SPSD/M

Introduction and Overview Guide

This guide is designed to give an overview and introduction to the SPSD/M. It describes the construction of an SPSM simulation run and contains several examples of model use. It should be read completely by anyone who will be using the SPSD/M for research purposes.

Statistics Canada   Statistique Canada

Canada

Table of Contents

Introduction

Welcome to the SPSD/M. This guide presents an introduction and overview of the product and a description of the basics required to use the model. **It should be read immediately after installing the SPSD/M and before you attempt to use the program for the first time.** Most of the material contained in this guide is covered in greater detail in the *User's Manual* and *Reference Manual*. What we do here is present you with an introduction sufficient to get you up and using the model quickly. Once familiar with the SPSD/M, users of the full package can restrict their use to the *User's Manual* and the *Reference Manual*.

This software is complex. The SPSD/M is a highly flexible tool with a vast number of independently controllable, interrelated parts. It is designed to be easy to use. With extensive use of default settings, the flexibility is controlled by first time users. However, implicit in these settings are numerous assumptions which could potentially affect interpretation of the results. Moreover, the institutions and programs modeled by the SPSD/M are themselves complex. **It is thus easy to make subtle mistakes which may result in significant estimation errors.** To make the best use of the SPSD/M, please read the available documentation. Then if you have any problems or you are unsure of the simulation results, please phone us at (613) 951-3774 or email us at spsdm@statcan.ca.

WHAT YOU SHOULD ALREADY KNOW

By now you should have the SPSD/M installed on your computer. If you do not, please do so now. The guide assumes that the user is already familiar with the use of Windows based applications. In addition, some knowledge of federal and provincial tax and government cash transfer programs as well as some familiarity with the data sources is essential. The SPSD/M is based on several disciplines including econometrics, statistics, computer science, mathematics, and microsimulation and training in any of these areas will be an asset.

While the SPSD/M can be used for analysis on its own, other software products can compliment and enhance analysis using the SPSD/M. Spreadsheets are good tools for manipulating output from SPSM standard and custom tables. In addition, statistical software packages such as PC-SAS and SPSS are useful for subsequent analysis of simulation results.

USING THIS GUIDE

Certain portions of this guide should be read by everyone who will be using the SPSM for research purposes from junior programmers to senior analysts. If you are unfamiliar with the SPSD/M and never plan to use the program or analyze and interpret the results directly, you can confine your reading to this introductory chapter.

The SPSD/M is a complex static microsimulation model and must be understood thoroughly if one is to make correct inferences. The computer program itself, however, is simple to use and results can be produced very quickly with it. Overview of the SPSD/M presents the fundamental concepts which must be understood prior to conducting any research with the model. This includes a description of all the basic parts of the model and their relationships to one another. By reading this, the user will also become familiar with the specific terminology which is used throughout the documentation.

The section titled Software and Facilities describes the use and function of the SPSM user facilities. These facilities allow users to modify parameters interactively, screen or select records, produce default and custom tables and distribution reports, create new variables, and conduct marginal tax rate analyses.

After reading through the above sections, the user should proceed to the tutorial examples. The examples are meant to be followed in order. They provide general information on using the SPSM as well as details on specific SPSM features. Some of the SPSD/M user facilities are illustrated in these examples.

The final section entitled Guide to SPSD/M Documentation gives an overview of the contents of the other help guides included with the SPSD/M package and provides some guidelines on their effective use.

OVERVIEW OF THE SPSD/M

The Social Policy Simulation Database and Model (SPSD/M) is a tool designed to analyze the financial interactions of governments and individuals in Canada. It allows the assessment of the cost implications or income redistributive effects of changes in the personal taxation and cash transfer system.

There are four basic elements to the SPSD/M: a database (the SPSD), a model (the SPSM which is comprised of a set of simulation algorithms), data retrieval and reporting software, and user documentation.

1. The SPSD is a non-confidential, statistically representative database of individuals in their family contexts, with enough information on each individual to compute taxes paid to and cash transfers received from government.

2. The SPSM is a static accounting model which processes each individual and family on the SPSD, calculates taxes and transfers using algorithms that simulate legislated or proposed programs, and reports on the results. A sophisticated software environment gives the user a high degree of control over the inputs and outputs to the model and allows the user to modify existing programs or to examine proposals for entirely new programs. Within the SPSM there are two models, configured as two distinct computer programs.

2a. The central program, the SPSM, is a microsimulation model which calculates taxes and transfers for individuals and families. These calculations are performed for everyone on the SPSD and then aggregated to obtain estimates. The SPSM is a static impact model and is therefore not intended to simulate how an individual's behaviour is likely to change in response to various policy options. The SPSM program includes software which allows the user to perform tabulation and database retrieval operations on the database information.

2b. The commodity tax model (COMTAX), is a macro-economic Input/Output based model. This model is not currently released with the SPSD/M package but rather the results from it are. COMTAX provides estimates of effective federal and provincial retail sales-tax-equivalent commodity tax rates by province, household expenditure category, and tax type. This model is required because many commodity taxes are levied at intermediate stages of

production, not at the final retail stage. Rates calculated by the COMTAX model can be input as parameters into the SPSM to obtain estimates of the commodity taxes paid, both directly and indirectly, by any given household.

3. The data retrieval and reporting software are configured as facilities accessed through the model. They allow the user to produce formatted output and to perform specific kinds of analysis.

4. The user documentation is extensive and comprehensive. It is divided into three manuals with a number of guides in each.

There are also two ways of running the SPSM: using the Visual SPSM interface and using the Classic SPSM mode.

1. Visual SPSM: The Visual SPSM interface allows users to modify model parameters directly, run simulations using SPSM, and examine outputs.

2. Classic SPSM: The SPSM can also be run from the command interpreter (cmd).

CAPABILITIES AND FEATURES

The SPSD/M algorithms operate on a database created using survey and administrative data from 2003. The SPSD/M comes with the necessary algorithms and parameters to simulate the tax/transfer systems in that year. The SPSD/M also performs projections to all other years from 1991 through 2012. The following is a list of the major programs which can be simulated with the SPSM. Users can add their own algorithms to this set.

Personal Income Taxes
-1991 to 2012 taxation year algorithms
-Major deductions and tax credits
-Payroll taxes (CPP/QPP, UI/EI).
-Both Federal and Provincial tax algorithms.
Cash Transfers
-Unemployment Insurance/Employment Insurance benefits
-Old Age Security benefits including partial benefits
-Guaranteed income supplement benefits including partial benefits
-Provincial GIS supplement programs
-Federal Child Benefit
-GST Tax Credit
Commodity Tax Model
-Ten different tax types accounted for including:
    – Federal Import Duties
    – Provincial profits on liquor commissions
    – Federal Fuel Taxes
    – Provincial Fuel Taxes
    – Provincial Retail Sales Taxes

-Cascading effects between taxes
-Determination of household expenditures net of implicit commodity taxes

The SPSM is written in the C++ programming language. Changes to the tax/transfer model algorithms can be made in two distinct ways: the "black box" or the "glass box" mode. The majority of SPSM analyses operates in "black box" mode.

- When using the model in the **black box** mode, a user would specify changes to one or more of the over 800 tax/transfer parameters. In this mode the user can change levels and rates affecting various benefits, taxes, and eligibility requirements, and can also specify a sub-population that will be modeled.

- The **glass box** mode is designed for use by persons who wish to develop entirely new algorithms or to change existing algorithms. Glass box users can write a new (or modify an existing) algorithm in the C++ programming language, and then recompile the model. The new model may then be used in black box mode. The user need not be a C++ language expert in order to use glass box mode, but should be experienced with some high level computer language.

The SPSD/M may be projected to represent a variety of different years using a method of static aging. A set of **database adjustment algorithms** and alternate weight files are included with the model to this end. The algorithms allow the user to "grow" or project the value of the money items on the base year database to some future year. The user can specify a series of growth rates which are then applied by the model. In most cases a single scaling factor is applied. However, for certain important or complex variables more complicated rates are applied. For example, employment earnings are "grown" by rates specific to province and sex of the individual. These algorithms are used in conjunction with annual demographic weight files included as part of the database.

