2 THE INDIVIDUAL INCOME TAX SIMULATION MODEL

James M. Cilke and Roy A. Wyscarver

I. INTRODUCTION

The staff of the Office of Tax Analysis (OTA) has the dual responsibility of projecting estimates of Federal income tax receipts and of analyzing and estimating the revenue, distributional, and economic impact of enacted and proposed legislation. One of the most useful tools employed by OTA to estimate and evaluate individual income taxes is the individual income tax simulation model.

Over the last four years, the model was used to estimate the effects of thousands of proposed changes to the tax code that arose during the course of the tax reform initiative. The tax model was used to analyze and estimate:

- Effects of a single tax provision or set of provisions,
- Separate effects of the components of a set of multiple provisions,
- Interaction among provisions,
- Sensitivity of the tax system to certain tax parameters,
- Average marginal tax rates,
- Effects of behavioral responses in certain instances,
- Distributional effects of the proposed tax system, and
- Effects of a proposal in terms of who gains and who loses.

Moreover, the model provided these analyses on a tax return or a family basis; distributed by adjusted gross income, economic income, or some other income concept; and including or excluding the non-filing population.

Last, but not least, the model allowed the analysis and estimation of proposals for fundamental tax reform; proposals that required information which was not tied to a particular tax law or limited to what was reported on tax returns filed under that tax regime.

The authors express their appreciation for the valuable comments and suggestions from Don Fullerton, Susan Nelson, Tom Neubig and Jim Nunns. The authors also thank Eunice Taylor for assistance in the preparation of the manuscript.

The individual income tax model consists of two major components: (a) a probability sample of individual income tax returns merged with other demographic and economic data and (b) a computer program that manipulates the sample, according to a set of input parameters that define the tax laws, for the analysis of enacted and proposed legislation. The tax model is a microsimulation model, and works by recalculating the Federal individual tax liability for each return in the sample. The effects of changes in tax laws can be simulated by changing the set of parameters or computer coding that define the tax laws.

The construction of the data component of the tax model is described in Section II of this paper. The data base was constructed by merging tax return data from the Statistics of Income (SOI) file with other demographic and economic data from the Current Population Survey (CPS) file, imputing additional information not available on either file, and extrapolating the data through a five year budget period.

The other component of the tax model, the computer program, is described in Section III. The computer program consists of three sets of instructions. The first set of instructions reads input parameters that contain changes to (a) tax parameters affecting capital gains, exemptions, credits, etc.; (b) parameters affecting the definition of itemized deductions; and (c) the tax rate schedules. These parameters collectively define the tax laws that are applied in the second set of instructions. The second set of instructions sequentially reads income tax returns from the data base, recalculates tax liability under the tax laws, and computes the change between the base law and the proposed law. Updates to aggregate tabulations are performed on either a return or family basis and retained for the final set of instructions. The last set of instructions prints the results of the simulation in a tabular format. A more detailed description of Sections II and III can be found in the individual tax model documentation.[1]

Section IV describes how the tax model was used to answer various kinds of tax policy questions during tax reform. This section also summarizes some of the quantitative results and their qualitative significance. Presented here are estimates of average marginal tax rates weighted by selected income sources under 1980 law, 1986 law, and tax reform. Also presented here is a table showing the distribution of gainers and losers from the Tax Reform Act of 1986.

Finally, Section V outlines and discusses additional developments that are needed to further improve the accuracy and usefulness of the tax model. Of particular importance is the need to more reliably extrapolate the data file to future years.

II. DESCRIPTION OF THE TAX REFORM DATA BASE

The increasing sophistication of tax proposals and analysis has gradually outdated the use of aggregate data. Distributional effects of tax changes are important in the assessment of the probable and overall impact of any proposal and have always been a factor in assessing individual income tax proposals. These distributional effects can best be estimated through the use of a microdata file. In addition, Federal individual income taxes are so complex that estimates of many interacting variables (necessary for the assessment of the effects of proposed changes in the tax code) cannot be handled in the aggregate. To capture these effects in our analyses, OTA develops and maintains a microdata file representative of the filing and non-filing U.S. population.

The basic data in the tax model's data base comes from information reported on individual tax returns. However, this information proved to be inadequate for analyzing many of the fundamental tax reform proposals offered in the early 1980's. Therefore, OTA expanded the tax return based data file by adding information such as non-taxed income sources, consumption, and wealth. The tax reform data base was constructed through four processes.

° Preparation of tax return data from the Statistics of Income (SOI) and demographic and economic data from the Current Population Survey (CPS) file,
° Merging the SOI and CPS files,
° Imputing additional information not available on either file, and
° Extrapolating the data through a five year budget period.

In the first process, considerable preparatory work was required to ensure that the SOI and CPS files were consistent with each other and representative of the U.S. population. This process is outlined for each file in Subsection A.

The second process, merging the SOI and CPS data files, is the topic of Subsection B. The merge process expands the data base to include forms of income that are not presently taxed, primarily, transfer payments; to incorporate non-filers, families and single individuals who do not file tax returns (largely because their incomes are too low); to combine taxpayers into families; and to unite dependent filers with their corresponding family.

Imputing additional information is covered in Subsection C. This additional information, when combined with the SOI and CPS data, allows for the construction of a measure of "economic income."

Subsection D covers the final process, the extrapolation of the data through a five year budget period. This extrapolation was necessary to provide to policymakers estimates of the revenue effects of tax law changes over the five year period immediately following enactment.

A. Preparation for the Merge

1. SOI Preparation

The starting point for developing the tax reform data base was a stratified random sample of approximately 144,000 tax returns filed for calendar year

1981. This was the same sample employed by the Internal Revenue Service (IRS) to produce the tabulations published in the Statistics of Income – 1981 Individual Income Tax Returns.[2] The data on each tax return in the sample include the major entries on Form 1040, 1040A, and the auxiliary schedules. When weighted, these data represent the U.S. aggregate population of taxpayers.

a. Appending Social Security Data

The first enhancement made to the data base was the appending of Social Security Administration (SSA) data on age, sex, and social security earnings of the taxpayer(s). This was accomplished by providing SSA with the social security numbers (SSNs) of the taxpayers in the SOI sample. SSA used the SSNs to extract the requested data from its master file. OTA then merged the extracted data with the SOI sample.

b. Reducing the Sample Size

The use of large samples for microsimulation is computationally expensive. One way to limit this expense is to reduce the size of the sample. This approach, however, may result in a reduced sample that no longer accurately represents the population. Therefore, the objective is to reduce the size of the sample while minimizing the distortion this reduction introduces.

The usual methods for reducing a data file are variants of stratified and clustered sampling. Yet sampling theory does not provide error estimates for many of the distributional questions that are of concern in formulating tax policy. OTA used a deterministic optimization model that endeavored to maintain the characteristics of the enhanced 1981 SOI as closely as possible.

The optimization model was an optimal cluster analysis procedure that employed Lagrangian relaxation, an efficient variation of the K-median algorithm and dynamic programming search algorithms.[3] The objective of this procedure was to select "the" subset of points (75,422 observations) out of the original sample of 144,322 observations so as to minimize the total loss of information as measured by an explicit objective function.

Since this technique requires the computation of an information loss for every possible pairing of observations in the original sample, solving the total problem would involve 20.7 billion variables ($144{,}000^2$); an impossibility on conventional computers.

An obvious solution would be to decompose the total problem into relatively independent subproblems that could be solved separately. However, none of the many existing methods for decomposing a large scale optimization problem was applicable to the SOI reduction because of its enormous size. Therefore, a "heuristic" was employed for subdividing the aggregate problem into manageable pieces. The heuristic was similar to the formulation of strata boundaries for stratified sampling.

Table 2.1 **Essential Attributes Preserved In Reducing the 1981 SOI File**

Adjusted gross income or loss
Taxes paid Wages and salaries
Dividend income
Interest income
Business income or loss
Farm income or loss
Capital gains or losses
Total tax credits
Pension and annuity income
Itemized deductions
Total tax preference income

For simplicity, twenty subproblems were generated according to the strata boundaries defined by IRS, i.e., the presence of business/non-business activity stratified by adjusted gross income or largest single income item. One subproblem was constructed solely for returns with farm income or loss since this attribute was very sparse in the other subproblems. The two remaining subproblems consisted of individual taxpayers who filed singly and everyone else. Thus, the aggregate problem was divided into twenty-three subproblems and each subproblem was solved separately.

Before actually solving each subproblem, the staff of OTA selected the twelve items in Table 2.1 as essential attributes to be preserved in the reduced SOI sample. Summary statistics–aggregate total, mean, standard deviation, and coefficient of variation–were computed before and after the reduction for each subproblem. Analysis of these statistics revealed that the smallest percentage deviations between the full and reduced sample occurred when adjusted gross income (AGI), taxes paid (TP), and business income/loss (BI) were included in the objective function. The inclusion of additional attributes in the objective function produced summary statistics that were generally, but not always, better than the results obtained when the additional attributes were excluded. Although including additional attributes improved the result, this gain was often offset by a degradation in computational efficiency.

