An Introduction to the TAXSIM Model

Daniel Feenberg
Elisabeth Coutts

The TAXSIM model is now 17 years old. The original tax calculator was written by Amy Taylor to estimate the effects of tax deductibility on charitable giving [Feldstein and Taylor, 1976]. The model was given its present form several years later by another Harvard student, Daniel Frisch, for a study of the incidence of a proposed integration of the corporate and personal income tax systems [Feldstein and Frisch, 1977]. Subsequent studies by Martin Feldstein, his students, and others proved the data so rich and the model so useful that TAXSIM has been updated every year since and used in scores of NBER projects. This note provides a short history and description of TAXSIM, with special emphasis on the state tax calculator used by six of the papers in this issue of the *JPAM*.

WHAT IS TAXSIM?

"TAXSIM" refers to a collection of programs and data sets implementing a microsimulation model of the U.S. federal and state income tax systems. It divides naturally into three components.

The first component is a database of real tax returns. We are fortunate to have, from each year since 1973, a large sample of actual tax returns (at least 80,000 records) prepared for public use by the Statistics of Income Division of the Internal Revenue Service. This file, dubbed the "Tax Model" by the IRS (although it contains only data), includes almost 200 variables for each taxpayer. Virtually everything on Form 1040 is available, together with several items from each of most supporting schedules. However, certain easily traceable data items—such as alimony and real estate taxes—have been partially obscured to protect taxpayer identity [Strudler et al., 1987]. This is a stratified random sample, with high sampling rates for wealthy taxpayers and small states. The sampling rate can be as high as one in three taxpayers with AGI of $200,000 or greater, while only one in 10,000 low income taxpayers may be included. The data are as poor in demographic information (age, race, sex, and hours worked are not available) as they are rich in income information.

The data are sufficient to calculate federal tax liabilities to a high degree of accuracy and state liabilities to a somewhat lower standard. Indeed, the second part of the model is the tax calculator. It takes the raw data on incomes and deductions and calculates the tax liabilities. The tax calculator is simply the recreation of each year's tax law in FORTRAN and does, in effect, what H&R Block does. Although the actual amount of federal income tax owed is available from the file, we wish to calculate revenues from alternate tax regimes and marginal tax rates, which necessitates being able to compute liabilities ourselves. An expression for the total marginal tax rate where state taxes may be deductible on the federal return (and vice versa) is given in Feenberg [1987].

The final part of TAXSIM is the table generator, which produces tables of population-weighted aggregate statistics by any specified tabulation variable, usually some measure of income.

USES OF MICROSIMULATION MODELS

Like similar models used by the Treasury, the Joint Committee on Taxation, and Brookings Institution, TAXSIM can be used to prepare estimates of the static revenue effect—in total and by income class—of proposed changes in the tax law. Provided that additional data are not required, the procedure is as simple as calculating tax liabilities under the old and proposed regime, and tabulating the result. The microsimulation model overcomes the difficulty of estimating revenue effects from aggregate data when the tax law is nonlinear and depends in a complex way on a large number of variables with strong correlations. Since these joint distributions are typically unknown, a large sample of returns provides an empirical substitute for analytical knowledge. Some imputation of omitted variables is almost certain to be required, however. For example, in the founding study the distribution of corporate ownership was imputed from dividend data.

A more common use of TAXSIM, in contrast to the typical use of models maintained at other institutions, has been the study of the behavioral effects of taxation. Examples of these include studies of charitable giving [Feldstein and Taylor, 1976; Feenberg, 1987], interregional labor supply [Gyourko and Tracy, 1989], capital gains realizations [Feldstein, Yitzahki, and Slemrod, 1980], and the sheltering of income from taxation [Lindsey, 1990]. These studies emphasize price rather than income effects as the object of study, and therefore fall into what has been called the "New Public Finance" [Boskin and Stiglitz, 1977]. More parochially, these studies can be characterized as "tax-price regressions" because each attempts to estimate the behavioral response to a proposed change in the tax law from a cross-section data set in which differences in tax rates across individuals provide the only experimental variation in prices.

The first successful tax-price regression was Feldstein and Taylor's regression of charitable contributions on personal characteristics and the after-tax marginal price of charitable giving. The crucial contribution made then was the use of a "first dollar" tax rate (the tax rate an individual would have paid on his or her first dollar of charitable giving) to overcome the endogeneity of the observed marginal tax rate. This endogeneity stems from the fact that sufficiently large gifts can reduce the taxable income enough to move the taxpayer into a lower marginal bracket rate.

A characteristic limitation of any cross-section regression is the lack of any source of true experimental variation. As long as the variation in tax is a consequence of variation in personal characteristics, it may be difficult to distinguish price effects from the direct effects of the personal characteristics themselves. A fuller discussion of this argument is presented in Feenberg [1987], together with a possible solution based on cross-state variation in tax rates in an instrumental variables framework.