Users may add their own adjustment routines to augment or even replace the supplied adjustment algorithms. For example, if users wish to grow employment income at rates differing by age and sex, they could set the existing employment income scaling parameters to one and create a new income scaling algorithm.

SOFTWARE AND FACILITIES

The SPSD/M comes with a powerful software support environment designed to enhance the range, speed and ease of policy analysis in both black box and glass box modes. The comprehensive and flexible set of reporting options allow for detailed examination and verification of the database, algorithms and results. The software has been designed to work in an integrated fashion at different levels of analysis (e.g. household, census family, individual).

The Visual SPSM program allows for people unfamiliar with the command environment to use the SPSM by providing a Windows based method to modify parameters, run the model, and examine the resulting output. A help facility provides information on the use of the model as well as a handy reference to parameter lists. Some highlights of the model are:

- The **user expression facility** allows the user to create, name, and access in subsequent runs, new variables that can be used for reporting purposes, without the need for any C++ language programming.

- The **record selection facility** allows the user to specify a database subset in terms of any combination of database or modeled variables. For example, a user could create a condition with an alternate definition of income and then select all economic families with less than a specified amount of that income.

- **Marginal tax rate analyses** can be performed in a single run of the model.

- **Two distinct tax/transfer systems can be run simultaneously.** The two systems can be compared using any combination of variables in the same run. Base and variant runs may be differentiated on the basis of parameter inputs, alternate algorithms (glass box), or both. Additionally, through the use of auxiliary input databases, users may compare the results of any number of different simulation scenarios in a single execution of the SPSM.

- A comprehensive set of **default tables** includes dollar amount and persons reporting for over 40 variables with breakdowns by province, income class, family type, and proportion above or below specified income levels by family size and type.

- A **cross tabulation facility** allows the user to specify custom N-dimensional tables using database, model, or user-defined variables. The user is given full control over output formats, precision, level of analysis, and labeling of the tables.

- **Output facilities** allow the user to create custom files in compressed binary format, or any of three ASCII file formats. These files are intended for subsequent analysis using the SPSM or other software packages and can contain any combination of database or modeled variables.

- **Performance** has been given high priority in the design of the SPSM but still varies depending on the hardware environment, the simulation request, and the types of output specified. The SPSM can be run in batch mode to allow a series of simulations to run unattended.

We have tried to make the documentation as clear and concise as possible, and the software has been extensively tested. Still errors are possible. If you encounter any problems with the software or documentation, contact Statistics Canada directly at (613) 951-3774.

Quick Start

The SPSM can be very simple to operate and can give fast results. A basic understanding of the SPSM structure is all that is required. It should be kept in mind, however, that the SPSM is also a highly complex tool. The multi-level database and numerous parameters and facilities give great flexibility but can lead to confusion if many simultaneous changes are attempted. For this reason, the SPSM makes extensive use of default parameter files and

built-in documentation features. The user can always begin from the 2003 defaults, alter various settings, and examine the results.

The supplied SPSD/M 2003 default parameter files are configured to simulate the 2003 legislated tax/transfer system while producing two default tables. After first briefly examining the SPSM structure, this section will describe how to make a program run using these default files. The section concludes with an examination of the output and a general discussion of how to view your results.

SPSM STRUCTURE

As already mentioned, the SPSD/M models are configured as two distinct executable computer programs. The central program, SPSM, calculates personal income taxes, sales taxes, and government cash transfers. The other program, COMTAX, executes the macro-economic commodity tax model that provides parameters for input to the SPSM. The COMTAX program is not available and its execution is not discussed here. However, results from it are incorporated into parameter files in the form of effective sales tax rates. Default SPSM files apply these parameters to household expenditures to estimate individual commodity taxes.

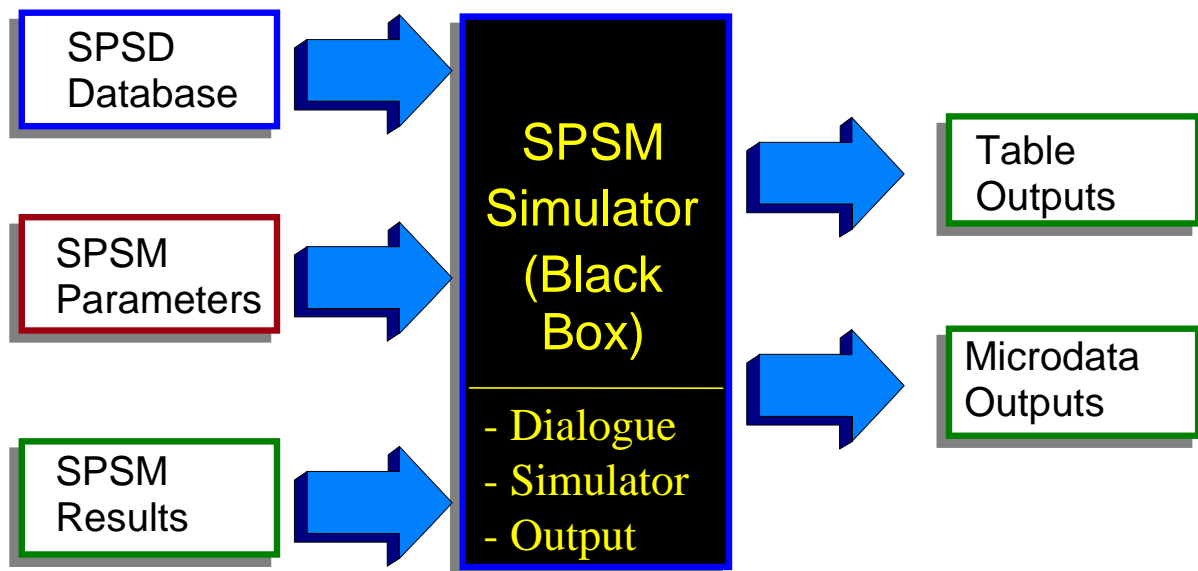Any execution of the SPSM program has the following basic structure.

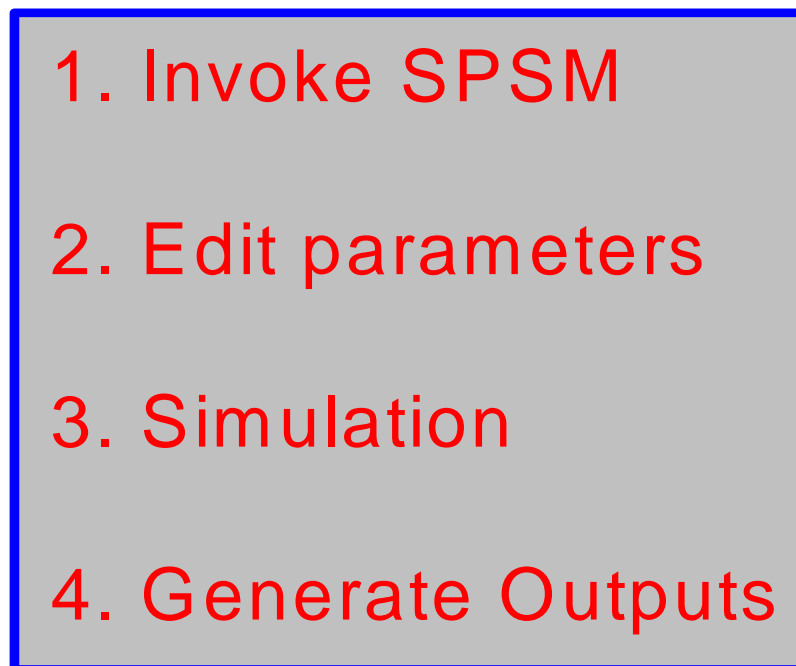Figure 2. SPSM Program Run Structure

Microdata on individuals and households as well as parameters are specified inputs to the SPSM. (Optionally, results from a previous SPSM run may also be input into subsequent runs). The SPSM then performs the simulation, calculates results, and outputs reports and data files.

