the current system. Indeed, as we write, the tax structure in 2013 under recent amendments is much closer to the 2000 structure than in previous years. For instance, top marginal rates are back to 2000 levels.

References

[1] Akaike, H. (1973). "Information theory as an extension of the maximum likelihood principle." in B. N. Petrov, and F. Csaki, (Eds.) Second International Symposium on Information Theory, pages 267-281 . Akademiai Kiado, Budapest.

[2] Alesina, A., Ichino, A. and L. Karabarbounis, "Gender Based Taxation and the Division of Family Chores," *American Economic Journal: Economic Policy*, 3(2), 2011, pp. 1-40

[3] Altig, D., Auerbach, A., Kotlikoff, L., Smetters, K., and J. Walliser, "Simulating Fundamental Tax Reform in the United States," *American Economic Review*, 91(3), 2001, pp. 574-595.

[4] Apps, P. and R. Rees, "Taxation of Couples," in A Cigno, Pestieau, P. and R. Rees (eds), *Taxation and the Family*, MIT Press (forthcoming), 2010.

[5] Barro, R. T. and C. Sahasakul. "Measuring the Average Marginal Tax Rate from Individual Income Tax," *Journal of Business*, 56(4), 1983, pp. 419-452.

[6] Barro, R. and C. Redlick. "Macroeconomic Effects from Government Purchases and Taxes". Mimeo, 2010.

[7] Black, T., Bloomquist, K., Emblom, E., Johns, A., Plumley, A. and E. Stuk. "Federal Tax Compliance Research: Tax Year 2006 Tax Gap Estimation." IRS Research, Analysis & Statistics Working Paper.

[8] Bick, A. and N. Fuchs-Schundeln. "Taxation and Labor Supply of Married Women across Countries: A Macroeconomic Analysis." Mimeo, 2012.

[9] Boskin, M. J. and E. Sheshinski, "Optimal tax treatment of the family: Married couples," *Journal of Public Economics*, 20(3), 1983, pp. 281-297.

[10] Caucutt, E. , Imrohoroglu, S. and K. B. Kumar, "Growth and Welfare Analysis of Tax Progressivity in a Heterogeneous-Agent Model," *Review of Economic Dynamics*, 6(3), 2003, pp. 546- 577.

[11] Castañeda, A, Díaz-Giménez, J. and J. V. Ríos-Rull, "Accountig for the US Earnings and Wealth Inequality", *Journal of Political Economy*, 111(4), 2003, pp. 818-857.

[12] Chakraborty, I., Holter, H. A. and S. Stepanchuk, “Marriage Stability, Taxation and Aggregate Labor in the US vs. Europe.” Mimeo, 2012.

[13] Conesa, J. C. and D. Krueger, “On the Optimal Progressivity of the Income Tax Code,” *Journal of Monetary Economics*, 53(7), 2006, pp. 1425-1450.

[14] Conesa, J. C., Kitao, S. and D. Krueger. “Taxing Capital? Not a Bad Idea after All!,” *American Economic Review*, 99(1), 2009, pp. 25-48.

[15] Congressional Budget Office. *The Distribution of Household Income and Federal Taxes*, 2008 and 2009. Statistical Supplement. 2012, Washington, DC.

[16] Diaz-Gimenez, J. and J. Pijoan-Mas. “Flat Tax Reforms in the U.S.: A Boon for the Income Poor,” mimeo, 2005.

[17] Díaz-Giménez, J., A. Glover and José-Víctor Ríos-Rull. “Facts on the Distributions of Earnings, Income, and Wealth in the United States: 2007 Update." *Federal Reserve Bank of Minneapolis Quarterly Review*, 34(1), 2011.

[18] Erosa, A. and T. Koreshkova, “Progressive Taxation in a Dynastic Model of Human Capital,” *Journal of Monetary Economics*, 54(3), 2007, pp. 667-685.

[19] Gouvieia, M. and R. Strauss. “Effective Federal Individual Income Tax Functions: An Exploratory Analysis,” *National Tax Journal*, 47(2), 1994, pp. 317-39.

[20] Guner, N., Kaygusuz, R. and G. Ventura. Taxation and Household Labor Supply. *The Review of Economic Studies*, 79(3), 2012, pp. 1113-1149.

[21] Guner, N., Kaygusuz, R. and G. Ventura. "Taxing Women: A Macroeconomic Analysis" *Journal of Monetary Economics*, 59(1), 2012, pp. 129-132.

[22] Guvenen, F., Kuruscu, B. and S. Ozkan, “Taxation of Human Capital and Wage Inequality: A Cross-Country Analysis,” mimeo 2009.

[23] Heathcote, J., Storesletten, J. and G. L. Violante. “Quantitative Macroeconomics with Heterogeneous Households,” *Annual Review of Economics*, 1, 2009, pp. 319-354.

[24] Heathcote, J., Storesletten, J. and G. L. Violante. "Redistributive Taxation in a Partial-Insurance Economy," mimeo, 2011.

[25] Huggett, M. and J. C. Parra. "How Well Does the US Social Insurance System Provide Social Insurance?" *Journal of Political Economy*, 118 (1), 2010, pp. 76-112.

[26] Joines, D. H. "Estimates of Effective Marginal Tax Rates on Factor Incomes," *Journal of Business*, 54(2), 1981, pp. 191-226.

[27] Johns, A. and J. Slemrod, "The Distribution of Income Tax Noncompliance," *National Tax Journal*, 63(3), 2010, pp. 397-418.

[28] Kaygusuz, R. "Taxes and Female Labor Supply," *Review of Economic Dynamics*, 13(4), 2010, pp. 725-741.

[29] Kleven, H., Kreiner, C.T. and E. Saez, "The Optimal Taxation of Couples," *Econometrica*, 2009, 77(2), pp. 537-560.

[30] McCaffery, E. J. *Taxing Women*. The University of Chicago Press: Chicago, IL. 1997.

[31] Mendoza, E. G., A. Razin, L. L. Tesar, "Effective Tax Rates in Macroeconomics: Cross-Country Estimates of Tax Rates on Factor Incomes and Consumption," *Journal of Monetary Economics*, 34(3), 1994, 297-323.

[32] Nishiyama, S. and K. Smetters, "Consumption Taxes and Economic Efficiency with Idiosyncratic Wage Shocks," *Journal of Political Economy*, 113(5), 2005, pp. 1088-1115.

[33] Piketty, T. and E. Saez. "Income Inequality in the United States, 1913-1998", *Quarterly Journal of Economics*, 118(1), 2003, pp. 1-39.

[34] Prescott, E. "Why Do Americans Work So Much More Than Europeans?," *Federal Reserve Bank of Minneapolis Quarterly Review*, 2004, pp. 2-13.

[35] Seater, E. M. "Marginal Federal Personal and Corporate Income Tax Rates in the U.S., 1909-1975," *Journal of Monetary Economics*, 10(3), 1982, 121-135.

[36] Ventura, G., "Flat Tax Reform: A Quantitative Exploration," *Journal of Economic Dynamics & Control*, 23(9-10), 1999, pp. 1425-1458.