The objective function used in the optimization model measured the loss of information via the functional relationship below:

$$L_{ij} = [\ (AGI_i - AGI_j)^2 + (TP_i - TP_j)^2 + (BI_i - BI_j)^2\ ] \qquad (2.1)$$

Where i is a point in the original SOI sample.

j is a point in the reduced SOI sample.

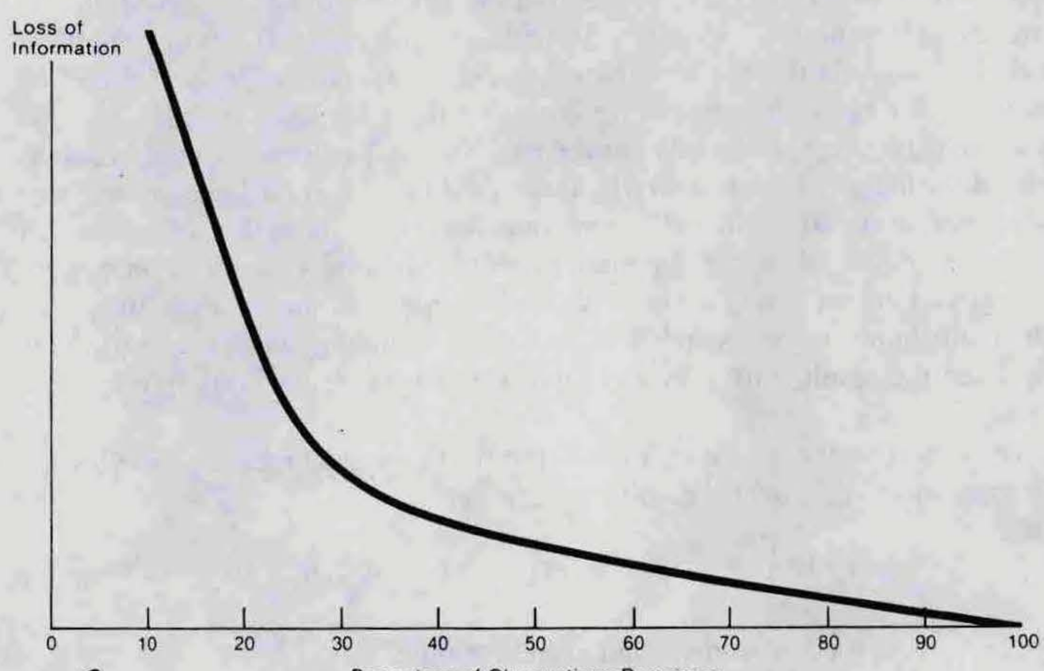
L_{ij} is the loss that results from letting point j represent point i.

AGI_i is adjusted gross income for point i.
AGI_j is adjusted gross income for point j.
TP_i is taxes paid for point i.
TP_j is taxes paid for point j.
BI_i is business income/loss for point i.
BI_j is business income/loss for point j.

Figure 2.1 presents a graphical representation of a typical loss function. The loss function in equation (1) does not in and of itself guarantee that the twelve attributes in Table 2.1 will be preserved. However, the aggregate amounts, standard deviation, and coefficient of variation for these attributes are well preserved because the attributes not specifically included in the loss function are highly correlated with either AGI or taxes paid.

The number of returns retained in each of the twenty-three subproblems was determined on a case-by-case basis. In general, the subproblems were reduced in size until the loss function yielded an information loss which was less than or equal to an arbitrary standard. However, no subproblem was allowed to go below the 20 percent retention level because this produced unacceptable information losses. Further, the weight per observation was not allowed to exceed a trigger weight, so that no single observation represents an unusually large portion of the population. All subsequent steps in forming the tax reform data base were made to this reduced SOI sample.

Figure 2.1 **A Typical Loss Function**



c. Imputing Itemized Deductions and Other Data

The first actual modifications to the data base were the imputations of itemized deductions for non-itemizers and the lesser earning spouse's share of wage and salary income. The former imputation was required in order to analyze tax proposals that treat itemized deductions differently from present law, and the latter was required to simulate the deduction for married couples enacted in the Economic Recovery Tax Act of 1981. Itemized deductions were imputed to non-itemizers by analyzing 1981 tax returns that itemized, and the share of wage and salary income was imputed by analyzing 1980 CPS data.

d. Extrapolating the Data to 1983

The second modification was a statistical extrapolation or "aging" of the data base to 1983 levels of income and deductions. The extrapolation process consisted of two stages. The first stage was designed to reflect uniform growth and inflation, and was accomplished by imposing constant weight changes and proportional increases in all the dollar amounts in the data base. The second stage reflected more complicated trends superimposed on the uniform growth, and was achieved by adjusting only the weights of each return to meet target amounts for a set of selected variables. The net effect of this process ensured that the levels of income, deductions, credits, and exemptions were consistent with aggregate economic measures such as components of national income, inflation, employment, and population. The following subsections describe the extrapolation in further detail.

For the liability year 1983, OTA projected 96.7 million returns with total adjusted gross income of $1,997 billion. The 1983 file reflects these estimates. OTA also developed projections of several sources of income and other aggregates, as well as the distribution of AGI by broad AGI classes, and are reflected in the 1983 file. These estimates or "targets" are presented in Table 2.2.

The twenty-three targets were developed by Office of Tax Analysis staff using time series techniques, including regressions of aggregate Statistics of Income data against data from the National Income and Product Accounts. The targets were intended to be consistent with the short-range economic forecasts underlying the federal budget for fiscal 1984.

In choosing the items for which targets were needed, OTA relied on intuition and experience. In general, an item was not targeted unless time series techniques seemed likely to give a better projection than the first-stage extrapolation described below.

Stage I of the Extrapolation to 1983. The first step in the extrapolation process was designed to reach the targets for number of returns by filing status and for total adjusted gross income. This step consisted of simply multiplying the sample weight of each return by a factor based on filing status and multiplying every dollar amount (except itemized deductions) on each return by a second factor which was the same for all returns. The weight

Table 2.2 **Targets for the Extrapolated 1983 Tax Model**

Description	Target (Millions)
$ AGI < $ 0	-25,000
$ 0 < = AGI < $ 5,000	45,552
$ 5,000 < = AGI < $ 10,000	127,565
$ 10,000 < = AGI < $ 15,000	177,933
$ 15,000 < = AGI < $ 20,000	194,184
$ 20,000 < = AGI < $ 30,000	419,624
$ 30,000 < = AGI < $ 50,000	603,675
$ 50,000 < = AGI < $ 100,000	282,210
$ 100,000 < = AGI < $ 200,000	87,756
AGI of $ 200,000 & Over	83,901
Number of single returns	41.0
Number of joint returns	46.1
Number of married-separate or surviving-spouse returns	1.3
Number of head-of-household returns	8.3
Number of taxpayer and dependent exemptions	219.6
Number of aged and blind exemptions	15.1
Pensions in adjusted gross income	65,300.0
Net capital gains in adjusted gross income	36,900.0
Dividends in adjusted gross income (1979 Law)	52,400.0
Interest in adjusted gross income (1979 Law)	166,700.0
Wages & Salaries in adjusted gross income	1,660,100.0
Earned income credit	1,620.0
Investment tax credit	4,140.0

factors were chosen to reach targets for the number of returns by filing status. Once these targets were achieved, the dollar-amount factor was chosen to reach the target for total adjusted gross income.

The dollar amounts of itemized deductions were adjusted separately by multiplying each deduction type by its own growth factor. In addition, extraordinary growth or legislation required that four tax credits be adjusted exogenously: (a) the child care credit, (b) the elderly credit, (c) the foreign tax credit, and (d) the jobs tax credit. These changes achieved the targets for number of returns by filing status, aggregate AGI, itemized deductions, the four tax credits, and brought all other aggregates into a more plausible range, so that fewer iterations would be required in the second stage of the extrapolation process.

Stage II of the Extrapolation to 1983. The purpose of the second stage of the extrapolation was to change the sample weights in the presample in a way that would achieve all of the targets in Table 2.2. This amounts to solving 23 linear equations in 75,422 variables. All that is required is a criterion by which to choose one of the many solutions. Our criterion is stated in terms of an objective function that measures the change in the sample weights. Simply stated, the mathematical method selects the new weights in a way that will "hit" the targets and simultaneously minimize the objective function.