A third strain of studies has emphasized the effect of tax laws on state decisionmaking, including the response of state tax systems to changes in federal law [Feldstein and Metcalf, 1987], or inflation and the effect of state tax structures on public sector growth [Feenberg and Rosen, 1987]. These studies use the state as the unit of econometric analysis, but take data on state average marginal tax rates and other variables from individual data in TAXSIM. Because the net price of state and local taxes depends on federal deductibility, which in turn depends on the probability of itemizing deductions, individual microdata are required to generate the exogenous measures of state and local tax prices used in the Feldstein and Metcalf study.

The emphasis on estimating behavioral relations explains the general neglect in TAXSIM of statistical merges and imputations of data not available on the Tax Model file. Whatever the value of imputed data in the day-to-day business of making revenue estimates, it adds nothing to a tax-price regression. (The estimated coefficient on an imputed variable would not be informative.) Nor are elaborate mechanisms to age data to match known or projected aggregates necessary in this activity.

THE STATE TAX CALCULATOR

The first TAXSIM package did not include any estimates of state tax liabilities, as residence information was not then available on the public use file. Starting with the 1974 tax year, state of residence (actually, the state field from the taxpayer's return address) has been available for most returns. Starting in 1981, we made an effort to calculate state tax liabilities, and we currently have calculators for every state for the years 1977 to 1988. A full description of the tax calculator and a comparison of the state tax systems based on it is provided in Feenberg and Rosen [1986], from which this section is adapted.

Although state tax regimes differ significantly from one another, most share the basic structure of the federal tax. That is, deductions and exemptions are subtracted from adjusted gross income to obtain taxable income. A schedule converts taxable income to income before credits, from which a variety of credits, some refundable, are subtracted.

Even so, the state taxes are not clones of the federal tax. As of 1989, only three states use federal taxable income or federal tax as their tax base. Eleven states allowed a deduction for federal taxes paid (several limit the deduction), while all but four disallow the federal deduction for state income taxes paid. Eighteen have separate schedules for couples and individuals, but 35 allow income splitting or separate filing. Child-care credits, rent credits, property tax credits, general credits, various elderly and pension credits, minimum and maximum taxes, percentage standard deductions, and many other features have each found expression in one or more states. The most ubiquitous provisions in state laws that have no correspondence to federal law are the

property-tax credits included in 20 states and the rent credits and deductions found in 20 (mostly overlapping) states.

We code the tax laws using information obtained chiefly from the tax forms distributed by the states to their residents, and secondarily from summaries published by the Commerce Clearing House, the Advisory Commission on Intergovernmental Relations (ACIR), and the Tax Foundation. The tax forms are especially useful because they present a snapshot of the actual law in effect for a particular year. Secondary sources are rarely sufficiently detailed, often omitting such basic information as bracket rates for a filing status, and are often forward looking, presenting tax laws scheduled to go into effect in the future, but which may be withdrawn [Feenberg and Rosen, 1988].

METHODS AND LIMITATIONS

We have attempted to code every aspect of the tax system that our data would allow. However, data limitations have forced us to impute several variables that have an impact on state tax liabilities: (1) Federal tax returns provide no data on household rent payments, but rent credits are an important component of state tax systems. We have assumed that families with few or no property tax deductions and modest income were renters, and estimated their rent based on consumer expenditure data. (2) Social Security benefits are not reported until 1984, but rent and low-income credits often depend upon benefits. In earlier years households with age exemptions are imputed a benefit from data reported in the Survey of Consumer Finance. (3) In some states separate filing is often advantageous, even for couples filing a joint federal return. Since federal returns do not list husbands' and wive's income separately, we have not allowed this election. In one state requiring separate returns, we have presumed a one-third to two-thirds split within the household. (4) Taxpayers with no state identification (because their income exceeded $200,000) were assigned randomly to states, such that the number of those returns by state matched figures provided by the Joint Committee on Taxation.

For some other missing variables we could not arrive at a satisfactory imputation scheme. Certain aspects of state tax systems are therefore ignored. The most important of these are as follows:

1. *Tax exempt interest*. Because federal tax returns did not include interest from municipal debt until 1987, and still do not distinguish between taxable (at the state level) out-of-state debt and nontaxable (in some states) in-state debt, we do not compute any state tax liability generated by such interest.
2. *Interest from federal securities*. This is not taxable at the state level, but is included in our data as interest income.
3. *Property tax credits for nonitemizers*. For nonitemizers we have no estimate of property tax liability, and such taxpayers are treated as renters for the purpose of calculating rent and property tax credits.
4. *Itemized deductions for nonitemizers*. Many states have smaller standard deductions or more generous treatment of personal itemized deductions. These data are not available for taxpayers not itemizing on the federal return, and are uniformly treated as zeros by the state tax calculators.

5. *Nontaxable household income.* Some states predicate clawbacks of low-income credits on a concept of household income more inclusive than AGI. Aside from the imputation of Social Security benefits, we use taxpayer gross income for household income.