Figure 1. Average Tax Rates

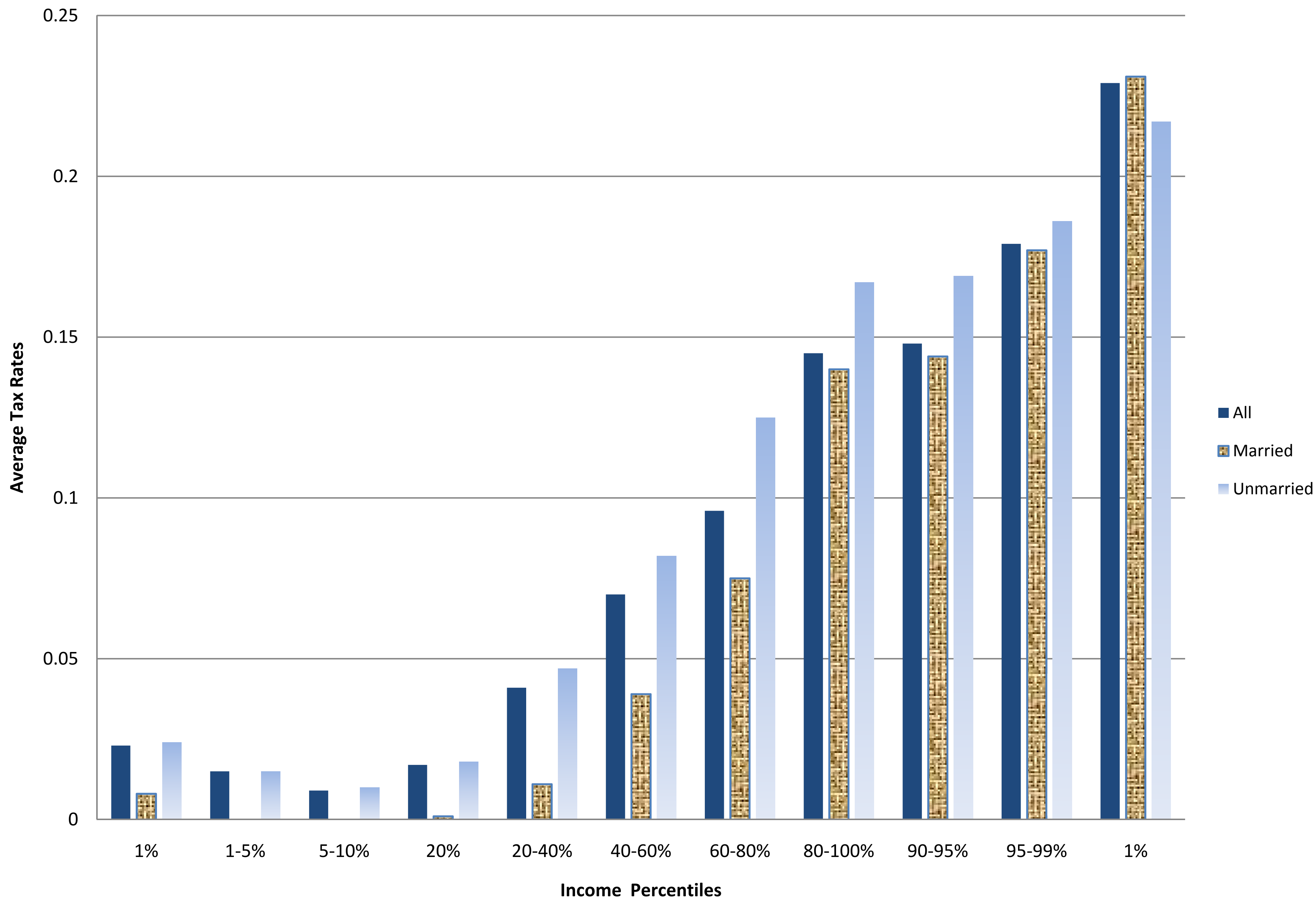


Figure 2. Average Tax Rates for Married Households (data and the parametric estimates)

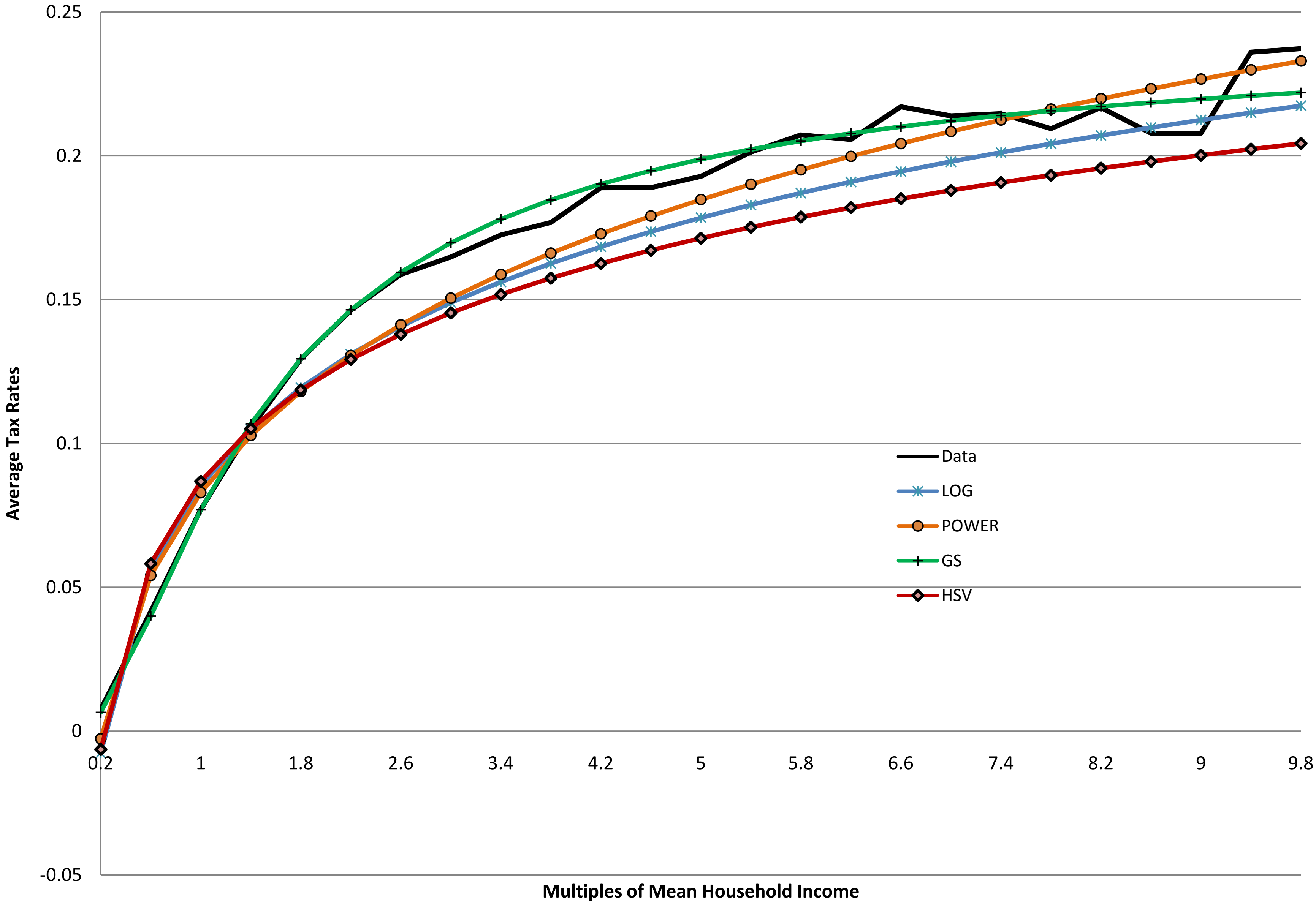


Figure 3a. Log Tax Functions (All Households with and without Social Security Income)