Since stage I of the extrapolation adjusted all data items proportionally for inflation (excepting special cases), the correlation among the data items on each return have, for the most part, remained unchanged and simultaneously, the important targets have been achieved. Achieving additional targets in this stage can only be accomplished by introducing distorting adjustments to the sample weights. Thus, our objective function contains an arbitrary measure of the distortion caused by changing the weight on any return. This measure depends on the ratio of the new weight to the old weight: if the ratio is close to 1.0, there is not much distortion but if the ratio is closer to zero or exceedingly large, the distortion is large. As the ratio approaches zero, the distortion approaches infinity. Let x represent the ratio of new to old weight for a particular return, and $\phi(x)$ represent the associated distortion (by our arbitrary measure).

For t10he distortion function ϕ, we chose the following:

$$\phi(x) = x^4 + x^{-4} - 2 \qquad (2.2)$$

This choice was arbitrary and can only be justified by the results and by these desirable features:

$$\phi(1) = 0 \qquad (2.3)$$

so there is no distortion when the weight is not changed.

$$\lim_{x \to 0} \phi(x) = \lim_{x \to +\infty} \phi(x) = +\infty \qquad (2.4)$$

so weights too far from 1 are not tolerated.

$$\phi(x) = \phi(1/x) \qquad (2.5)$$

so that the amount of the percentage change is independent of direction. For example, it is as distorting to multiply a weight by 4 as to divide it by 4. It can be shown that any function having these properties must be of the form:

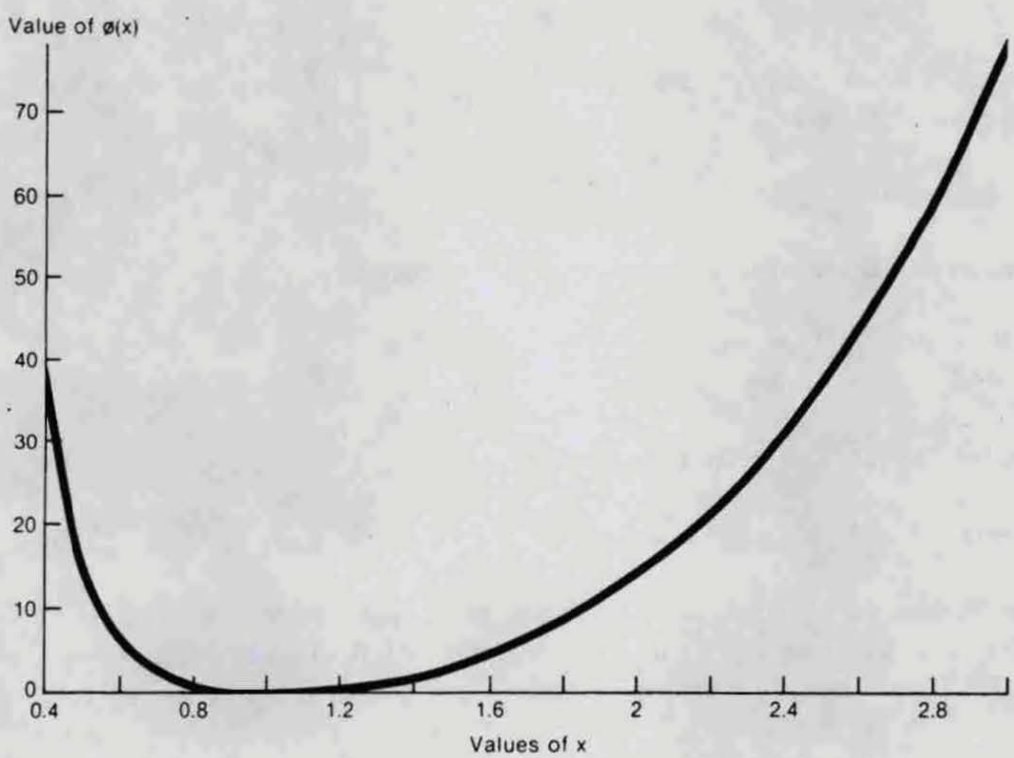
$$\sum^{+\infty} \beta_i \, (x^i + x^{-i} - 2) \qquad (2.6)$$

It was decided to choose $\beta_4 = 1$ and all other $\beta_i = 0$. The graph of ϕ is shown in Figure 2.2. The complete objective function is obtained by calculating the distortion $\phi(x)$ for each sample return and taking the sum; weighted by the original weights.

Results of the Extrapolation. All of the targets were achieved in Stage II with an average distortion of 0.198 per population return. On average, this was the amount of distortion that resulted from multiplying (or dividing) a sample weight by 1.1166.

Some of the weight changes, however, were extreme (the largest ratio of new to old weight was 8.67 and the smallest was 0.342). Increasing the exponent (that is, the power 4) in our objective function would have forced the extreme values closer to 1, with slightly more distortion among the less extreme returns. Decreasing the exponent would have allowed the extreme values to be more extreme. When the exponent 2 was used in a trial run, the largest factor by which a weight was increased was about 30.

Figure 2.2 **Graph of ϕ(x)**



e. Partitioning the File

The final step in preparing the SOI file was the creation of two extracts: one for taxpayers 16 years of age and over, and one for taxpayers under 16 years of age. This was done to ensure that comparable files would be merged and is discussed further in Section B.

2. CPS Preparation

Tax return data do not contain information on non-taxable income sources, demographic characteristics of the population, or people not required to file tax returns. To obtain these types of information, OTA used the data collected by the Bureau of the Census in its Current Population Surveys. Every March, the Census Bureau surveys approximately 50,000 households selected to represent the U.S. population, asking them questions about themselves, their family composition, incomes for the previous year, and employment experiences.

A somewhat involved set of adjustments was performed on the CPS data file before it was combined with the tax return data. First, the CPS file was extrapolated from 1982, the most recent year for which data were available, to 1983 levels, the same year as the extrapolated tax return data.

Second, transfer payments (except for welfare payments) were corrected for nonreporting and underreporting in order to produce aggregates consistent with national estimates. For example, public and private pensions, including social security, were adjusted. The correction process grows existing transfer payments and assigns actual transfers to new units based on a vector of socio-economic attributes (a procedure known as "hot decking").

Third, both welfare payments received and the number of recipients were simulated based on the characteristics of the households on the CPS file and the criteria for eligibility for welfare programs in 1983. This simulation effectively replaces all of the original data. However, units that originally received transfer payments, in general, still receive transfer payments.

Next, households in the CPS were converted into potential tax filing units by applying tax filing rules to each person age 16 and over in the family. This placed the CPS file on a comparable basis with the file of taxpayers age 16 and over. This process created two CPS extracts: (a) potential filers and (b) potential non-filers (CPS tax units who do not need to file a return under 1983 law because their incomes are too low). In families with multiple tax units, units other than the family head were deemed to be "dependent filers."

The final set of adjustments, called the premerge alignment, consisted of two steps. The first step, the filer/non-filer alignment, ensured that the CPS file represented the same number of tax returns as on the SOI file. This process created additional CPS filers from the extract of potential CPS non-filers. The second step corrected certain forms of income for underreporting and nonreporting on the CPS. This process ensured that the files of both tax returns and CPS filers had the same amounts of income from major sources. It also ensured that certain forms of income and the number of people reporting income from these sources were consistent with national aggregates. For example, the CPS file consistently showed smaller amounts of interest, dividends, and other forms of capital income than what is reported on the SOI file. Consequently, as part of the premerge alignment the amounts of capital income and the number of units reporting capital income on the CPS file were increased.

B. The Merge Process[4]

At the beginning of the merge process there were four 1983 level files:

- The SOI with taxpayers age 16 and over,
- The SOI with taxpayers under age 16,
- CPS filers, and
- CPS non-filers.

The preparatory work ensured that the SOI files and the CPS files were consistent with each other and were both representative of the U.S. population.

Two of these files, the SOI with taxpayers age 16 and over, and the CPS filers, were merged on the basis of a common set of core variables, principally the dollar amounts of major income sources and key demographic variables. The merge process was a "soft match," or a "statistical match," which applied a state-of-the-art extended transportation algorithm capable of determining which SOI return should be matched with which CPS return. This determination was made by minimizing a "penalty function" that expresses the statistical cost of mismatching, i.e., matching a SOI return and a CPS return with dissimilar common core variables. Some CPS records were matched to multiple SOI records and vice versa because the SOI file had a high proportion of returns from upper-income classes while the CPS file had many tax filers from low-income classes.

Once these two files had been merged, the CPS non-filers were appended. Next, the data base was sorted by CPS family identification numbers to reconstruct families. Finally, taxpayers under age 16 were merged with families judged to have appropriate income and demographic characteristics. This was accomplished by partitioning further the SOI file with taxpayers under age 16 into three sets:

- Returns with income less than $200,000 and with at least some wages and salaries,
- Returns with income less than $200,000 and from capital income only, and
- Returns with income in excess of $200,000.

Returns from partition 1 were randomly merged with families from all income levels while returns from partition 2 were randomly merged with families earning more than $50,000 annually. Finally, returns in partition 3 were randomly merged with families whose income exceeded $200,000. To qualify for selection, the age of a taxpayer in a SOI partition had to exactly equal the age of a dependent in the family. Furthermore, the sample weight for a return in the SOI partition had to be equal or similar to the family weight.