The SPSM is driven by over 800 parameters that control three main processes. **Control parameters** specify input and output files for a specific model run and are used to activate model software facilities. **Database adjustment parameters** control the growth of database money items. **Tax/transfer parameters** control the specific functioning of the tax and transfer programs.

The SPSM creates a set of output files that the user has specified. Some of these outputs can be seen using the Visual SPSM interface. Others contain microdata for further analysis using the SPSM or some other software package.

The SPSM can be run in two different ways. Users can choose to use Visual SPSM to access the program. This will be the method described in this guide, as it does not require knowledge of the command interpreter environment. Alternatively, the classic method of using SPSM is still available. See *How to Run the SPSM* for details on using the classic method. This guide assumes that all new users will be using the visual interface.

Any SPSM program run consists of the following four basic steps:

1. Invoke SPSM

2. Edit parameters

3. Simulation

4. Generate Outputs

Brief descriptions of these steps are given below.

1. To invoke the SPSM, go to Start, Programs, SPSM 15.0, Visual SPSM. This invokes the visual interface which will be used to edit parameters, run the simulation, and examine outputs.

2. The parameters can be edited using Visual SPSM. The parameters are divided into four groups, control parameters, database adjustment parameters, variant tax and transfer parameters, and base tax and transfer parameters.

3. Once parameters have been changed, the simulation can be started. During this phase, the database is read, records are screened, and tax/transfer calculations are made based on the specified parameter settings.

4. When the simulation is complete, SPSM generates the specified output files. Though SPSM results are always written to files, most output can also be displayed in the interface.

RUNNING THE SIMULATION

We are now ready to try the first simulation.

The SPSM is now invoked by clicking on Start, Programs, SPSM 15.0, Visual SPSM. You'll see a screen with information about Visual SPSM and a link to the licence agreement. Click on **Close** to get the following screen.:
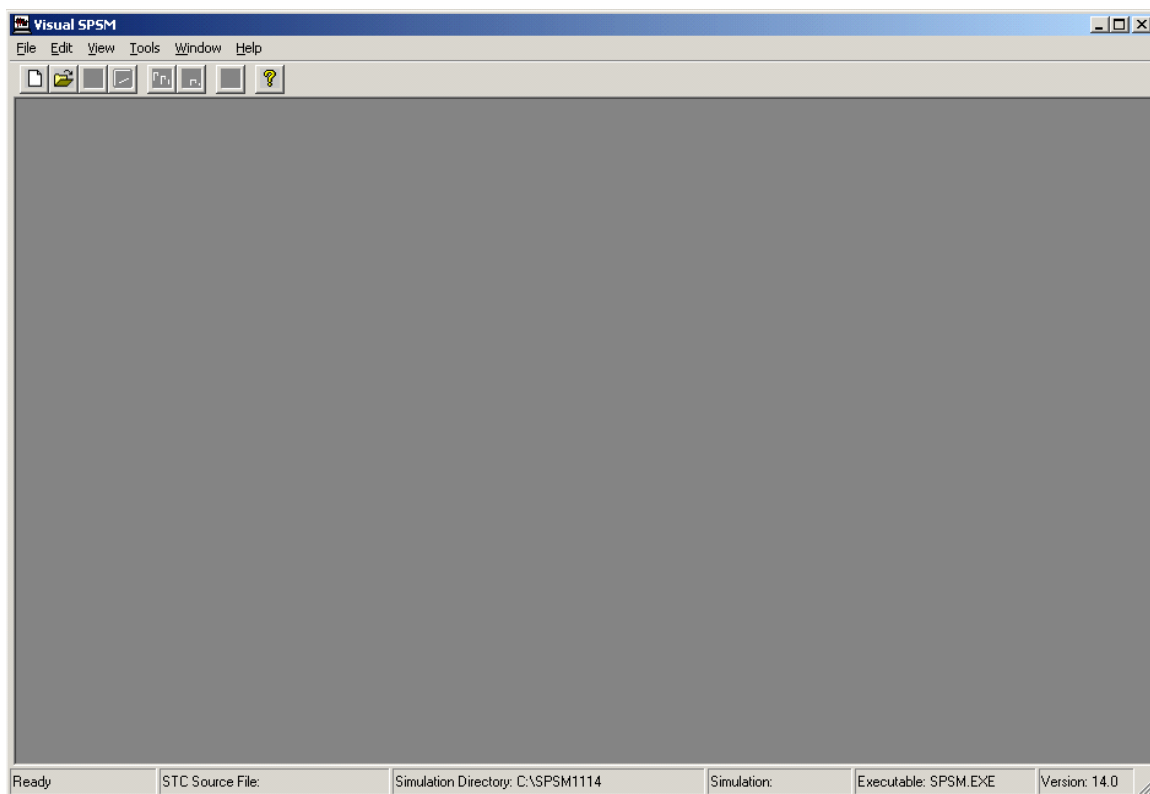
Open a new simulation by clicking **File**, and then click **New Simulation.** A box will show up which gives the possible starting points. Choose the base year, **ba02.cpr** and the click **OK.**
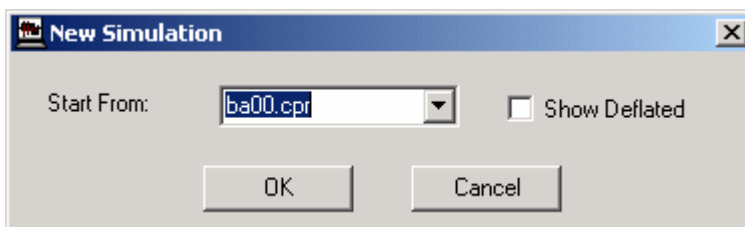
You now get the following interface which allows you edit parameters, if wanted.
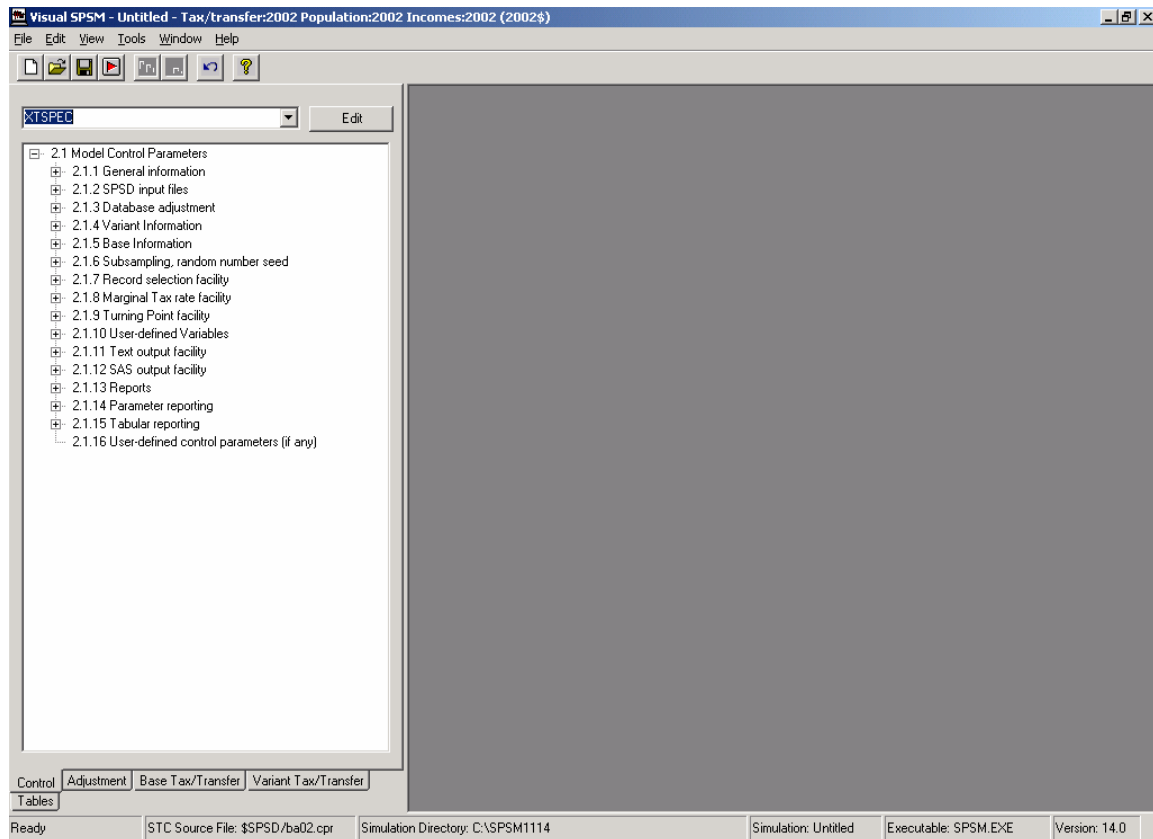
We will run the SPSM using the default parameters. So click on **File**, **Save Simulation** and save the simulation somewhere. We strongly suggest that you save all your simulations in appropriately named subdirectories. Then click on **File, Execute Simulation.** Once you have saved the simulation, the Visual SPSM will call the spsm executable and run the model. A simulation status box will appear at the end of the execution.
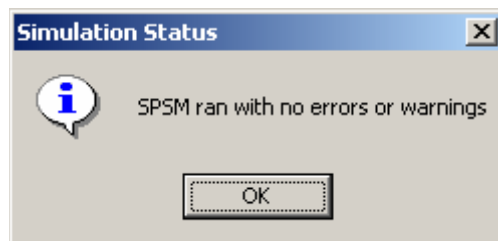
If the simulation contains errors, the status box will inform you and a log window will open.