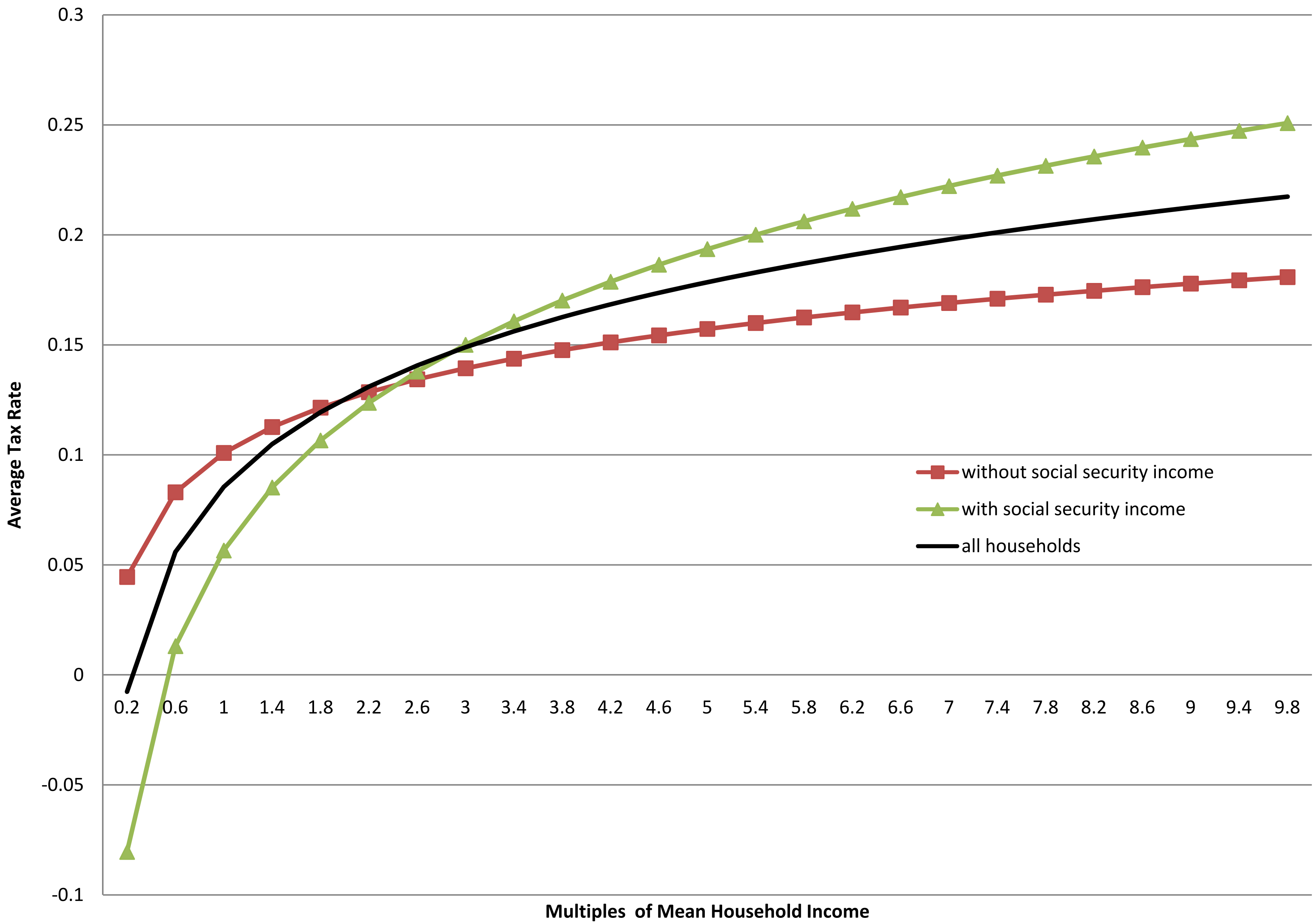


Figure 3b. Log Tax Function (All Households with and without State and Local Taxes)

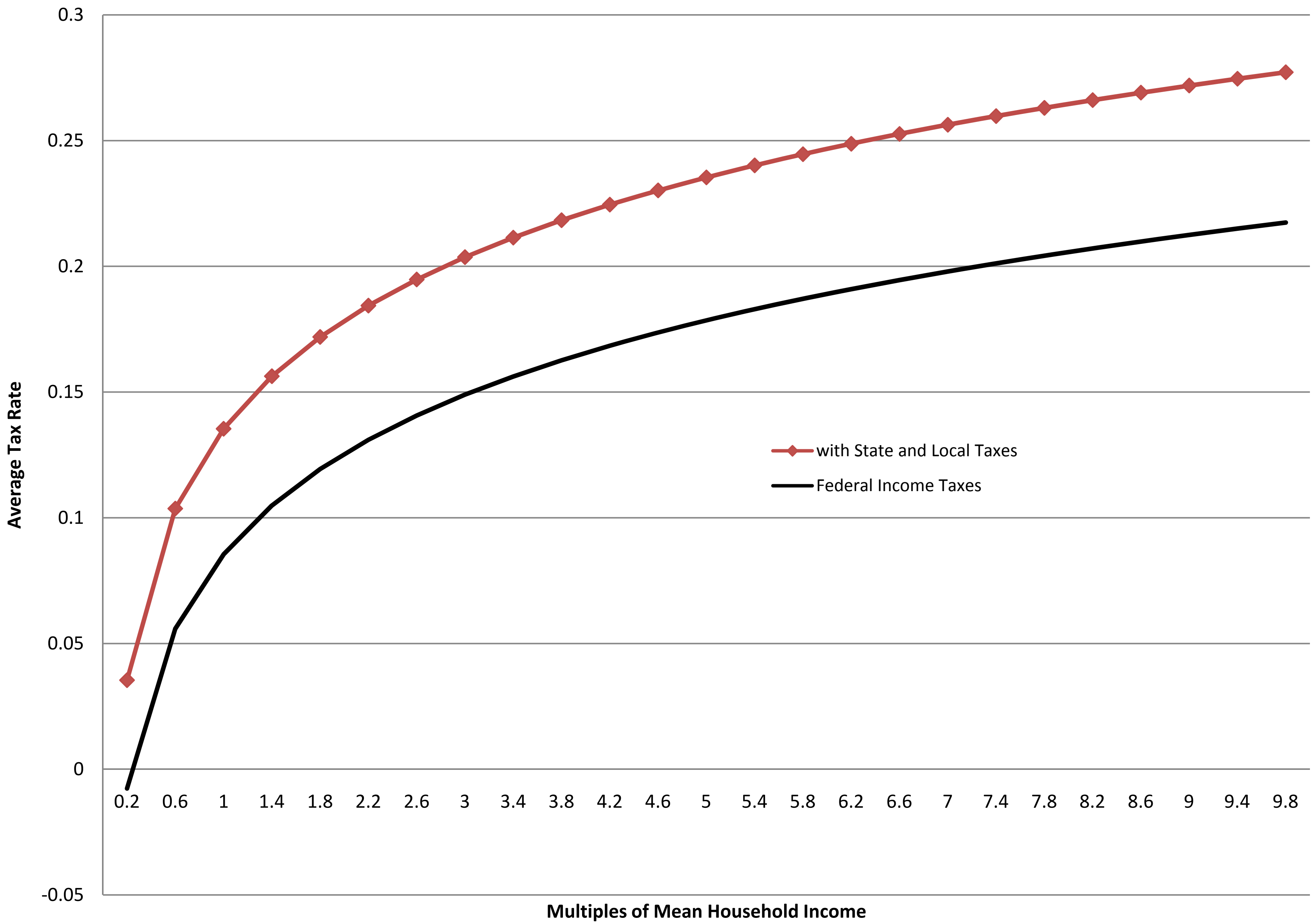


Figure 3c. Log Tax Function (All Households with Labor Income Only)