The resulting merge file contained nearly 200,000 records which represent 122 million potential tax units: 97 million tax filers (of which 13 million are dependent filers) and 25 million non-filers. At this point in the process, each record contained approximately 500 data items from the tax return, SSA, and CPS data.

The merge process has both strengths and weaknesses. One of the main advantages of merging two data sets through a statistical match is that the procedure preserves the relationships among the variables in each file, i.e., it maintains the variance-covariance matrices. However, one of the shortcomings is that, beyond the common core set of variables, nothing is really known about how the SOI variables are correlated with CPS variables. The statistical reliability of the match depends, moreover, on the assumption that for each merged observation, the values for the non-core SOI variables were statistically independent of the non-core CPS variables.[5]

C. The Imputation Process

1. Imputations to the Merge File

Once the MERGE file had been completed, additional information was imputed to the data file. Approximately 250 data items were imputed to provide information such as:

- Industry where the taxpayer is employed,
- Employer contributions to entitlement programs, insurance plans, pension plans, and fringe benefits,
- Tax exempt interest,
- Unreported and underreported income for filers and non-filers,
- Earnings on pension and life insurance funds,
- Home ownership variables such as market value, rent, depreciation, and maintenance expense,
- IRAs, KEOGHs, and unemployment compensation,
- Taxpayer's share of corporate profits,
- Types, holding periods, selling prices, and capital gains or losses of capital assets sold during the year,
- Unemployment compensation,
- Military benefits,
- Food stamps, and
- Consumption for 32 different items.

For most variables, the imputation process consisted of five steps.

(1) Selecting the Item to be Imputed. In this step, the exact item to be imputed was conceptually defined, and the aggregate amount to be attributed to the MERGE file was determined.

(2) Selecting a Data Source. A reliable exogenous source of data that contained attributes in common with variables already on the MERGE file was obtained. Sources included special surveys made by other government agencies or private organizations, or tax returns filed in another year. Although another micro-data file was the preferable source, sources were sometimes cross-tabulations or if necessary, simple tabulations.

(3) Determining a Statistical Relationship. An estimate was obtained of the statistical relationship between the imputation item and the attributes in common with variables in the MERGE data and the source data. If the data source was a micro-data file, the relationship was usually a regression equation but sometimes was a table of probabilities and average amounts. If the data source was a cross-tabulation or simple tabulation, the tabulation had to provide the statistical relationship.

(4) Designing the Imputation Methodology. Using the estimated statistical relationship, a methodology was designed to impute the item to the appropriate records on the MERGE file. If the source data were extensive, this step applied sophisticated econometric techniques to reproduce some of the unexplained variation in the original item. Even if the source data were very simple, a two-step procedure was often used: (a) records were randomly selected, in accordance with a probability distribution, to receive the imputed item, and (b) the amount of the item was calculated in accordance with its distribution about some mean amount.

(5) Calibrating the Imputation. To assure that the total for any imputed item (summed across all the records in the file) equaled the correct aggregate for the economy, imputed amounts were usually adjusted or calibrated. These adjustments were necessary because the imputation process was imprecise and the populations represented by the source and MERGE data frequently differed.

2. Additional Imputations for Tax Reform

Several additional imputations were made in the middle and latter stages of tax reform. These imputations were frequently made in response to a particular tax reform proposal. So, the level of detail in the imputations often reflect the rules or definitions contained in the reform proposal. Further. the imputations were often adjusted because of new information, or to reflect changes in the original proposals. Unlike the imputations listed in the previous section, these imputations were not included in the data file. They were created in each simulation where they were required. Some of the more important additional imputations were:

- Interest expense on second homes,
- Meals and entertainment expenses (for wage and business returns),
- Adjustments to percentage depletion preferences,
- Adjustments to intangible drilling costs preferences,
- Health insurance for the self-employed,
- Passive partnership income,
- Passive farm income,
- Passive rental income,
- Active and passive partnership rental incomes,
- Passive business interest expenses,
- Adjustments to income from changes to at-risk rules for real estate activities,
- 401(k) pension contribution levels, and
- Pre-retirement distributions (withdrawals) of pensions.

D. Extrapolation Over Multiple Years

During the tax reform process, policymakers became interested in the revenue effects of certain tax law changes over the five year period immediately following enactment. In particular, policy makers wanted to know the revenue effects of tax reform proposals between 1986 and 1990, and subsequently between 1987 and 1991.

To address these questions, a procedure was developed whereby the data base could be extrapolated from its 1983 levels to any year up to 1995. The procedure, largely developed by the staff of the Joint Committee on Taxation, consists of adjusting the values of variables for each return by a set of growth factors for each year. The weights of each return are adjusted to reflect changes in the demographic profile of the tax filing population. The growth factors for important variables are set by income class, while less important variables have a single growth factor. Further, the growth factors are easily adjusted to reflect new or alternative economic forecasts.[6] The revenue estimates of individual tax provisions in the Tax Reform Act of 1986 (P.L. 99-514) are largely based on model simulations using files extrapolated by this procedure.

III. DESCRIPTION OF THE SIMULATION MODEL

The computer program component of the Treasury's individual income tax simulation model consists of approximately 12,000 ASCII FORTRAN statements and 250 control stream instructions. The FORTRAN program contains 114 subprograms which are grouped to form three parts or phases. Each phase is executed separately after executing the previous phase.

Since January 1981, the Treasury individual income tax simulation model has been implemented on a UNIVAC 1100/82 computer. On this system, the program's execution time varies with the size of the tax sample and with the complexity of the tax proposal under analysis. A simulation on the full Tax Reform Data Base of 200,000 records requires approximately 25-35 minutes of processing time while the smaller SOI sample of 75,000 returns requires approximately 10-12 minutes.

Section A below provides the reader with an overview of the three phases in the individual income tax model. Phase I, the input of parameters is briefly explained in Section B. The tax calculator, Phase II, is discussed in some detail in Section C and the outputs produced by the model, Phase III, follow in Section D. Finally, Section E describes the post-simulation adjustments to the model results.

A. Design Philosophy of the Tax Model

In the tax model, the tax law currently in effect is defined by various tax parameters which collectively are called Plan X. Similarly, the set of tax parameters which define a proposed tax law are referred to as Plan Y. Normally, the number of differences between Plan X and Plan Y are small in relation to the total number of tax parameters in Plan X. The design of the tax model program incorporates this feature by initially assuming that no differences exist between Plan X and Plan Y, i.e., that Plan X = Plan Y. This design simplifies data preparation for the user, who has only to specify how Plan Y differs from Plan X in order to completely define Plan Y.

Another design feature of the tax model program is its capability of comparing two alternative tax proposals with Plan X. Normally, a user will want to compare Plan Y with Plan X and analyze the results. In some cases, however, the user may want to compare one tax proposal, Plan Y, and a second tax proposal, Plan Z, with Plan X in the same simulation to determine which option individual taxpayers would elect.

Finally, it is desirable to have a tax model program that could produce summary results, detailed results, or both, while minimizing execution time. Consequently, all of the statistical tables, except the "standard" tables, are optional and must be specifically requested.

1. Phase I

Phase I of the simulation is basically an initialization and preparation stage. It begins by inputting operational data such as the run description, the characteristics of the data sample that will be inputted, the desired output tables, the editing required on these tables (if any) and which tax plans will change in the simulation. Next, all tax parameters in Plan Z, if requested, are equated to those of Plan X. (Plan X and Plan Y are already equated to each other unless specified otherwise). The remaining data are inputted and used to modify the tax parameters in one or more of the tax plans so that Plan Y does not equal Plan X (and Plan Z if relevant). The remaining execution of Phase I prints the tax parameters under the alternative plans and prepares the stubs for the tables of taxable income by marginal tax rates, if they were requested.

2. Phase II

The primary function of Phase II is to process the data from the sample of individual income tax returns under each tax plan that was specified in Phase I. First, Phase II calculates the amount of storage space needed to produce each of the tables requested by the user and acquires this storage. Second, each data record is inputted and tax liability for the record is calculated under Plan Y and if requested, Plan X. Third, the requested statistical tables are updated. After all returns have been processed, operational statistics are printed, and the statistical tables are outputted to a temporary file.

3. Phase III

Phase III of the tax model is designed to process and output the results of Phase II. Initially, Phase III calculates the maximum amount of additional storage required to produce the requested tables and acquires this storage. Immediately thereafter, the program will input each statistical table (requested in Phase I) from the temporary file (created in Phase II). After the particular table has been inputted, a subprogram designed to perform the final processing and output for that table is executed. The subprogram will output each page of the table that the user requested in the Phase I input.

Finally, Phase III permanently stores summary results of the tax model run and prints the processing and elapsed times required for the execution of the simulation.