EXAMINING THE LOG WINDOW

Though you can use Visual SPSM to modify parameters, run the simulation, and examine output, the model itself still invokes spsm.exe to do the simulation. The log window shows how changes made in the Visual SPSM interface get converted to the input needed for spsm.exe. This section will go through the log window to explain its contents. To open the window after a successful simulation, go to **View, Log**. If the simulation was unsuccessful

the window will open automatically.

The information that we input using the Visual SPSM is converted into the requisite steps for the classic SPSM.

In [1], you can see the starting point for the simulation. In our case, the undeflated 2002 tax and transfer system.

In [2], you see the new name of the simulation. In this case, Example1.

```
                    Welcome to the Statistics Canada

                Social Policy Simulation Database / Model
                              ( SPSD / M )
                   Program Version              : 14.0
                 Database Adjustment Algorithm   : Standard
              Standard Tax/Transfer Algorithm : Version 14.0: 91-10
                 Alternate Tax/Transfer Algorithm: none/aucun
               Compiler / Operating System      : MSC 13.10 / WIN/NT

                 Copyright (c) Minister of Industry 1984-2005

        Press 'F1' for context-specific help.
        Press Ctrl-Break to terminate.
=================== Control Parameters =======================
[1]     Enter name of input control parameter file ==> \spsd\ba02
        Loading control parameters from '\spsd\ba02.cpr'

[2]     Enter specification for generating output files ==> Example1
        Generating default output file names.
```

The SPSM next displays some of the key content of the default control parameter file

```
        Current values of selected control parameters:
        Descriptive
          CPRDESC    Tax/transfer:2002 Population:2002 Incomes:2002 (2002$)
          LICENSEE   Statistics Canada  # SPSD/M licensee
        Selection/Sample
          SELFLAG          0            # Selection facility activation flag
          SAMPLEREQ   1.000000000       # Size of sample requested
        Input Databases
          INPSPD     $SPSD/v140y02.spd # Name of SPSD file (in)
          INPFXV     $SPSD/v140y02.fxv # Name of SHS vector file (in)
          INPWGT     $SPSD/v140y02.wgt # Name of weight file (in)
        Growth
          INPAPR     $SPSD/ba02_02.apr  # Name of database adjustment parameter file
        Scenario Control <no base, variant computed using parms>
          BASMETH          0            # Method of creating base variables
          VARMETH          2            # Method of creating variant variables
          INPVARMPR  $SPSD/ba02.mpr     # Name of variant tax/transfer parameter file
         Output Files
         OUTMRSFLAG       0             # Variant results file creation flag
         OUTCPR     Example1.cpr        # Name of control parameter file (out)
```

Next, if we had made changes to any control parameter files, these would be read in [3].

Since we made no such changes, the defaults are kept. The step [4] is never used when running the model using Visual SPSM, it is there for people who use the model using Classic SPSM.

```
[3]    Do you wish to modify any control parameters ? ==> n
       No files will be overwritten with these control parameters
[4]    Any further control parameter changes ? ==> n
       Writing out control parameters to 'Example1.cpr'
```

The next step is editing database adjustment parameters. A few of the key parameters are displayed. In our example, all growth factors have been set to 1 since this is the base year, so effectively the database is not altered. If we had made changes to any control parameter files, these would be read in [5].

```
:================ Database Adjustment Parameters =================
 Loading database adjustment parameters from '$SPSD/ba02_02.apr'
 Current values of selected database adjustment parameters:
  GFISENF        10        # Growth Factor: Self-employed income -non
    0.65990
    0.79030
    0.95680
    0.71600
    0.86120
    0.75230
    0.78490
    0.72210
    0.73510
    0.79560
  GFIEMP         10                      # Growth Factor: Employment income
    0.96060  0.96060
    0.94740  0.94740
    0.97410  0.97410
    0.91340  0.91340
    0.93240  0.93240
    0.91260  0.91260
    0.90560  0.90560
    0.84730  0.84730
    0.93850  0.93850
    0.92600  0.92600
```

```
[5]    )Do you wish to modify any database adjustment parameters ? ==> n
```

The next step is editing the variant tax and transfer parameters. Once again, a few key parameters are shown. If we had made changes to any control parameter files, these would be read in [6].

```
==================== Tax/Transfer Parameters ====================
        Loading variant tax/transfer parameters from '$SPSD/ba02.mpr'
        Current values of selected variant tax/transfer parameters:
        MPRDESC       Current values for 2002
        CTFLAG              1              # Commodity tax activation flag
        CTOPT               2              # Commodity tax calculation method
[6]     TARGETYEAR       2002             # Year of analysis
        Do you wish to modify any variant tax/transfer parameters ? ==> n


        ================================= Execution =================================

        SPSM          started   on Mon Apr 05 11:27:56 2004
[7]     SPSM          completed on Mon Apr 05 11:27:57 2004
        Elapsed time= 00:00:01
        Number of households processed=  3999
        Number of persons       selected= 11483
        Variant consumable income ($000,000):   336942.4
        Control parameter file . . . . . . . . Example1.cpr
        Table output file  . . . . . . . . . . Example1.tbl
```

The simulation is then run [7].

EXAMINING OUTPUT

Though all output from the SPSM is written to files, the tables it produces can also be viewed from Visual SPSM. Click on the **Tables** tab on the bottom left of the tree.
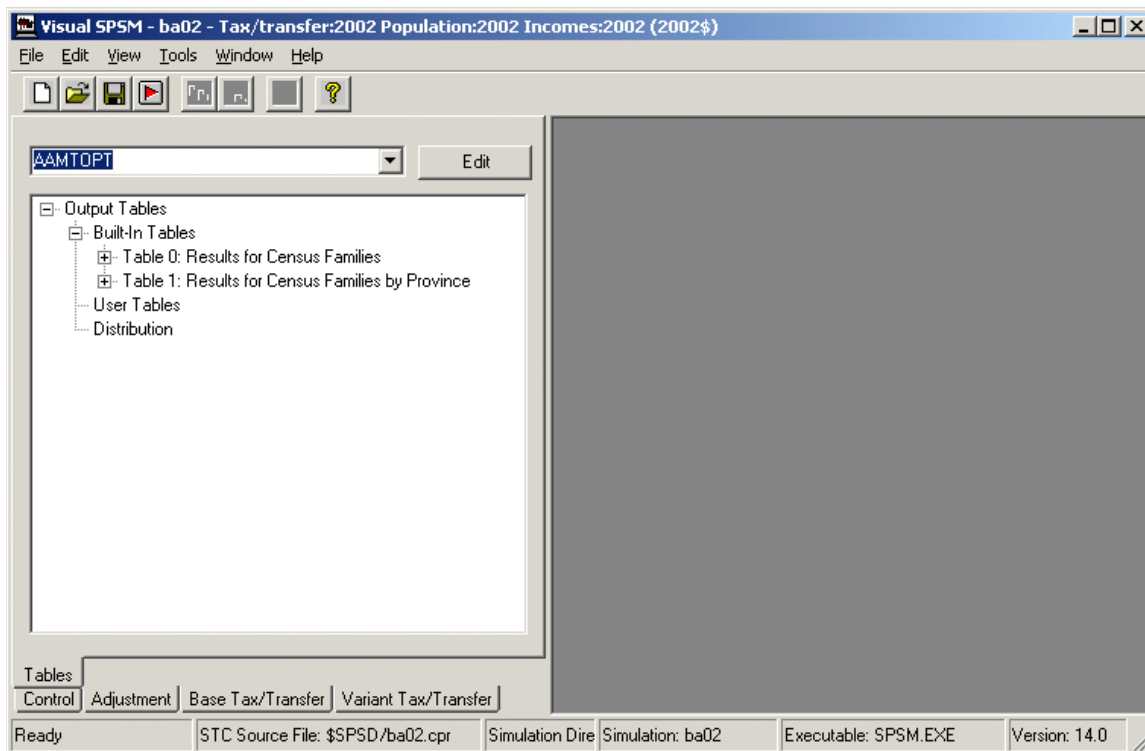
Two of the ten possible default tables were created, Table 0 and Table 1. The format and contents of these tables are fixed. Users may not change either their rows or their columns of the table. Table 0 is a table for all Canada while Table 1 is a table by province. The next snapshot shows the first rows of Table 0.