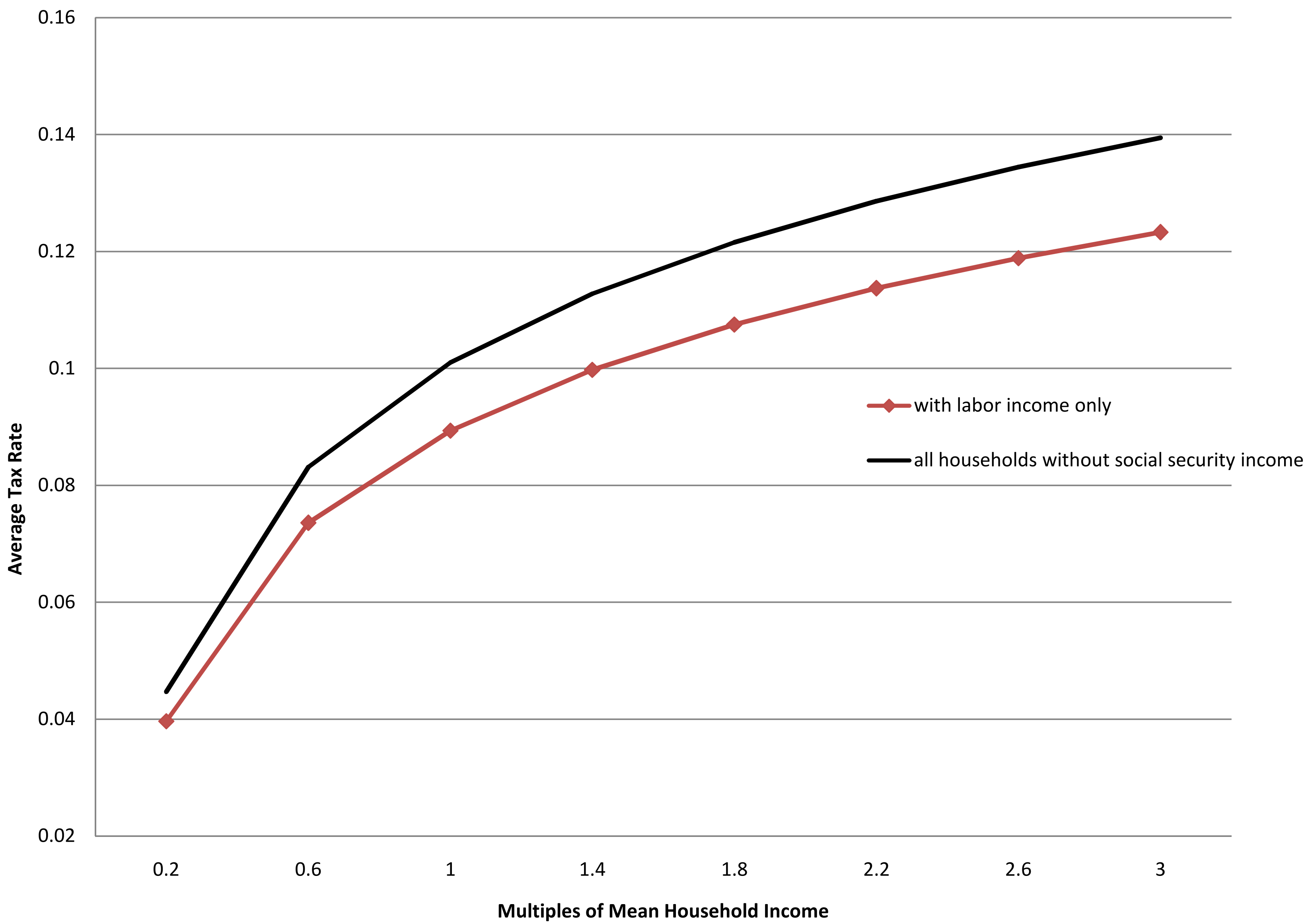
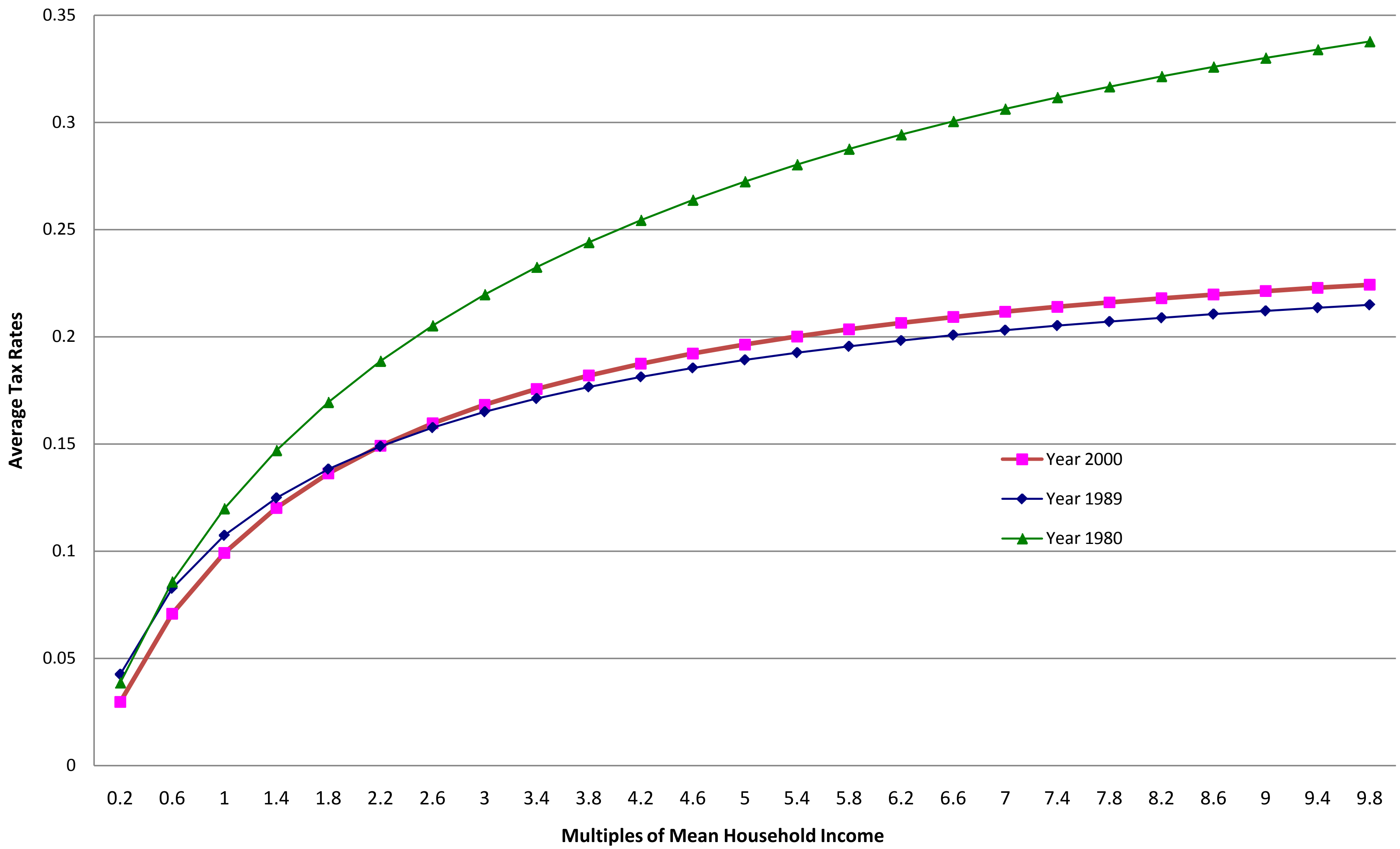