B. Description of the Tax Model Input Parameters

Each tax model simulation solicits operational data in eight steps. This data provides information necessary to prepare the run, select what output will be produced and determine how certain tax parameters will be changed. This data when combined with changes to the tax calculators, completely document a tax model run. Briefly, the eight specification steps are:

(1) Describe and identify the run.
(2) Specify the data sample to be used and the number of observations to be processed.
(3) Specify which basic tax regime will be used for Plan X (e.g., 1986 law).
(4) Specify which tables are to be printed and the unit of analysis.

(5) Specify for the tabular output the income concept used for distributional classification, the stubs for the income classifier, the specific pages to be printed, and the desired disaggregation of the tables.
(6) Specify which tax plans will have parameter changes or rate changes.
(7) Specify the changes in tax parameters for each plan.
(8) Specify any changes in the regular and alternative minimum tax rate schedules.

The tax model currently contains tax parameters for six tax regimes; 1981 law through 1984 law, 1986 law, and a fully phased-in law at 1984 levels. (A parameter set for the Tax Reform Act of 1986 is currently being formed.) The desired tax regime is specified in step 3.

An important input parameter specified in step 7 is an inflation indexing value. Consider the case where the tax regime being examined is 1986 law. Since the basic data is at 1983 levels, the dollar values of the tax parameters need to be deflated to 1983 levels (i.e., 1986 law at 1983 levels). The analyst only needs to specify that 1986 parameters will be used and the appropriate indexing factor between 1983 and 1986.

Now then, consider the case where the tax regime being examined is 1988 law. Some, but not all parameters are indexed for inflation by statute beginning in 1984. In this case, those parameters that are indexed need to be inflated to their (expected) 1988 levels, then all parameters need to be deflated to 1983 levels. This procedure has been routinized so that the analyst only needs the indexing factors (CPI) for 1983 through 1988.

C. The Tax Calculator

The heart of the individual tax model are three almost identical tax calculators that are used in Phase II to calculate Plan X, Plan Y, and Plan Z tax liability. The only difference between the calculators is the names of the variables they determine. Thus, AGI in Plan X (AGIX) can easily be distinguished from AGI from Plan Y (AGIY).

The calculator takes information from each potential tax filing unit in the data file, and using a set of tax parameters, calculates that unit's Federal individual income tax liability under the appropriate tax plan. The tax calculator also computes the values of several variables that affect tax liability. Table 2.3 presents a list of variables that are endogenous to the tax calculator. (Several more items will be added to this list when tax reform law becomes a standard part of the model.) Note that the calculator does not endogenously determine levels of capital gains (or losses), unemployment compensation, or social security income. It only determines the portion of these items included in AGI.

Not every variable in Table 2.3 is determined in every simulation. Some variables in Table 2.3 may not be applicable to the tax law requested by the

Table 2.3 **Items Recalculated by the Tax Calculator (Prior to the Tax Reform Act of 1986)**

Capital Gains included in AGI.
Dividend (and interest) exclusion.
All savers interest included in AGI.
Deduction for a married couple when both work.
Unemployment compensation included AGI.
Social Security income included in AGI.
Adjusted gross income (AGI).
Zero bracket amount (ZBA).
Personal exemption.
Medical expense deduction(s).
Casualty and loss deduction.
Total and excess itemized deductions.
Unused ZBA (for returns required by law to itemize).
Earned income (for dependents with unearned income).
Taxable income.
No benefit exclusions for minimum and alternative minimum taxes.
Itemized deductions for minimum tax purposes.
Regular tax liability.
Statutory marginal tax rate.
Personal service income (PSI).
Taxable PSI.
Tax savings from the maximum rate on PSI.
Tax liability after tax savings.
Child care credit.
Energy credits.
Political contributions credit.
Tax liability after credits.
Total itemized deduction for minimum, alternative minimum, maximum, and alternative maximum tax purposes.
Total preference items for minimum and alternative minimum tax purposes.
Taxable income for the minimum and alternative minimum taxes.
Tentative minimum tax on tax preferences.
Minimum tax and alternative minimum tax.
Earned income credit.
Final tax liability.

user. For example, the maximum tax on personal services income is only calculated when pre-1982 law is requested. Similarly, the individual minimum tax is only determined when pre-1983 law is requested.

There are three basic methods for simulating a tax regime other than current law: change tax parameters, change the coding of the tax calculator, or change both. Tax parameters are changed using a set of input parameters (described in the next section). The tax calculator is changed by deleting certain lines of code, inserting new lines of code, or by doing both.

Two implicit behavioral assumptions are used in the simulation model. First, all filers choose tax options that minimize their tax liabilities. Second, there are no behaviorial or "feedback" effects on variables exogenous to the model. These variables include the level and distribution of pre-tax income or expenditures on deductible items. This latter assumption implies that only first round effects are simulated. Although one may believe some feedback effects will occur, assuming no feedback effects is often the logical starting place for analysis.[7]

Limitations. The tax calculator is quite robust, and can trace through most of the interactions between any income source and the various provisions of the Tax Code. However, it does have limitations. These include:

- The calculator does not simulate the effects of changes in tax law on a number of small but important provisions in the tax code. Among these are state and local tax refunds, income averaging, the foreign tax credit, the elderly credit, and the investment tax credit. Values for these variables are assumed to remain constant within the simulations.
- The calculator does not capture behavioral responses of individuals to marginal changes in disposable income or prices. For example, the model would not capture changes in charitable giving due to changes in tax rates or incomes.
- Simulations often require imputed data (data not contained on original tax returns but necessary in computing tax liability under one or more plans). Although every effort is made to correctly impute data items, errors are unavoidable. Within any simulation, variables containing imputed data will interact with other variables in the model. To the extent that imputation errors were made to one or more variables in some systematic way, the results of simulations will be biased.
- Intertemporal interactions are not accurately captured by the tax calculator (at least at the micro level). For example, changes to rules relating to loss or credit carryovers not only affect the current tax regime, but also affect future tax regimes.
- Finally, many of the imputations for non-itemizers are based on values of variables on tax returns such as AGI. For example, the imputation for state and local income tax deductions was based, in part, on AGI. Therefore, these imputations should be adjusted when the "base" variables change. For simplicity, such adjustments are not made.

While these limitations are easy to identify, they are difficult to correct. So, these limitations may be corrected "off model." That is, if the analyst believes the tax model does not accurately portray the effects of a tax law change, he may make adjustments to the final output of the simulation run. Post adjustments to tax model simulation results are discussed in Section E.

D. Outputs Produced by the Model

The output from the simulation model consists of six sections but most of these are of a operational nature and will not be discussed here. One section, however, is of paramount importance to the analyst: the tabulated tax statistics.

The tabulated tax statistics consist of "standard" output tables and optional output tables. The standard output tables are produced by default. These tables present information that, through experience, analysts have found to be most useful. The standard tables are produced on either a tax return or family basis.

Table 2.4 lists the tax model tabulations for analysis on a tax return or, family basis. The Appendix provides a more detailed description of the tables

Table 2.4 **Tax Model Tabulations for Analysis by Output Table Number**

Output Table No.	Contents of the Tabulation
Tax Return Basis	
1	Distribution of Income Tax Liabilities
2	Distribution of Average Tax Liabilities
3	Sources of Income and Adjustments
4	Taxable Income, Deductions, and Credits
5	Distribution of Tax Change and Returns
6A	Cumulative Taxable Income and Tax by Marginal Tax Rate
6B	Taxable Income by Marginal Tax Rate
7	Distribution of AGI, Tax, Effective Tax Rates, and After Tax AGI
8A	Summary Effects of the Proposal by Filing Status
8B	Aggregate Summary Effects of the Proposal
9	Tax Burdens for Hypothetical Tax Returns
10	Distributions by Filing Unit Size
11	Computation of Iles (e.g., deciles).
Family Basis	
12	Distribution of Income Tax Liabilities
13	Distribution of Tax Change and Families
14A	Summary Effects of the Proposal by Family Size
14B	Aggregate Summary Effects of the Proposal

produced for analysis on a tax return and family basis. These descriptions are not intended to be comprehensive but rather to provide an overview of the model's capabilities.