Several of the authors have used simulated data. By using the same sample of taxpayers across states, a comparison of tax laws is not compounded by (possibly endogenous) cross-state differences in the income distribution. Nevertheless, the above limitations are still present.

The significance of these omissions is difficult to judge. In this issue of the *Journal*, Helen Ladd has criticized the model for its failure to reproduce aggregate revenue figures by state. In addition to the missing data noted above, the following considerations reduce our ability to reproduce published revenue totals:

1. *Sampling error.* At least 10 states have estimated sampling errors of 10 percent or greater for aggregate 1985 revenues.
2. *Apples and oranges.* Our estimates are for calendar year liabilities; states uniformly publish fiscal year receipts data.
3. *Nonuniform treatment of credits.* In our estimates, credits are always negative tax revenue. States often treat one or more credits as positive expenditure. These considerations do not affect the tax-price regressions, although they are relevant for some of the studies in this volume.

One may also speculate on the source of the differences Ladd has noted between the ACIR estimates of the state revenue windfall from the Tax Reform Act of 1986 and those derived from TAXSIM. The crucial difficulties here may include the following:

1. *Active and passive income.* The distinction is not available in the data for tax years prior to 1987.
2. *Depreciation and other business deductions.* Changes in depreciation allowances are not modeled; business income is available only net of depreciation.
3. *Windfall concept.* TAXSIM does not know the statutory basis for income and deduction definitions; that is, a line on a state tax form may be based on a definition in the federal law as of a certain date, or may change in step with changes in the federal law. Even in the former case, custom may dictate that state law be updated to conform to current federal law, and these changes may fairly be characterized as windfalls, if widely accepted.

CONCLUSION

TAXSIM currently offers an elaborate set of databases and tax calculators for a variety of microsimulation and econometric uses. We expect to see further use in both the evaluation of the likely response of taxpayers to changes in the law at both the state and federal law, and to revenue estimates incorporating this behavior.

194 / *Taxsim*

DANIEL FEENBERG is Research Associate and ELISABETH COUTTS is Research Economist at the National Bureau of Economic Research offices in Cambridge and Palo Alto, respectively.

REFERENCES

Boskin, Michael J., and Joseph E. Stiglitz (1977), "Some Lessons from the New Public Finance," *The American Economic Review, Papers and Proceedings* 67(1) (February), pp. 295–301.

Feenberg, Daniel R. (1987), "Are Tax Price Models Really Identified? The Case of Charitable Giving," *National Tax Journal* 40(4) (December), pp. 629–633.

Feenberg, Daniel, and Harvey Rosen (1986), "State Personal Income and Sales Taxes: 1977–1983," *Studies in State and Local Public Finance* (Chicago: University of Chicago Press), pp. 135–186.

Feenberg, Daniel, and Harvey Rosen (1987), "Tax Structure and Public Sector Growth," *Journal of Public Economics* 32(1) (February), pp. 185–201.

Feenberg, Daniel, and Harvey Rosen (1988), "Promises, Promises: The States' Experience with Income Tax Indexing," *National Tax Journal* 41(4) (December), pp. 525–542.

Feldstein, Martin, and Daniel Frisch (1977), "Corporate Tax Integration: The Estimated Effects on Capital Accumulation and Tax Distribution of Two Integration Proposals," *National Tax Journal* 30(1).

Feldstein, Martin, and Gilbert Metcalf (1987), "The Effect of Federal Tax Deductibility on State and Local Taxes and Spending," *Journal of Political Economy* 95(4) (August), pp. 710–736.

Feldstein, Martin, Joel Slemrod, and Shlomo Yitzhaki (1980), "The Effects of Taxation on the Selling of Corporate Stock and the Realization of Capital Gains," *The Quarterly Journal of Economics* 114(4) (June), pp. 777–791.

Feldstein, Martin, and Amy Taylor (1976), "The Income Tax and Charitable Contributions," *Econometrica* 44(6) (November).

Gyourko Joseph, and Joseph S. Tracy (1989), "On the Political Economy of Land Value Capitalization and Local Public Sector Rent-Seeking in a Tiebout Model," *Journal of Urban Economics* 26, pp. 152–173.

Lindsey, Lawrence B. (1988), "Federal Deductibility of State and Local Taxes: A Test of Public Choice by Representative Government," in Harvey S. Rosen (ed.), *Fiscal Federalism: Quantitative Studies* (Chicago: University of Chicago Press).

Lindsey, Lawrence B. (1990), *The Growth Experiment: How the New Tax Policy Is Transforming the U.S. Economy* (New York: Basic Books).

Studler, Michael, H. Lock Oh, and Fritz Scheuren (1987), "Protection of Taxpayer Confidentiality on the IRS Tax Model," *Statistics of Income and Related Administrative Record Research* (Washington, DC: Internal Revenue Service).