Figure 4. GS Tax Functions (All Households)



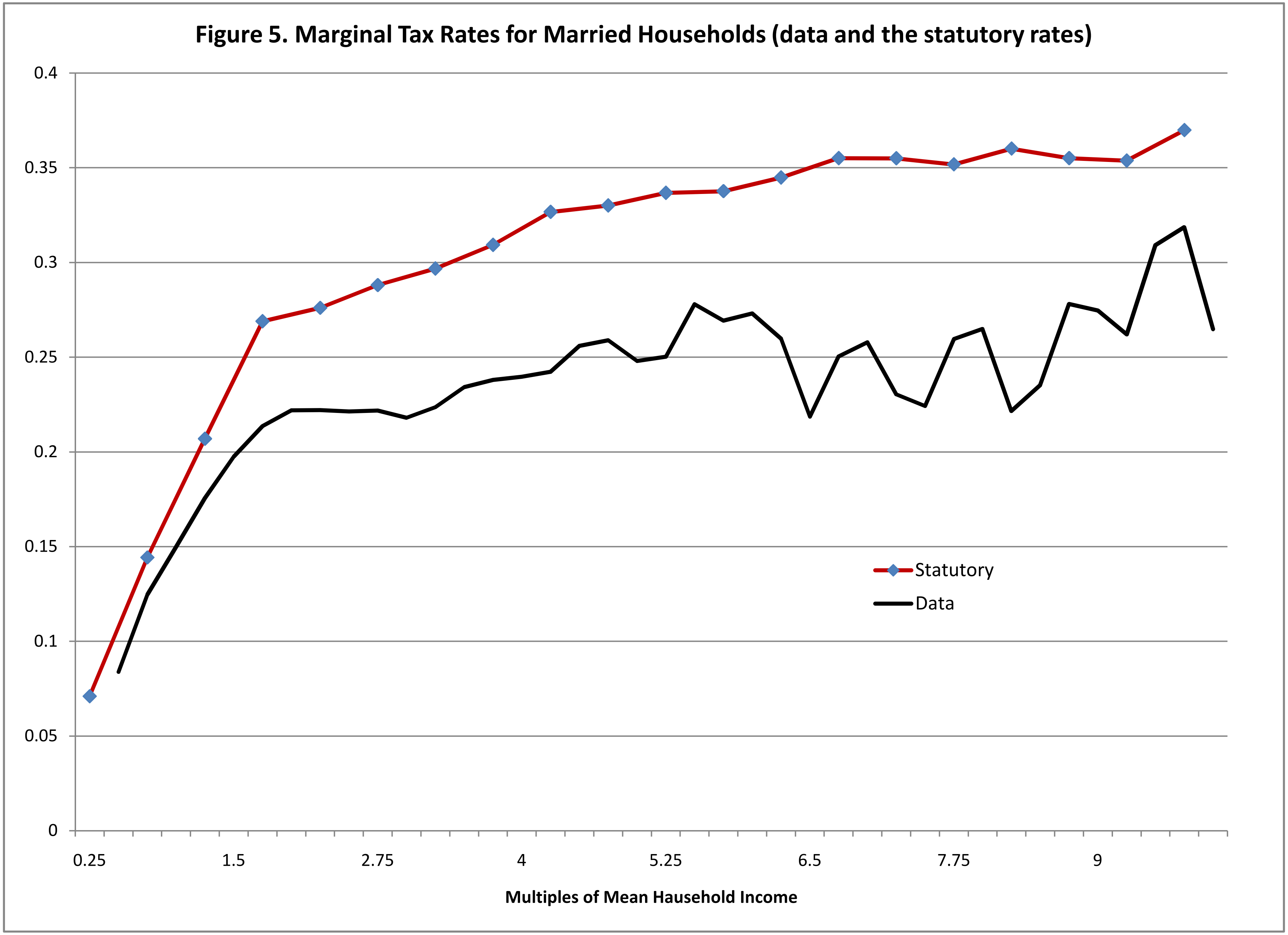


Figure 5. Marginal Tax Rates for Married Households (data and the statutory rates)
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0.25
1.5
2.75
4
5.25
6.5
7.75
9
Multiples of Mean Household Income
Statutory
Data

Figure 6. Marginal Tax Rates for Married Households (data and the parametric estimates)