E. Post-Simulation Adjustments to the Model Results.

The tax model does not always capture some known effects of a given tax regime or the effects of a change in the tax code. The tax model may also be incapable of providing the exact information desired by the analyst. There are several reasons why this may occur:

- The data may not be at the level desired by the analyst. Tax data, like any other data, is rarely as current as the analyst would like it to be.
- The data may be incomplete. For example, information on alternative minimum tax preference items is only available from those returns that actually paid minimum or alternative minimum taxes. No imputations were made to people who have these preferences but did not pay these special taxes.
- The appropriate data may not be on the data file. For example, the data file does not contain information on scholarship or fellowship income.
- Inappropriate targets may have been used in the extrapolation procedures. The target values in any extrapolation process are simply estimates of future levels of variables. Unavoidably, the target values will not be the same as the actual values. This problem is exacerbated when the target levels are determined before some change in the economic environment (e.g., a change in tax law). Thus, distributions and aggregate amounts in an extrapolated file will not be exactly the same as those found in an actual file. Further, the extrapolation process is expensive and time consuming, and so a new extrapolation is not made every time new information about the future becomes available.
- The individual model is used to make tax revenue estimates under various economic scenarios (e.g., different interest rates, inflation rates, or rates of economic growth). Again, the costs of creating a new projected file for each variant in economic conditions is prohibitive.
- The tax calculator currently does not capture behavioral responses or the interaction among certain variables.
- In addition to analyzing tax liabilities, the tax model is used to estimate changes in tax receipts. Tax liabilities are generally analyzed on a calendar year basis, while receipts are analyzed on a fiscal year basis. In addition, there is a time lag between the time when tax liabilities are incurred and when the taxes are actually paid. So, calendar year tax liability estimates are converted to fiscal year tax receipts.

To deal with these problems, off-model adjustments are made to the tax model results. For example, the analyst may inflate or deflate values, or he may adjust the output values of certain variables based on some separately acquired elasticity estimates. The adjustments may be simple or sophisticated, depending on the questions being answered. Further, the tax model will often be used in making off-model adjustments.

This emphasizes the fact that the individual tax model is only one tool used by the analyst. It does not and cannot provide the definitive answer to all questions relating to individual income taxes. So the analyst must use other tools to answer questions he faces.

IV. SIMULATING THE EFFECTS OF TAX REFORM

The individual tax model became a veritable workhorse during the tax reform process. Over the last four years, the model was used to estimate the effects of thousands of proposed changes to the tax code in developing Treasury I and II and that arose before and during the House mark-up, the Senate mark-up, and the Conference Committee deliberations. The proposed changes could be very simple (e.g., changing the floor on medical deductions from 5% of adjusted gross income to 7.5%), or they could be very complex (e.g., imposing a 28% top rate on capital gains).

Although the individual income tax model is used to answer many questions, the kinds of analyses that it addresses can be classified into ten general groups. These are enumerated below.

(1) Estimates for a Single Provision or Set of Provisions. To examine the effects of a single provision or set of provisions, Plan X and Y are set so that they only differ by the provision or set of provisions being examined. Note that Plan X does not need to be present law, so that, for example, the model could be used to estimate the effect of a single variant of the House bill.

(2) Stacking Series. When examining a set of multiple provisions, the analyst may want to separate the effects of the package into the component provisions. This is accomplished through a stacking series. In a stacking series, each provision is estimated separately. Three types of stacking series are possible; stacked-first, stacked-last, and stacked-sequentially. Under stacked-first, each provision is compared to the same Plan X (usually current law). Under stacked-last, Plan X contains all of the provisions except the one being examined, while Plan Y contains all provisions. Stacked-sequentially starts out with present law (or some other base tax regime) in Plan X and the same law plus one provision in Plan Y. In the next simulation, the old Plan Y becomes the new Plan X, and a provision is added to the new Plan Y. This continues for all provisions. In some cases, it is possible to stack the provisions recursively so that each added provision does not interact with any of the previously stacked provisions.

(3) Interaction Among Provisions. The tax model is an excellent tool for estimating the interrelationships among individual tax provisions. Because of

interactions among provisions, adding up the effects from each of the stacking series will not give the combined effect of all of the provisions. For example, changing tax rates will affect the values of deductions, and changing the levels of deductions will affect the amount of revenue generated by the rate structure.

(4) Sensitivity Analysis. Often the analyst is interested in how responsive the tax system is to certain tax parameters. Using the model, the analyst can determine the change in tax liabilities due to a change in a tax parameter. The analyst may also be interested in the sensitivity of the tax system to economic variables. For example, the analyst can get a sense of how interest rates affect certain tax variables by adjusting interest income and interest expense items.

(5) Estimating Average Marginal Tax Rates. One of the most useful attributes of the individual tax model is its ability to estimate average marginal tax rates. To obtain marginal tax rates, the Federal income tax liability under a given tax regime is calculated for each return on the data file. Then, the value of an income source is increased by a small amount (usually one percent). The model recalculates each return's tax liability and the change in tax associated with the change in income. The tax and income changes are summed over all returns, and their ratio computed to give the average marginal tax rate for that income source. This, in effect, gives a rate that is weighted by the amount of the particular income source appearing on each return. Alternatively, the marginal tax rate for each return can be multiplied by that return's weight in the population. This leads to an average marginal tax rate that is weighted equally among all returns (or all returns with that particular income source).

Table 2.5 contains estimates of average marginal tax rates under 1980 law, 1986 law, and P.L. 99-514, the Tax Reform Act of 1986 (TRA). Under each tax regime, marginal tax rates are presented for wages, interest income, dividend income, noncorporate business income, and net capital gains. All computations were performed at the 1983 level and distribution of income. To be consistent, all dollar-valued tax parameters were inflated or deflated to 1983

Table 2.5 **Average Marginal Tax Rates on Selected Income Sources under 1980 Law, Current Law, and the Tax Reform Act of 1986**
(1983 income levels)

Income Source	1980 Law	Pre-Tax Reform Law	Tax Reform Act of 1986
Wages	29.1 %	25.8 %	21.7 %
Interest	28.5	25.5	21.5
Dividends	39.0	32.9	25.7
Non-Corporate Business	29.6	25.0	21.0
Net Capital Gains	16.6	13.8	21.5

levels using the consumer price index (CPI). So, to calculate 1980 law, all tax parameters were inflated from 1980 levels to 1983 levels. Similarly, the calculations for 1986 law deflated all parameters to 1983 levels. For most parameters under TRA, the dollar-valued parameters were deflated from 1988 levels to 1983 levels using the 1983 CPI and the (estimated) 1988 CPI. Exceptions occurred when transition rules extend beyond 1988 (e.g., transition rules apply to the personal exemption until 1989). In these cases, the parameters were deflated from their fully phased-in levels to 1983 levels.

The average marginal rates under each of the three laws in Table 2.5 are fully phased-in. That is, the rates are calculated without regard to any transition rules. For example, under fully phased-in 1986 law, there is no above-the-line deduction for charitable giving since it expires after 1986. Similarly, the TRA provisions assume all of the transition rules have already occurred. (Under the fully phased-in assumption, it was implicitly assumed that all dollar parameters are indexed for inflation beginning in 1986.)

(6) Incorporation of Behavioral Responses. As mentioned earlier, the tax calculator does not account for any behavioral responses to changes in the tax code other than tax-minimizing choices like the decision to itemize. Estimates of behavioral responses are generally made "off model." On occasion however, the analyst may explicitly define a behavioral response to a particular change in the tax code. For example, In one project the price and income elasticities of charitable giving were incorporated into the model to estimate how particular changes in the tax code affect levels of charitable contributions.

(7) Distributional Analysis. The model is also extensively used for distributional analysis. For example, policymakers are interested in how individual income tax liabilities, and changes in these liabilities are distributed among the population. As part of the standard output of the model, the distribution of tax liabilities of tax units or families is presented in every model simulation.[8] However, because the model uses microdata, the effects of tax reform can be seen on almost any subgroup of the population.

Because of the tax model's rich data base, individual income tax liabilities can be distributed by almost any income classifier desired. The most common classifier used during tax reform was AGI. However, AGI is often an inappropriate income classifier because it is tied to a particular tax law, and because it does not accurately measure "economic well-being." So, an "economic income" measure was constructed that approximates the Haig-Simons definition of income (consumption plus change in net worth).[9]

(8) Gainers vs Losers. A common question that arises with any tax reform proposal is, who gains and who loses. A gainer is usually defined as a tax return or family that pays less tax, while a loser pays more tax. A special class of gainers are those returns or families who are removed from the tax rolls (tax liability changes from positive to non-positive) under the proposal. Conversely, a special class of losers are tax returns or families who were previously non-taxable, but must pay some tax under the proposal.

As part of the standard output of the model, the number of gainers and losers and the amount of their gain or loss are tabulated. The output includes the special gainers and losers who are either removed or added to the tax rolls. Additionally, the model can break down the gainers and losers by filing status (or family type), and whether the return itemizes or not. A portion of standard output of the individual tax model is shown in Table 2.6 The table shows at 1983 levels the number of families with tax increases or tax decreases from the individual provisions of the Tax Reform Act of 1986.

(9) Analysis of Non-Income Taxes. Because of its rich data base, the tax model has also been used for analysis of non-income taxes. For example, the tax model has been used to determine the distribution of Federal excise taxes on gasoline, as well as the level and distribution of social security taxes. Also, as part of Treasury I, the model was heavily used in simulating a value added tax.[10]

(10) Special Cross Tabulations. Because the tax model relies on microdata, cross tabulations can be made on any set of variables before and after a tax change. Frequently, the analyst may want to know how a particular subgroup of the population fares under a tax proposal. The analyst may also want to know the characteristics of gainers or loser (e.g., what variables cause a person to be a gainer or loser). The analyst often wants information on the levels of certain variables before and after the proposal (e.g., the level of medical expenses deducted or IRAs purchased). The tax model is a highly flexible tool for addressing such questions.