| Visual SPSM - ba02 - Tax/transfer:2002 Population:2002 Incomes:2002 (2002$) |

SPSD/M (Database 14.00)   Fri Jan 13 14:19:20 2006
Base Description: No base results
Variant Description: Current values for 2002   [Driver: Version 14.0 : 91-10, File: ba02_V.mpr]
Sample: 1.0000 AGENAME='Standard'

Results for Census Families

Variable (x1,000,000)	TOTAL
Family Units (x1000)	14810.1
Persons (x1000)	30972.0
SLID Survey Records	35407.0
SPSD Records	94838.0
Income (Base)	538155.4
Income (Variant)	538155.4
Change	0.0
Number of Gainers (x1000)	0.0
Number of Losers (x1000)	0.0
No Change (x1000)	14810.1
Gainer's Gain	0.0
Loser's Loss	0.0
Total Income	769421.3
Market Income	676419.2
Wages and Salaries	518669.4
Self-Employment Income	41452.8
Investment Income	46588.3
Other Income	69708.8
Transfer Income	93002.1
Total Tax	231685.4
Net Transfers	-138683.2
Disposable Income	606206.4
Consumable Income	538155.4
Federal Taxes	146397.4
Federal Income Tax	86255.7
UI/EI Contributions	8175.6
UI/EI Benefit Recovery	181.9
CPP/QPP Contributions	16939.1
Other Recoveries	0.0
Federal Commodity Taxes	34426.0

Ready   STC Source File: $SPSD/ba02.cpr   Simulation Directory: C:\SPSM1114   Simulation: ba02   Executable: SPSM.EXE   Version: 14.0

Understanding Basic Components

Having performed a complete SPSM run, we leave aside for the moment the specifics of performing a program run and turn our attention to the several key components that form the basis of the model and any given simulation.

At a very general level, the SPSM accepts as input a set of database variables and applies various algorithms to them which in turn produce a set of simulated variables. The database and simulated variables are then output in different forms for analysis. The entire process is controlled by parameters. Any complete simulation involves all of the following four components.

- Database
- Parameters
- Simulation
- Outputs

Each of these four components must be understood on its own as well as in terms of its relationship to the other components as shown broadly below.
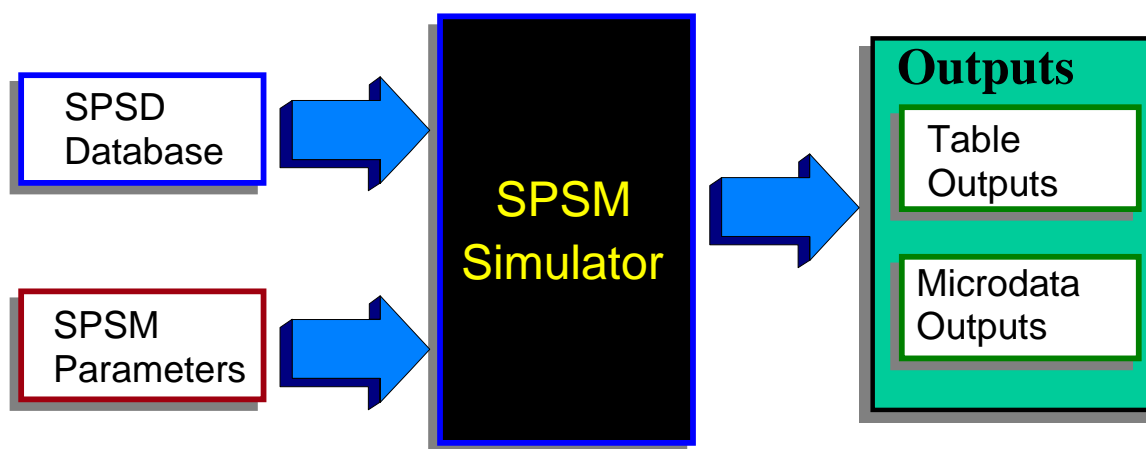
This section describes the function and most important features of each of these four components. It also describes the ways that these components interact with each other.

The SPSD database forms the foundation for the entire SPSD/M and is discussed first in terms of its scope, organization, structure, and variables. The simulator performs all operations on the database including adjustment and screening, simulation of modeled variables, and reporting results. The parameters that control this process are discussed next. This is followed by a general discussion of the outputs available from the SPSM. The section concludes with an examination of the SPSD/M files and their relationship to the components just discussed.

DATABASE

The SPSD was constructed by combining individual administrative data from personal income tax returns and unemployment insurance claimant histories with survey data on family incomes and on expenditure patterns. The techniques used to create the database and avoid confidential data disclosure include various forms of categorical matching and stochastic imputation. While no one record on the database contains information for the same individual from the four bases, the database has been constructed in such a way as to provide a micro-statistically representative sample of Canadians.

The SPSD can be used to represent the population of Canada commencing in 1991. This is done through the use of different sets of household weights that embody demographic data and projections by age, sex and province for years beyond the database year.

Certain adjustments have been made in SPSD to force agreement between data and known control totals. For example, survey weights have been adjusted to ensure that the population by age and sex corresponds to census estimates, and the number of high income Canadians corresponds to the number reported by Canada Revenue Agency. Further adjustments have been made to compensate for item non-response in the surveys (e.g. to increase the number of families receiving social assistance). It should be noted that the SPSD coverage does not extend to the Yukon, Nunavut or Northwest Territories, persons residing on reservations, or

armed forces personnel residing in barracks.

The richness and complexity of the database permit comprehensive simulations but require some degree of familiarity if the data are to be processed correctly and valid conclusions drawn. The fundamental family structure of the data as well as the organization and range of variables and their sources are briefly presented here. For a full discussion of the variables and database consult both the *Variable Guide* and the *Database Creation Guide* in the *Reference Manual.* A listing and brief description of the six major types of database variables follows.

Demographic Variables

Demographic variables include age, sex, province, and family structure. A number of other classification variables, such as industry, occupation, educational status, labour force characteristics and housing tenure are also present. These variables are taken from the Survey of Labour and Income Dynamics (SLID) of about 40,000 households.

Income Variables

A number of variables giving individual income (for persons aged 15 years or older) by source are present. Sources include employment income, self-employment income, dividends, interest, and capital gains. These variables are drawn from the SLID except for high income individuals where they are derived from personal income tax information.

UI/EI (Unemployment Insurance/Employment Insurance) Variables

Variables providing some detail on the structure of up to two UI/EI claims for each individual in receipt of UI on the SPSD are given. Included are data related to the start date of a claim, the type of claim, and weeks on UI in the various phases. Variables come from matching SLID EI claimants with a 10% random sample (about 400,000 records in the base year) of Human Resources Development Canada administrative data files on individuals in receipt of UI/EI.