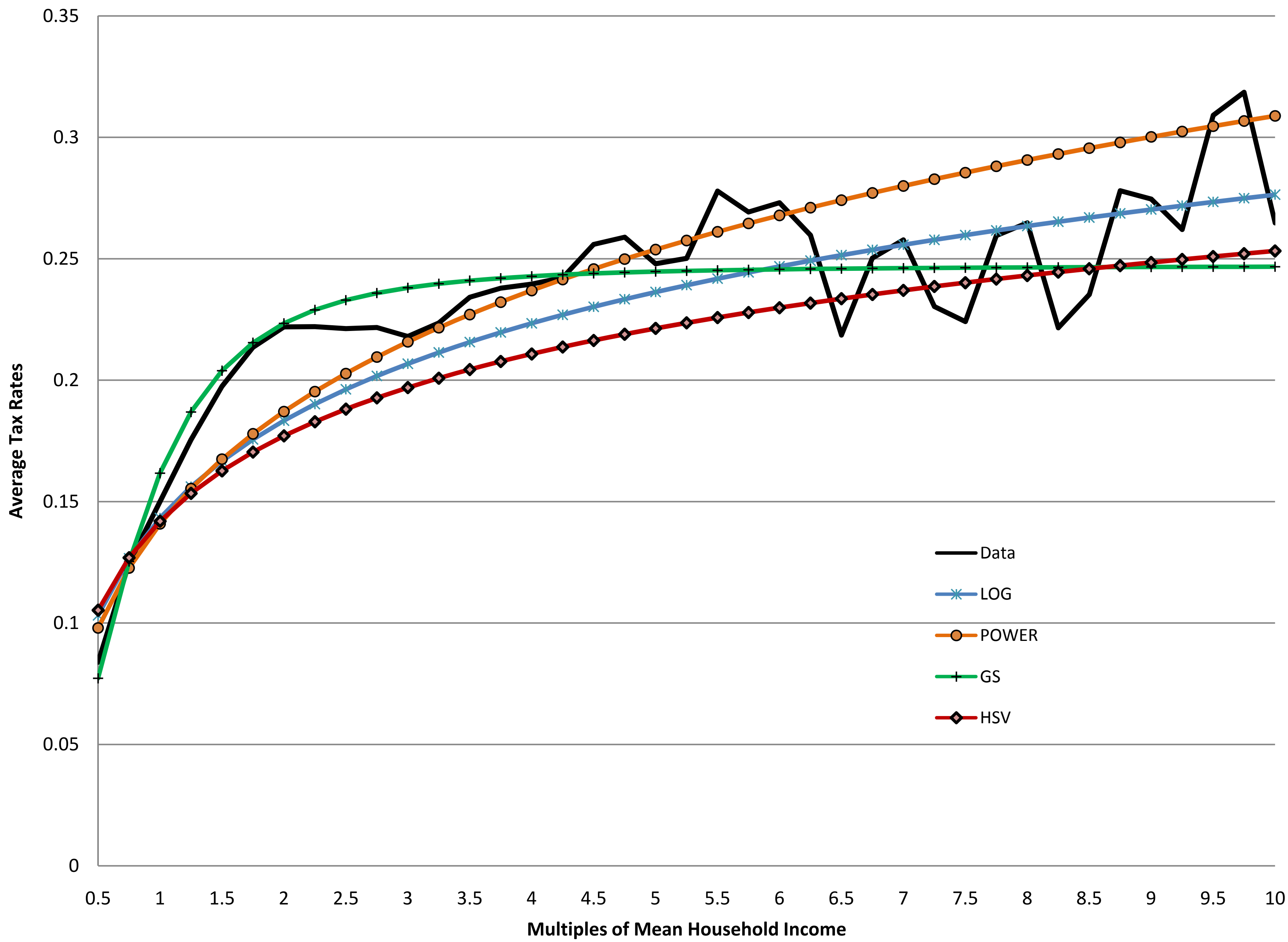


Table 1: 2000 Income Tax Schedule

Marginal Tax Rate	Married Filing Jointly Tax Brackets (Taxable Income)	Single Tax Brackets (Taxable Income)	Head of Household Tax Brackets (Taxable Income)
15.0%	0 - 43,850	0 - 26,250	0 - 35,150
28.0%	43,850 - 105,950	26,250 - 63,550	35,150 - 90,800
31.0%	105,950 - 161,450	63,550 - 132,600	90,800 - 147,050
36.0%	161,450 - 288,350	132,600 - 288,350	147,050 - 288,350
39.6%	over 288,350	over 288,350	over 288,350
Standard Deduction	$7,350	$4,400	$6,450
Personal Exemption	2,800	2,800	2,800

Note: This table displays the income tax schedule in the year 2000 for different filing categories.

Table 2: Income Distribution Statistics

Quantiles	Share of Income	Share of Adjusted Gross Income	Share of Taxable Income
Bottom			
1%	0.0%	0.0%	0.0%
1-5%	0.1%	0.1%	0.1%
5-10%	0.4%	0.4%	0.2%
Quantiles			
1st (bottom 20%)	2.0%	2.1%	1.4%
2nd (20-40%)	6.1%	6.2%	5.1%
3rd (40-60%)	11.3%	11.3%	10.4%
4th (60-80%)	19.1%	19.6%	18.2%
5th (80-100%)	61.3%	60.8%	65.0%
Top			
90-95%	10.6%	10.7%	10.5%
95-99%	15.0%	14.5%	15.4%
1%	20.9%	20.4%	24.4%
Other Statistics			
Gini Coefficient	0.59	0.585	0.63
Var-log Income	1.50	1.46	2.04

Note: This table shows summary statistics for the distribution of income, adjusted gross income and taxable income in the sample.

Table 3: Sources of Income

Quantiles	Labor Income	Capital Income (I)	Transfer Income	Capital Income (II)
Bottom				
1%	88.7%	5.3%	6.1%	14.4 %
1-5%	86.6%	12.6%	0.8%	14.3%
5-10%	89.2%	9.7%	1.1%	12.5%
Quantiles				
1st (bottom 20%)	88.0%	10.2%	1.8%	14.2%
2nd (20-40%)	88.6%	8.2%	3.2%	15.5%
3rd (40-60%)	89.1%	6.1%	4.9%	12.4%
4th (60-80%)	85.4%	8.8%	5.8%	16.6%
5th (80-100%)	81.6%	15.7%	2.7%	24.2%
Top				
90-95%	82.5%	15.1%	2.4%	22.7%
95-99%	74.0%	24.3%	1.7%	34.3%
1%	57.9%	41.4%	0.6%	54.6%

Note: This table shows the contribution of labor, capital and transfer income at different income levels in the sample. Both notions of capital income introduced in the text are presented.

Table 4: Descriptive Tax Statistics

	Income Level										
	Bottom			Quantiles					Top		
Households	1%	1-5%	5-10%	20%	20-40%	40-60%	60-80%	80-100%	90-95%	95-99%	1%
<u>All</u>											
Average Tax Rate	0.023	0.015	0.009	0.017	0.041	0.070	0.096	0.145	0.148	0.179	0.229
(Std. Deviation)	(0.065)	(0.048)	(0.033)	(0.039)	(0.041)	(0.043)	(0.045)	(0.059)	(0.043)	(0.059)	(0.097)
Marginal Tax rate	0.018	0.020	0.032	0.049	0.121	0.155	0.198	0.279	0.277	0.309	0.360
<u>Married</u>											
Average Tax Rate	0.008	-	0.000	0.001	0.011	0.039	0.075	0.140	0.144	0.177	0.231
(Std. Deviation)	(0.055)	-	(0.002)	(0.011)	(0.020)	(0.034)	(0.032)	(0.056)	(0.037)	(0.055)	(0.095)
Marginal Tax rate	0.003	0.001	0.002	0.001	0.076	0.135	0.161	0.277	0.275	0.308	0.363
<u>Unmarried</u>											
Average Tax Rate	0.024	0.015	0.010	0.018	0.047	0.082	0.125	0.167	0.169	0.186	0.217
(Std. Deviation)	(0.066)	(0.049)	(0.033)	(0.040)	(0.041)	(0.040)	(0.043)	(0.063)	(0.064)	(0.076)	(0.104)
Marginal Tax rate	0.019	0.021	0.033	0.052	0.131	0.162	0.243	0.286	0.285	0.312	0.346

<u>Note:</u> This Table shows average tax rates and statutory marginal rates at different income levels. The statutory marginal rate reported is the average of the corresponding statutory marginal tax rate for each taxpayer unit within the income category.