V. FUTURE MODEL DEVELOPMENTS

As the focus of tax policy shifts, the individual income tax simulation model requires new capabilities that often cannot be anticipated. Furthermore, data is constantly replaced with more current, more accurate, and more comprehensive data. Since the development of a modeling system is dynamic, there is always room for improvement and the individual tax model is no exception.

This section outlines improvements to the individual tax model that Treasury hopes to pursue and achieve in the near future. These improvements can be grouped into two categories: data improvements and simulation model improvements.

A. Data Improvements

Extrapolation. Of particular importance to OTA is the ability to accurately and efficiently extrapolate the tax model's data base to future levels. This involves simultaneously adjusting data values and/or weights to expected levels and distributions of target variables, while maintaining the "integrity" of each data record. To accomplish this, an efficient convergence algorithm is required. Further work is also needed in accurately predicting the levels and distributions of the target variables.

Table 2.6 **Distribution of Tax Change and Number of Families Affected by Economic Income for Families**
(1983 Levels of Income)

	Presently Taxable Families Made Nontaxable		Presently Nontaxable Families Made Taxable		Families with a Change in Tax Liability					
					Tax Decreases			Tax Increases		
Economic Income (Dollars)	Number of Families (Units)	Amount of Tax Decrease ($ Mil)	Number of Families (Units)	Amount of Tax Increase ($ Mil)	Number of Families (Units)	Amount of Tax Decrease ($ Mil)	Plan X Tax Liability ($ Mil)	Number of Families (Units)	Amount of Tax Increase ($ Mil)	Plan X Tax Liability ($ Mil)
Under 0	7459.	-.641	15671.	321.687	54008.	-13.205	29.237	29020.	522.181	115.953
0 – 10000	731277.	-121.852	88411.	48.034	3957447.	-493.511	929.921	528548.	183.883	293.569
1000 – 15000	1597172.	-1002.689	84778.	38.086	7499738.	-1947.672	4061.943	1076631.	277.802	741.804
15000 – 20000	1343858.	-822.211	80509.	55.685	7912581.	-2374.442	7780.005	1404243.	488.746	1361.253
20000 – 30000	871054.	-487.649	166339.	140.556	13172149.	-4365.714	24260.029	3416366.	1282.606	4858.896
30000 – 50000	219448.	-107.142	132777.	174.299	15639579.	-8009.392	56326.663	4938437.	2871.225	13083.117
50000 – 100000	26886.	-11.833	50309.	190.032	8407182.	-7067.475	61556.928	3686113.	4400.159	21516.679
100000 – 200000	20.	-.002	5253.	79.029	1066485.	-3215.352	22828.720	479248.	2656.358	7325.575
200000 and Over	90.	-1.654	3777.	390.354	277420.	-7535.238	27255.161	119302.	4794.185	10844.304
All Family Types	4797264.	-2555.672	627825.	1437.761	57986589.	-35022.001	205028.604	15677909.	17477.145	60141.147

Note: Behavioral changes not included. Numbers reflect many but not all the provisions of the individual income tax code.

Imputations. Several improvements to the imputation process are expected to occur when the next tax reform data base is created. For imputations to non-itemizers, OTA is exploring the possibility of using a maximum likelihood technique for estimating values of missing variables. In order to have a more internally consistent set of imputations, OTA is also examining the possibility of using a simultaneous imputation technique.

The results from several recent studies sponsored by the Treasury Department will be used for many new tax model imputations. For example, OTA is currently developing a data set which contains a sample of tax returns from the same taxpayers over a five year period. This data will be useful in estimating behavioral responses to changes in tax laws. OTA will probably incorporate data from IRS's Taxpayer Compliance Measurement Program (TCMP). This will aid in the analysis and estimation of tax compliance and evasion issues. Finally, OTA will soon obtain information from a W-2 study, which will provide information on wage and pension withholdings and the industries of taxpayers.

The Match Process. OTA will continue to develop and use state-of-the-art techniques when statistical matching data sets. Recent literature suggests that the conditional independence assumption (required in statistical matching) may be relaxed when additional information on the joint distributions of non-matching variables is available.[11]

OTA is considering the possibility of matching SOI data with something other than CPS data. In particular, a match with the Survey of Income and Participation is being considered. in order to place consumption on the data file, OTA is also considering a match between SOI data and data from the Bureau of Labor Statistics' Consumer Expenditure Survey.

B. Simulation Model Improvements

Behavioral Effects. One limitation of the tax model which OTA is attempting to correct is incorporating behavioral responses to changes in the tax code or the economic environment directly into the model, rather than relying on off-model adjustments. For a simple example, it may be possible to determine a relationship between charitable giving and marginal tax rates. Thus, values for charitable deductions would be adjusted whenever tax rates change.

Integration with OTA's Other Models. OTA currently maintains several tax models including a corporate tax model and a depreciation model. It would be useful if results from these models could be directly incorporated into the individual tax model. For example, changes in rules relating to depreciation are accurately captured in the depreciation model. So, a method is needed where changes in depreciation deductions, as calculated by the depreciation model, are translated into changes in taxable business income on the individual model. Similar cases can be made for integration with the corporate, partnership, and general equilibrium models.

Multi-year Model. The effects of almost any change in the tax code will span several years. Also, the impact of a particular provision in the tax code will not be the same in all years. Therefore, it is desirable to have a multi-year tax model. Such a model would simulate the effects of changes in the tax code or economic environment over several years.

Model Efficiency. All users would like the model to run faster. So, OTA frequently examines ways to make the model more efficient. Beginning this year, the tax model will be run on a VAX-8800 computer instead of a UNIVAC 1100/82. This change will make the model faster, more versatile, and have less downtime.

Receipts Estimates. In addition to estimating tax liabilities, it is also important to know when the tax liabilities will be received. This information is very important in formulating the Federal budget, and in estimating how much money the Federal government will need to borrow to meet its financial obligations. With some modification, the individual model might be a useful tool for estimating such things as withholding allowances and tax payment patterns.

Other. Several other improvements to the individual tax model can be pointed to. The model could be modified to simulate the effects of changes in social security tax laws. In the long run, OTA may divide the individual tax model into two sub-models. The first submodel would be used for revenue estimating purposes and would simulate the tax code for a specific year. The second submodel would be used for analytical or distributional analyses, and would simulate hypothetical, fully phased-in tax regimes.

APPENDIX
OUTPUT TABLES FROM THE INDIVIDUAL TAX MODEL

I. Tables for Analysis by Tax Returns

Table 1 presents the distribution of income tax liabilities and the percentage distribution of income tax liabilities for Plan X, Plan Y and the change between Plan X and Plan Y. This table also presents the percentage change in income tax liabilities, i.e., (Plan X tax - Plan Y tax) divided by Plan X tax. The results in Table 1 are distributed by size of present law AGI and are disaggregated by itemizers and non-itemizers.

Table 2 presents the distribution of returns and their average income tax liability for Plan X, Plan Y and the change between Plan X and Plan Y. This table also presents the percentage change in average income tax liabilities, i.e., (average Plan X tax - average Plan Y tax) divided by average Plan X tax. The results in Table 2 are distributed by size of present law AGI and are disaggregated by itemizers and non-itemizers.

Table 3 shows sources of income and the adjustments to gross income for each filing status (single, joint, head-of-household, and all returns) and for each deduction type (itemized, non-itemized, and both) distributed by size of present law AGI. It is a summary of Tables 1.4 and 1.5 in the "Statistics of Income" publication. Table 3 is the only table produced by the tax model that is unaffected by changes in any of the tax plans. Its primary use is to provide the user with a detailed breakdown of the income sources derived from the Treasury's sub-sample for comparison with the "Statistics of Income" publication in the base year and for forecasting income levels in future years.

Table 4 is a tabulation of taxable income, tax credits, deductions, and exemptions for each filing status (single, joint, head-of-household, and all returns) and for each deduction type (itemized, non-itemized, and both) distributed by size of present law AGI. It is a summary of various tables from section 2 of the "Statistics of Income" publication. Table 4 can be used in the same manner as Table 3, that is, for performing comparisons with the "Statistics of Income" publication when forecasting the future levels of credits, deductions, and exemptions. In addition, Table 4 may be used to estimate the effects of changes in the levels of credits, deductions, or exemptions by comparing the new levels with the levels computed for present law.