Tax-related Variables

To complete a tax form, one requires certain individual non-income variables. These include RRSP contributions, tuition fees, medical expenses, and charitable donations. These variables are imputed from a stratified sample of Canada Revenue Agency T1 individual income tax returns (the same sample that underlies the Canada Revenue Agency's annual Greenbook publication).

Consumption Pattern Variables

Each household on the SPSD has an associated expenditure pattern, giving household expenditures by 48 distinct types of commodity. The Survey of Household Spending (SHS) is used to impute this information on expenditure patterns.

Household Weights

Household weights are stored on separate files. There is a series of weight files corresponding to population estimates for the years from 1991 to 2012.

A full list of the available variables can be found in the *Variable Guide*.

Level of Analysis

The tax/transfer system considers the family context of individuals in determining the value of many taxes and benefits. For example, child tax credits depend on the presence of children and the amount of family income. The SPSM processes households and reports results based on several different social units.

The SPSD is a hierarchical family file in which the family relationships among all the members of a household are known. The phrase "a level of analysis" refers to one of five hierarchically arranged types of "families". The individual is the smallest and a household is the largest. The SPSM five family levels of analysis are defined as follows:

Individual	Single person or record on the SPSD.
Nuclear family	A head, spouse if present, and never-married children under the age of 18 sharing the same dwelling.
Census Family	A head, spouse if present, and unmarried children under the age of 25 sharing the same dwelling.
Economic Family	A group of individuals living together who are all related by blood, marriage, or adoption and share the same dwelling.
Household	Any individuals or group of individuals who share the same dwelling.

Note that unattached individuals are considered to be special kinds of census and economic families in SPSD/M. This usage differs from most Statistics Canada publications.

The SPSD data are stored at either the individual or the household levels. For example, age and employment income are individual-level variables while province and shelter expenditures are household-level variables. The data from the individual level are aggregated to another higher specified level for reporting purposes. For example, individual disposable incomes can be aggregated to census family disposable income. Some of the information from the household level can be "passed-down" for reporting at lower levels. For example, a table on individual capital gains by province would draw on the household level province variable. All modeled variables are assigned at the individual level.

The level of analysis must be specified for the production of default tables, user-defined cross-tabulations, and text output files. A level of analysis must also be specified for the selection facility, which causes results for certain user specified units of analysis to be excluded from processing and/or results. The facility for displaying the univariate distributions of variables also makes use of a level of analysis.

The default level of analysis varies depending on the application.

Variable Types

Variables in the SPSD/M can be grouped in two broad categories. Nominal or ordinal variables whose values denote membership in distinct categories are called 'class variables' in

this documentation. They may also be referred to as classificatory, categorical or classification variables. An example of a class variable is *idsex*, which records the sex of an individual. Variables which take numeric values on which arithmetic operations can be meaningfully performed are termed analysis variables. An example of an analysis variable is *idiemp*, which records the employment income of an individual.

The distinction between class and analysis variables is an important one. Certain of the SPSM facilities require that either a class variable or an analysis variable be specified in certain contexts. Class variables can be converted into analysis variables, or analysis variables to class variables using the SPSM's facilities to create user-defined variables.

A further important distinction is database versus modeled variables. Database variables, as described above, represent responses to survey or administrative reports and reside on disk in the SPSD file. Modeled variables are those variables which are calculated during the actual simulation run and may also be subsequently stored in result files.

Because of the many different types of variables used by the SPSM, extensive and strict naming conventions have been adopted. For example all database variables stored at the individual level of analysis begin with the *id* prefix and all individual modeled variables begin with an *im*. Further details on naming conventions can be found in the *Users Guide*.

SIMULATOR

The simulator is the second basic phase of an SPSM program run. Once the user has started the model and provided a complete set of parameters the simulator phase begins. The simulator phase operates in five sequential steps. Each of these steps is described briefly below.

Adjust Data
Calculate Base Scenario
Calculate Variant Scenario
Calculate User Variables
Selection and Output Routines

Simulator
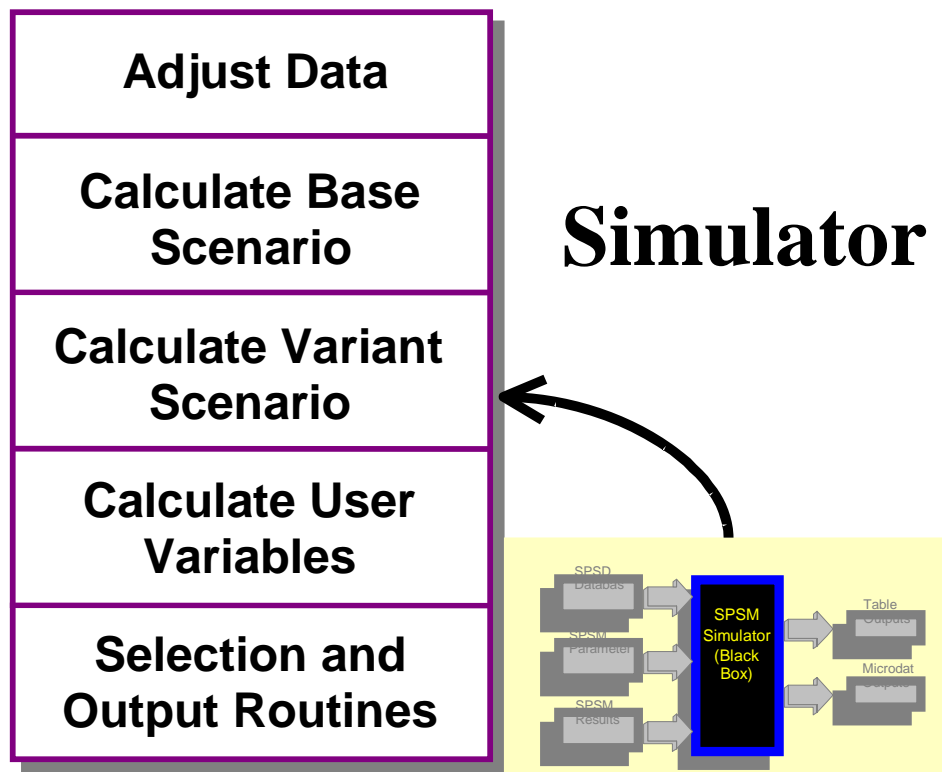
Database Adjustment

The SPSD is based on surveys and administrative data collected for the 2003 calendar year. Adjustment of this data may be required for several reasons. The user may wish the incomes to represent the incomes in a more recent year such as 2007. In this case the user would then supply estimates of growth rates for income by source between 2003 and 2007. The user may also wish to make adjustments for simulation purposes in the 2007 year. For example, known under-reporting of alcohol expenditures may be treated by scaling each household's alcohol expenses, or a deduction item could be scaled down to zero to represent a restriction of a certain tax measure. Finally the user may wish to adjust the weights of individuals and households to reflect alternative demographic assumptions.

Prior to the calculation of any taxes or transfers each dollar denominated database variable is multiplied by a single corresponding database adjustment parameter. There are two key exceptions to this. Employment income, comprising about 70% of total income, is given more rigorous treatment due to its relative importance in the overall tax/transfer system. Employment income is therefore adjusted by industry specific growth rates. CPP/QPP income is adjusted by differential rates according to age groups. This allows a simulation of the phasing in of that program. Users should refer to the *Parameter Guide* for further details on specific adjustment parameters.

**Calculate Base Results and Variant Results**

Microsimulation results are usually analyzed by comparing the results of one specified scenario with those of another. For example, if one was interested in examining the impact of increasing the maximum amount of Old Age Security (OAS) in 2003 from $5497.62 to $6000.00, two scenarios would have to be calculated. The first scenario would be the

tax/transfer system as currently legislated, let us call it Scenario 1. The second scenario has the exact same tax/transfer system with the exception that the maximum amount of Old Age Security is set to $6000. Let us call this second scenario Scenario 2. This could be done by performing two separate program runs and then comparing the outputs from the runs to determine the impact. However the SPSM is capable of performing the comparison in a single program execution. This is accomplished by applying the simulation algorithms to a given household twice in the same program run but with different parameter files. One such application is called the base and produces "base" results while the second application is called the variant and produces "variant" results. The first run, Scenario 1, would be calculated using the base set of parameters while Scenario 2 would be calculated using the variant parameters.