Table 5: Descriptive Tax Statistics: Married Households

	Income Level										
	Bottom			Quantiles					Top		
Households	1%	1-5%	5-10%	20%	20-40%	40-60%	60-80%	80-100%	90-95%	95-99%	1%
No Children											
Average Tax Rate	-	-	0.000	0.000	0.020	0.056	0.087	0.150	0.154	0.176	0.211
(Std. Deviation)	-	-	(0.003)	(0.004)	(0.023)	(0.034)	(0.033)	(0.053)	(0.039)	(0.059)	(0.103)
Marginal Tax rate	-	0.002	0.003	0.002	0.098	0.131	0.166	0.278	0.274	0.303	0.351
Two Children											
Average Tax Rate	-	-	-	0.000	0.000	0.019	0.062	0.132	0.135	0.177	0.255
(Std. Deviation)	-	-	-	(0.004)	(0.023)	(0.034)	(0.033)	(0.053)	(0.031)	(0.051)	(0.078)
Marginal Tax rate	-	-	-	0.001	0.043	0.142	0.154	0.277	0.276	0.312	0.377
Two+ Children											
Average Tax Rate	-	-	-	-	0.001	0.005	0.040	0.117	0.121	0.172	0.254
(Std. Deviation)	-	-	-	-	(0.009)	(0.011)	(0.024)	(0.066)	(0.034)	(0.054)	(0.082)
Marginal Tax rate	-	-	-	-	0.011	0.122	0.149	0.265	0.275	0.311	0.377

Note: This Table shows average tax rates and statutory marginal rates at different income levels for married households. The statutory marginal rate reported is the average of the corresponding statutory marginal tax rate for each household within the income category. Missing values reflect small number of observations in a given cell.

Table 6: Descriptive Tax Statistics: Unmarried Households

	Income Level										
	Bottom			Quantiles					Top		
Households	1%	1-5%	5-10%	20%	20-40%	40-60%	60-80%	80-100%	90-95%	95-99%	1%
No Children											
Average Tax Rate	0.026	0.017	0.011	0.022	0.070	0.098	0.137	0.173	0.171	0.186	0.217
(Std. Deviation)	(0.067)	(0.051)	(0.036)	(0.043)	(0.032)	(0.031)	(0.039)	(0.064)	(0.065)	(0.078)	(0.104)
Marginal Tax rate	0.021	0.023	0.038	0.064	0.142	0.167	0.261	0.289	0.285	0.312	0.346
Two Children											
Average Tax Rate	-	-	-	-	0.000	0.019	0.074	0.135	0.125	0.193	0.247
(Std. Deviation)	-	-	-	-	(0.042)	(0.024)	(0.026)	(0.054)	(0.053)	(0.064)	(0.080)
Marginal Tax rate	-	-	-	0.000	0.088	0.148	0.181	0.271	0.276	0.311	0.353
Two+ Children											
Average Tax Rate	-	0.005	-	0.001	-	0.011	0.053	0.122	0.116	0.174	0.281
(Std. Deviation)	-	(0.027)	-	(0.010)	-	(0.021)	(0.034)	(0.063)	(0.029)	(0.052)	(0.061)
Marginal Tax rate	-	-	-	-	0.042	0.147	0.169	0.277	0.255	0.312	0.388

Note: This Table shows average tax rates and statutory marginal rates at different income levels for unmarried households. The statutory marginal rate reported is the average of the corresponding statutory marginal tax rate for each household within the income category. Missing values reflect small number of observations in a given cell.

Table 7: Distribution of Tax Liabilities

Income Level	Share of Total Taxes Paid
Bottom	
1%	0.0%
1-5%	0.0%
5-10%	0.0%
Quantiles	
1st (bottom 20%)	0.3%
2nd (20-40%)	1.9%
3rd (40-60%)	5.7%
4th (60-80%)	13.1%
5th (80-100%)	79.1%
Top	
90-95%	11.2%
95-99%	19.4%
1%	35.8%

Note: This Table shows the share of total taxes paid at different levels of income in the sample.

Table 8: After-tax Distribution Statistics

Income Level	Before Tax Share of Total Income	After Tax Share of Total Income
Bottom		
1%	0.0%	0.0%
1-5%	0.1%	0.2%
5-10%	0.4%	0.4%
Quantiles		
1st (bottom 20%)	2.0%	2.3%
2nd (20-40%)	6.1%	6.9%
3rd (40-60%)	11.3%	12.2%
4th (60-80%)	19.1%	20.2%
5th (80-100%)	61.3%	58.5%
Top		
90-95%	10.6%	10.6%
95-99%	15.0%	14.3%
1%	20.9%	18.4%
Other Statistics		
Gini Coefficient	0.59	0.56
Var-log Income	1.50	1.39

Note: This Table shows statistics of the distribution of income before and after income taxes in the sample.

Table 9: Parameter Estimates: All and Married Households

Estimates	All	Married (all)	Married No Child.	Married One Child	Married Two Child.	Married Two + Child.
Log						
α	0.099	0.085	0.096	0.089	0.073	0.058
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
β	0.035	0.058	0.054	0.061	0.067	0.060
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
AIC	-458238.82	-341358.66	-168293.14	-56948.98	-80694.80	-47433.30
HSV						
λ	0.902	0.913	0.903	0.910	0.925	0.940
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
τ	0.036	0.060	0.058	0.064	0.070	0.058
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
AIC	-454289.41	-336293.15	-166921.73	-55710.24	-78843.81	-45796.08
Power						
δ	-0.089	-0.451	-0.829	-0.415	-0.495	-0.266
	(0.002)	(0.011)	(0.053)	(0.020)	(0.020)	(0.009)
γ	0.186	0.534	0.923	0.501	0.566	0.320
	(0.002)	(0.011)	(0.053)	(0.020)	(0.020)	(0.009)
ϵ	0.236	0.108	0.059	0.124	0.116	0.186
	(0.002)	(0.002)	(0.003)	(0.005)	(0.004)	(0.005)
AIC	-472252.24	-344987.74	-168759.00	-57955.38	-81968.21	-49976.70
GS						
b	0.264	0.247	0.227	0.251	0.271	0.278
	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
s	0.013	0.001	0.001	0.001	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
p	0.964	1.850	1.842	1.844	2.070	2.602
	(0.006)	(0.014)	(0.023)	(0.029)	(0.025)	(0.042)
AIC	-482426.01	-360639.42	-175942.65	-61049.20	-87249.78	-56009.23

Note: This Table shows the parameter estimates for all households as well as for married households for the three specifications considered. Standard errors are in parenthesis.

Table 10: Parameter Estimates: Unmarried Households

Estimates	Unmarried (all)	Unmarried No Child.	Unmarried One Child	Unmarried Two Child.	Unmarried Two + Child.
Log					
α	0.105	0.121	0.077	0.048	0.037
	(0.000)	(0.000)	(0.001)	(0.001)	(0.002)
β	0.034	0.035	0.042	0.028	0.022
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
AIC	-141800.11	-116720.40	-16514.01	-9944.23	-2909.50
HSV					
λ	0.897	0.882	0.926	0.954	0.965
	(0.000)	(0.000)	(0.001)	(0.001)	(0.002)
τ	0.034	0.036	0.042	0.027	0.021
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
AIC	-140705.17	-115646.18	-16309.34	-9877.80	-2899.00
Power					
δ	-0.068	-0.086	-0.101	-0.056	-0.049
	(0.002)	(0.003)	(0.005)	(0.003)	(0.005)
γ	0.180	0.212	0.183	0.114	0.093
	(0.002)	(0.003)	(0.005)	(0.003)	(0.006)
ϵ	0.296	0.243	0.345	0.468	0.422
	(0.004)	(0.004)	(0.010)	(0.013)	(0.024)
AIC	-147304.40	-120627.61	-17633.08	-10887.58	-3100.75
GS					
b	0.238	0.226	0.170	0.197	0.221
	(0.002)	(0.002)	(0.002)	(0.003)	(0.010)
s	0.008	0.019	0.000	0.000	0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
p	1.366	1.192	9.545	7.318	4.078
	(0.018)	(0.016)	(0.153)	(0.479)	(0.369)
AIC	-152448.50	-124590.48	-20396.16	-12891.46	-3509.09

Note: This Table shows the parameter estimates for unmarried households for the three specifications considered. Standard errors are in parenthesis.