Table 5 presents a summary of the number of returns affected by the tax proposal and their associated tax change for each filing status (single, joint, head-of-household, and all returns) and for each deduction type (itemized, non-itemized, and both), distributed by size of present law AGI. Each return that has a tax change will be tabulated in either the "tax decrease" or the "tax increase" columns. In addition, a return with a tax decrease may also be tabulated in the "presently taxable returns made non-taxable" column and a return with a tax increase may also be tabulated in the "presently non-taxable returns made taxable" columns. Returns with either a tax increase or tax decrease may also be tabulated in the "returns which changed their type of deduction" columns.

Table 6A presents Plan Y taxable income and Plan Y tax before credits distributed by the marginal tax rate for the specified tax rate schedule (single, joint, head-of-household, and all combined) and for a specified Plan Y AGI size class. In addition, Table 6A provides the cumulative and percentage distribution of Plan Y taxable income and Plan Y tax before credits summed from the lowest to the highest and summed from the highest to the lowest. Table 6A is primarily used to compute new tax rate schedules yielding a predetermined revenue gain or loss.

Table 6B presents Plan Y taxable income by size of Plan Y AGI and by all Plan Y AGI classes for each marginal tax rate in the specified tax rate schedule (single, joint, head-of-household, and all combined).

Table 7 presents the distribution of AGI, tax liability, the effective tax rates (computed as a percentage of a user specified income concept), and the after-tax AGI for Plan X and Plan Y by filing status (single, joint, head-of-household, and all returns) and by size of present law AGI. This table also provides the percentage change in the effective tax rates and in the after tax AGI. These results are disaggregated by itemizers and non-itemizers.

Table 8A provides a summary of the effects of the proposed change (i.e., Plan Y) in the tax law and consists of four pages, one for each filing status (single, joint, and head-of-household) and a summary page for all returns. Table 8A together with Table 8B and Table 5 provides a comprehensive aggregate picture of the revenue and distributional effects of the proposed tax law changes. The results in Table 8A are distributed by size of present law AGI and are disaggregated by itemizers and non-itemizers.

Table 8B presents a summary of most of the items that have already appeared in Tables 8A and 5 as well as some items that do not appear anywhere else in the tax model output. Some of the additional items that appear in Table 8B are:

- The number of taxable returns under Plan X and Plan Y,
- The amount of AGI under Plan Y and the change in AGI,
- The amount of taxable income under Plan X,
- The number of itemized returns under Plan X and Plan Y,
- The number of returns with outlays under Plan X and Plan Y,
- The amount of earned income credit, minimum tax, and alternative minimum tax under Plan X and Plan Y, and
- The amount and change in outlays under Plan X and Plan Y.

The results in Table 8B are distributed by size of present law AGI.

Table 9 allows the user to examine the impact of a tax proposal under specific assumptions with respect to marital status, number of dependents, itemized deductions, and the level of adjusted gross income. In other words, Table 9 presents the tax burdens of hypothetical individual income tax returns for the income and deduction characteristics specified by the user.

Table 10 is designed to provide the distribution of returns and tax liabilities under Plan X and Plan Y by an alternative income concept by size of the filing unit and all filing units, i.e., it approximates a family unit analysis. In addition, Table 10 also provides the percentage distribution of returns and tax liability for each of the tax plans.

Table 11 allows the user to determine in one tax model simulation the level of a ranked variable, usually some income concept, that corresponds to 1/N-th of the total amount of a second variable, usually the population. If N = 10 and the ranked variable is adjusted gross income and the second variable is the number of returns, then Table 11 presents the level of AGI that corresponds to each tenth of the population of returns, i.e., deciles. Table 11 will accommodate values of N between 2 and 20.

Finally, the model has the capability to perform special tabulations not covered by the standard or optional tax model tables. These tables are specifically designed for a particular model simulation and are relatively easy to construct. In general, special tabulations may contain an unlimited number of rows and up to 100 columns.

II. Tables for Analysis by Family

Table 12 presents the distribution of income tax liabilities and the percentage distribution of income tax liabilities for Plan X, Plan Y, and the change between Plan X and Plan Y. This table also presents the percentage change in income tax liabilities, i.e., (Plan X tax - Plan Y tax) divided by Plan X tax. The results in Table 12 are distributed by the size of family economic income and are disaggregated by type of family (single individuals, husband and wife, other, and all types).

Table 13 presents a summary of the number of families affected and their associated tax change for each family type (single individuals, husband and wife, other, and all types) distributed by size of family economic income. Each family that has a tax change will be tabulated in either the "tax decrease" or "tax increase" columns. In addition, a family with a tax decrease may also be tabulated in the "presently taxable families made non-taxable" column and a family with a tax increase may also be tabulated in the "presently non-taxable families made taxable" column.

Unlike Table 5, Table 13 contains a column which presents current law tax liability for families with a tax increase or a tax decrease.

Table 14A provides a summary of the effects of the proposed changes (i.e., Plan Y) in the tax law. The results are distributed by size of family economic income and are disaggregated by family type. Table 14A together with Tables 14B, 13, and 12 provide a comprehensive aggregate picture of the revenue and distributional effects of the proposed tax law changes.

Table 14B presents a summary of most of the items that have already appeared in tables 12, 13, and 14A as well as some items that do not appear anywhere else in the tax model output. Consequently, this table duplicates Table 8B but is tabulated on the basis of families. The results in Table 14B are also distributed by size of family economic income.

FOOTNOTES

[1] Roy Wyscarver, *The Treasury Individual Income Tax Simulation Model*. Office of Tax Analysis, 1985.

[2] Internal Revenue Service, *Statistics of Income - 1981 Individual Income Tax Returns*. Government Printing Office, 1983.

[3] For a detailed discussion of the sample reduction methodology, see John Mulvey, "Multivariate Stratified Sampling by Optimization." *Management Science*, vol. 29, no. 6, 1983.

[4] J. S. Turner, and Richard Barr, "A New Linear Programming Approach to Microdata File Merging." *1978 Compendium of Tax Research*, Washington, D.C.: Government Printing Office, 1978.

[5] This problem is documented in. Joseph Kadane, "Some Statistical Problems to Merging Data Files." *1978 Compendium of Tax Research*, Washington, D.C.: Government Printing Office, 1978.

[6] The Joint Committee on Taxation uses Congressional Budget Office forecasts while OTA uses Office of Management and Budget forecasts.

[7] For a discussion of this and other revenue estimating issues, see Howard Nester, "Guide to Interpreting the Dynamic Elements of Revenue Estimates." Chapter 1 in this volume.

[8] A discussion of the distribution of individual income tax liabilities can be found in Susan Nelson, "Family Economic Income and Other Income Concepts Used in Analyzing Tax Reform." Chapter 3 in this volume.

[9] For a discussion of the Haig-Simons economic income concept, see R. Goode, "The Economic Definition of Income." In J. Pechman ed., *Comprehensive Income Taxation*. Brookings Institution, 1977. For a more detailed discussion of the economic income concept used by the Office of Tax Analysis, see Susan Nelson, "Family Economic Income and Other Concepts Used in Analyzing Tax Reform Proposals." Chapter 3 in this volume.

[10] U.S. Department of the Treasury, *Tax Reform for Fairness, Simplicity, and Economic Growth: The Treasury Department Report to the President, Volume 3: Value-Added Tax*, Washington, D.C.: Government Printing Office, 1984.

[11] Gerhard Paass, "Statistical Record Linkage Methodology: State of the Art and Future Prospects," a paper presented at the 45th meeting of the International Statistical Institute, Amsterdam, 1985.

REFERENCES

Goode, Richard. "The Economic Definition of Income." In J. Pechman ed., *Comprehensive Income Taxation*. Brookings Institution, 1977.

Internal Revenue Service, *Statistics of Income - 1981 Individual Income Tax Returns*. Government Printing Office, 1983.

Kadane, Joseph. "Some Statistical Problems to Merging Data files." *1978 Compendium of Tax Research*, Washington, D.C.: Government Printing Office, 1978.

Mulvey, John. "Multivariate Stratified Sampling by Optimization." *Management Science*, vol. 29, no. 6, 1983.

Nelson, Susan. "Family Economic Income and Other Income Concepts Used in Analyzing Tax Reform." Chapter 3 in this volume.

Nester, Howard. "Guide to Interpreting the Dynamic Elements of Revenue Estimates." Chapter 3 in this volume.

Paass, Gerhard. "Statistical Record Linkage Methodology: State of the Art and Future Prospects," a paper presented at the 45th meeting of the International Statistical Institute, Amsterdam, 1985.

Turner, J. S. and Barr, Richard. "A New Linear Programming Approach to Microdata File Merging." *1978 Compendium of Tax Research*, Washington, D.C.: Government Printing Office, 1978.

U.S. Treasury Department, *Tax Reform for Fairness, Simplicity, and Economic Growth: The Treasury Department Report to the President*. Washington, D.C.: Government Printing Office, 1984.

Wyscarver, Roy. *The Treasury Individual Income Tax Simulation Model*. Office of Tax Analysis, 1985.