The results of any single simulation scenario are derived by a series of functions which sequentially calculate certain taxes and cash transfers as required by the tax/transfer system. For example, old age security benefits must be known before calculating net income. The sequence of these functions is controlled by a program called the driver (drv). The following figure represents the standard algorithm driver. Changes to the standard (or alternate) algorithm driver can be made in glass box mode.

The following figure shows the structure of the standard tax/transfer algorithm driver in more detail.
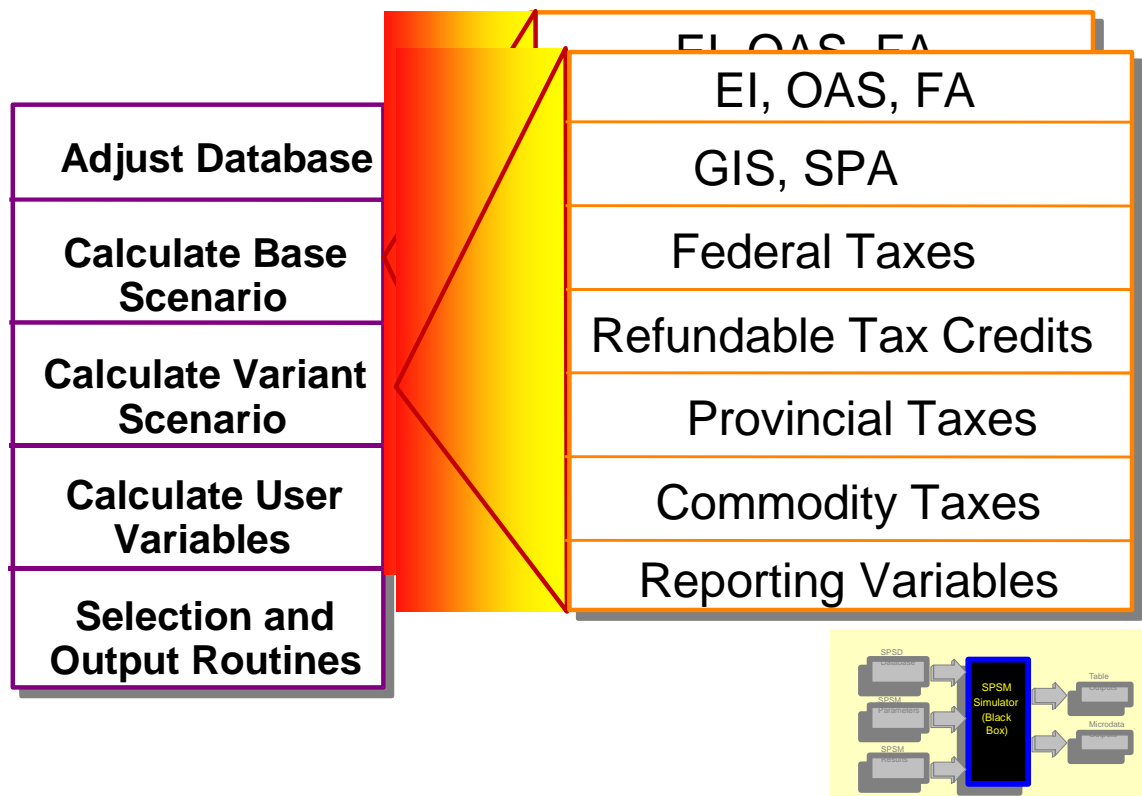
Tax/Transfer Algorithm Driver Structure

The first three functions simulate the Employment Insurance, Old Age Security and Family

Allowance programs whose benefits are determined by factors outside the tax system. This is followed by the calculation of net income. The Guaranteed Income Supplement, Spouses Allowances, and social assistance are calculated next based on net income. Federal taxes are then calculated. Then the federal refundable tax credits (child tax credit and sales tax credit) are calculated. This is followed by provincial taxes. Sales & excise taxes are calculated next by applying Input/Output based effective sales tax rates (calculated with COMTAX) to observed family expenditures and scaled to be consistent with simulated disposable income. Reporting variables for output are calculated as the last step of the driver algorithm.

Simulated variables, as opposed to database variables, are calculated either once or twice in the simulation phase. Base result variables, if requested, are calculated first, followed by variant variables. Base results, if generated at all, may be read from a file or calculated from the standard or alternate algorithm. Variant results, if generated, must be calculated from the standard or alternate algorithm (alternate algorithms may only be created in glass box mode). Base and variant results have independently specified tax/transfer algorithms and parameters. Distinct sets of result variables are maintained for both the base and variant so that the user may compare the two scenarios in terms of any combination of variables at any level of analysis.

User Variables

Although there are hundreds of variables available in the SPSD/M users will often wish to create their own variables for reporting purposes. This is done by manipulating the database and modeled variables. The user may create classification variables by specifying a continuous or discrete variable together with a set of breakpoints which define the class boundaries. Analysis variables may also be created by specifying an expression which manipulates logically and/or algebraically any of the database or modeled variables. User variables may be given names and labels as well as level labels for classification variables.

Screening Cases and Output Phase

Users may wish to restrict the calculation of results to a certain subset of households or to a certain type of individual or family on the database. This can serve both efficiency and analytical purposes. This screening can be based on modeled as well as database variables. Therefore it follows the calculation of base and variant results. For example, households having no children may be selected or individuals with no employment income excluded. Screening affects all of the SPSM output facilities in any given program run.

The output phase is the last step of the simulator and is designed to produce a series of outputs for different purposes. Output is either intended to be looked at or printed on the one hand or to be input to another computer program on the other. There are two main ways of producing printable aggregate tables, either by using one of the five supplied default tables or by creating your own customized tables. (See the user facilities section below for further details.)

Certain techniques useful in quantitative analysis (such as regression analysis or the calculation of gini coefficients) are not incorporated into the SPSM program. The SPSD database files are stored in a compact binary format which is interpretable only by the SPSM

programs. The output facility provides users the ability to create files containing subsets of the SPSD cases or variables in several different formats for subsequent data analysis using the SPSM or other software products. These formats can be read by the SPSM, SAS, spreadsheet programs, and other analytical software. Some output options also are suited to visual interpretation.