Table 11: Parameter Estimates: Special Cases

	w/SS Income			wo/SS Income			Only Labor Income			w/ State and Local		
Estimates	All	Married	Unmarried	All	Married	Unmarried	All	Married	Unmarried	All	Married	Unmarried
Log												
α	0.085	0.073	0.099	0.101	0.087	0.106	0.089	0.077	0.096	0.135	0.117	0.158
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)
β	0.057	0.065	0.059	0.035	0.058	0.034	0.031	0.047	0.033	0.062	0.075	0.068
	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)
GS												
b	0.196	0.200	0.194	0.276	0.261	0.251	0.182	0.204	0.276	0.317	0.305	0.287
	(0.000)	(0.000)	(0.000)	(0.002)	(0.001)	(0.003)	(0.005)	(0.007)	(0.011)	(0.002)	(0.001)	(0.003)
s	0.000	0.000	0.000	0.013	0.012	0.008	0.008	0.000	0.005	0.016	0.002	0.006
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
p	3.068	6.365	5.389	0.926	1.644	1.336	1.496	2.322	1.444	0.940	1.604	1.514
	(0.096)	(0.376)	(0.0468)	(0.006)	(0.012)	(0.018)	(0.048)	(0.135)	(0.041)	(0.011)	(0.020)	(0.051)

Note: This Table shows the parameter estimates for all, married and unmarried households under two specifications, *log* and *GS*. Four alternative cases are considered: with only social security income, without social security income, with only labor income and including state and local taxes. Standard errors are in parenthesis.

8 Appendix

8.1 Relation between Income and Taxable Income

The IRS micro data allows us to estimate a relation between household income and taxable income. To this end, let y represent household income and y_τ be the taxable income. If we know this relation for married and single households, one can use then statutory tax rates in Table 1 to figure out taxes paid for a household of income y. To this end, suppose the relation between y_τ and y is given by

$$y_\tau = \alpha + \beta y + \gamma y^2 + \theta_1 n + \theta_2 k + \varepsilon,$$

where n is the number of adults and k is the number of children in the household. Since single and married households face different tax schedules, we estimate this relation separately for married and single households. Table A1 shows the results.

Table A1: The Relation between Income and Taxable Income

Estimates	(Married)	(Single)
α	-11597.9***	-2103.9***
	(0.001)	(350.77)
β	0.877***	0.814***
	(0.001)	(0.013)
γ	-0.000**	-0.000***
	(0.000)	(0.000)
θ_1	-2549.8	-2924.2***
	(1168.6)	(344.6)
θ_2	-965.42***	-2223.3***
	(238.03)	(230.8)

***,** and * significant at the 1%, 5% and 10% significance level respectively.

Based on these estimates one can calculate taxable income and taxes paid for any household. Consider a married household with 2 children that earn about mean household income, 53000$. The estimates in Table A1 imply that the taxable income for this household is about 27853$ and according to Table 1, they would pay 15% on this taxable income, about 4178. As a result their effective tax rate is about 7.9%, which is close to tax rates we document in Figure 2. One can do similar calculations for different levels of income and household configurations. Hence, the empirical relationship between household income and taxable income, in conjunction with the tax schedule appears as a sensible alternative to our approach in the paper to estimate tax liabilities.

8.2 Additional Tables

Table A2: Taxes by Sources of Income

	Quintiles by Labor Income				
Quantiles by Capital Income	20%	20-40%	40-60%	60-80%	80-100%
20%	0.005	0.033	0.061	0.091	0.123
	(0.023)	(0.109)	(0.146)	(0.187)	(0.251)
20-40%	0.013	0.045	0.069	0.096	0.121
	(0.042)	(0.125)	(0.147)	(0.194)	(0.255)
40-60%	0.030	0.031	0.067	0.097	0.137
	(0.055)	(0.105)	(0.148)	(0.194)	(0.270)
60-80%	0.070	0.063	0.089	0.114	0.165
	(0.144)	(0.153)	(0.186)	(0.225)	(0.298)
80-100%	0.027	0.039	0.067	0.097	0.139
	(0.061)	(0.117)	(0.152)	(0.197)	(0.271)

Note: This Table shows average tax rates and statutory marginal rates at different quintiles of the joint distribution of capital and labor income. The statutory marginal rate reported is the average of the corresponding statutory marginal tax rate for each household within the income category.

Table A3: State and Local Taxes

Income Level	State and Local Taxes
Bottom	
1%	-
1-5%	-
5-10%	-
Quantiles	
1st (bottom 20%)	-
2nd (20-40%)	4.0
3rd (40-60%)	4.0
4th (60-80%)	4.2
5th (80-100%)	4.6
Top	
90-95%	4.6
95-99%	5.0
1%	5.3

Note: This Table shows the magnitude of state and local income taxes at different income levels.

Table A4: Tax Rate Distribution

Statistic	Married	Unmarried
% with zero taxes	14.5%	31.8%
Median Tax rate	8.5%	6.1%
Mean Tax rate	8.8%	6.4%
Tax Rate Defining		
Bottom 80%	14.0%	11.2%
Bottom 90%	17.1%	14.5%
Bottom 95%	19.8%	17.5%
Bottom 99%	27.7%	23.0%

Note: This Table shows properties of the distribution of average tax rates for married and unmarried households.

Table A5: Parameter Estimates – Special Cases

	w/SS Income			wo/SS Income			Only Labor Income			w/ State and Local		
Estimates	All	Married	Unmarried	All	Married	Unmarried	All	Married	Unmarried	All	Married	Unmarried
HSV												
λ	0.914	0.926	0.901	0.900	0.911	0.896	0.912	0.924	0.907	0.863	0.884	0.841
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)
τ	0.062	0.075	0.064	0.036	0.060	0.034	0.031	0.047	0.033	0.074	0.092	0.080
	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)
Power												
δ	-8.995	-9.327	-8.665	-0.084	-0.378	-0.063	-0.208	-0.044	-0.019	-9.025	-10.025	-8.989
	(0.000)	(0.000)	(0.001)	(0.002)	(0.010)	(0.002)	(0.002)	(0.004)	(0.001)	(0.000)	(0.000)	(0.000)
γ	9.080	9.400	8.763	0.183	0.462	0.178	0.133	0.121	0.151	9.160	10.141	9.136
	(0.000)	(0.000)	(0.000)	(0.002)	(0.009)	(0.002)	(0.002)	(0.005)	(0.002)	(0.000)	(0.000)	(0.000)
ϵ	0.006	0.007	0.007	0.246	0.126	0.314	0.508	0.603	0.680	0.007	0.007	0.007
	(0.000)	(0.000)	(0.000)	(0.002)	(0.002)	(0.004)	(0.014)	(0.027)	(0.014)	(0.000)	(0.000)	(0.001)

Note: This Table shows the parameter estimates for all, married and unmarried households under two specifications, *HSV* and *Power*. Four alternative cases are considered: with only social security income, without social security income, with only labor income and including state and local taxes. Standard errors are in parenthesis.