There are a large number of variables potentially available to the output routines. The diagram below indicates the broad types of variables and their relation to the SPSM simulator. The starting point is on the left with all database variables. Next, there are two possible sets of modeled variables, one for the base case and one for the variant. User variables are calculated next as a function of any of the variables created so far. Finally reference variables from previous model runs may be read and processed by the output routines.
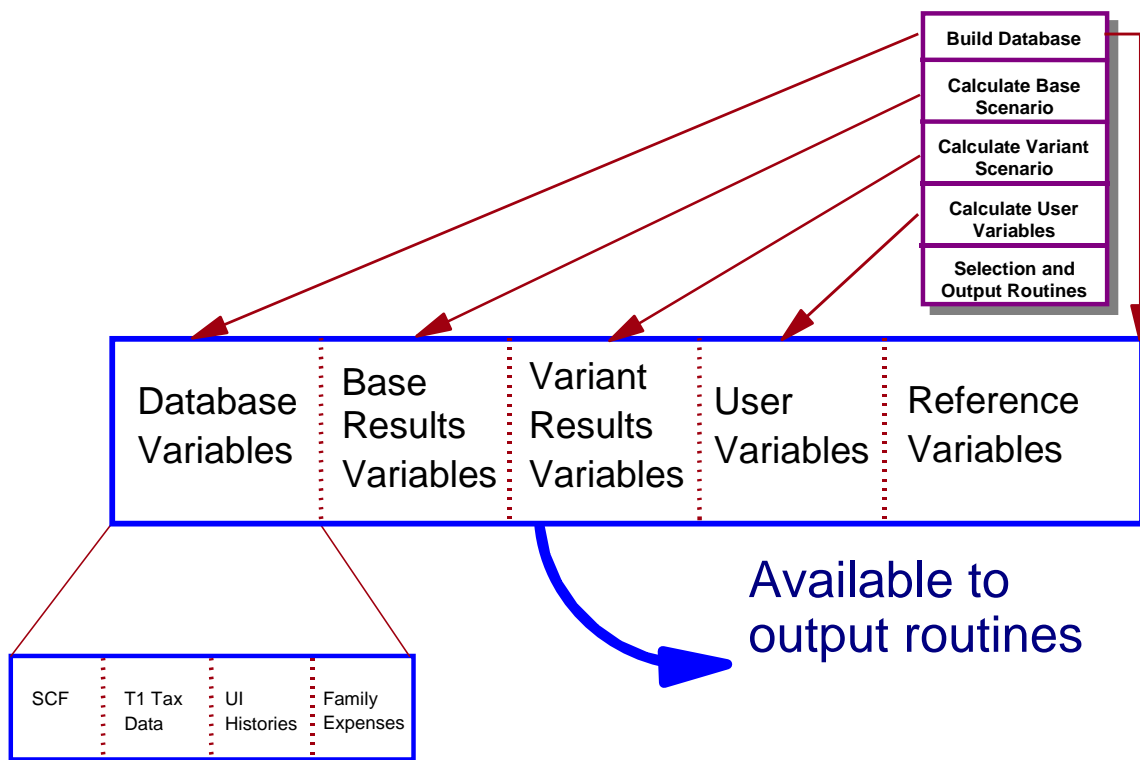
PARAMETERS

Parameters are used by the SPSM to control every aspect of a program run.
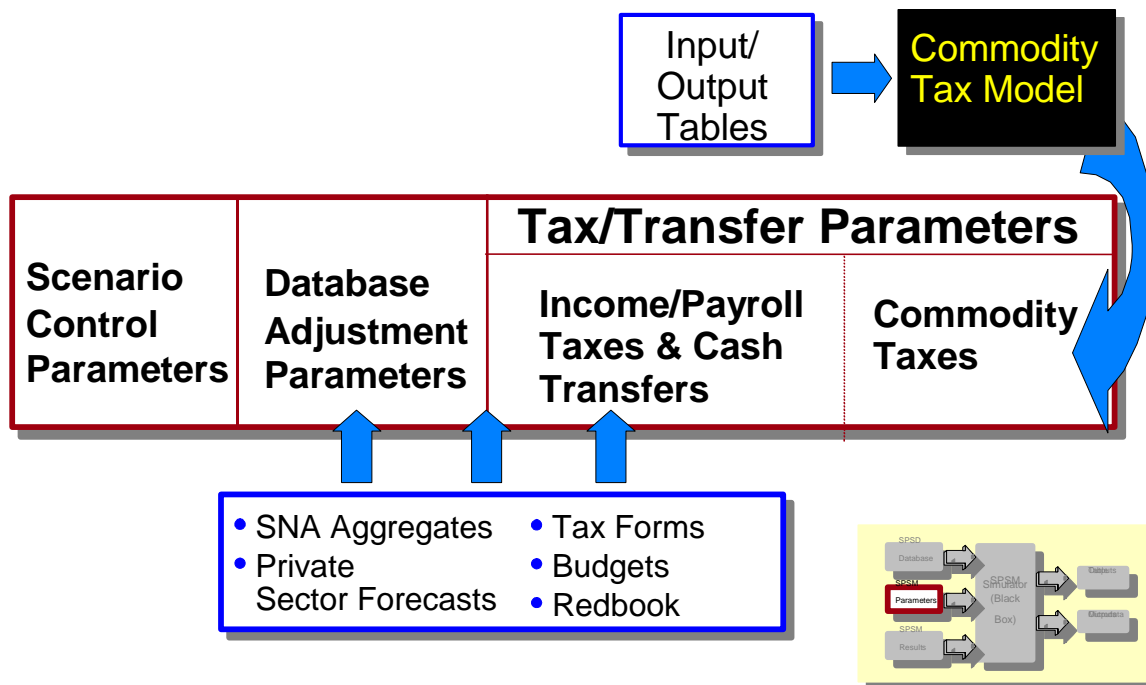
SPSM Parameter Types

As mentioned earlier, the SPSM is driven by over 800 parameters divided into three groups, corresponding to three types of files that control three main processes. **Control parameters** specify input and output files for a specific model run and are used to activate model software facilities. **Database adjustment** parameters control the inflation of database money items. **Tax/transfer** parameters control the specific functioning of the tax and transfer programs. In the case of commodity taxes, a set of default parameters is supplied.

A complete list of parameters organized by subject area can be referenced in the *Parameter Guide*. SPSM parameters provide instructions to the simulator in the following way.

| Scenario Control | Database Adjustment | **Tax/Transfer Parameters** | |
| | | Income/Payroll Taxes & Cash Transfers | Commodity Taxes |

- Get Data
- Calculate Base Scenario
- Calculate Variant Scenario
- Calculate User Variables
- Selection and Output Routines
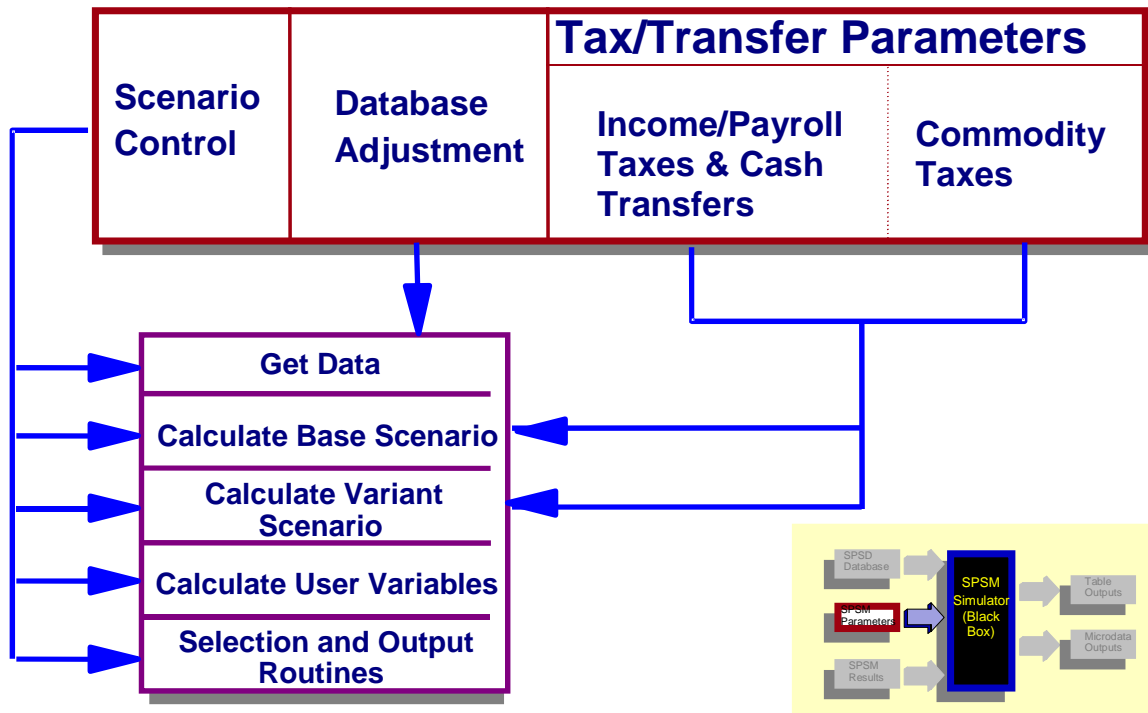
The Relationship of Parameters to the Simulator

Note that control parameters affect every step of the simulation whereas database adjustment parameters are used only once. Two sets of tax/transfer algorithm parameters are specified, one for calculating base results, and one for calculating variant results. In the Visual SPSM interface, the parameters are split into different tabs, as can be seen in the following figure.
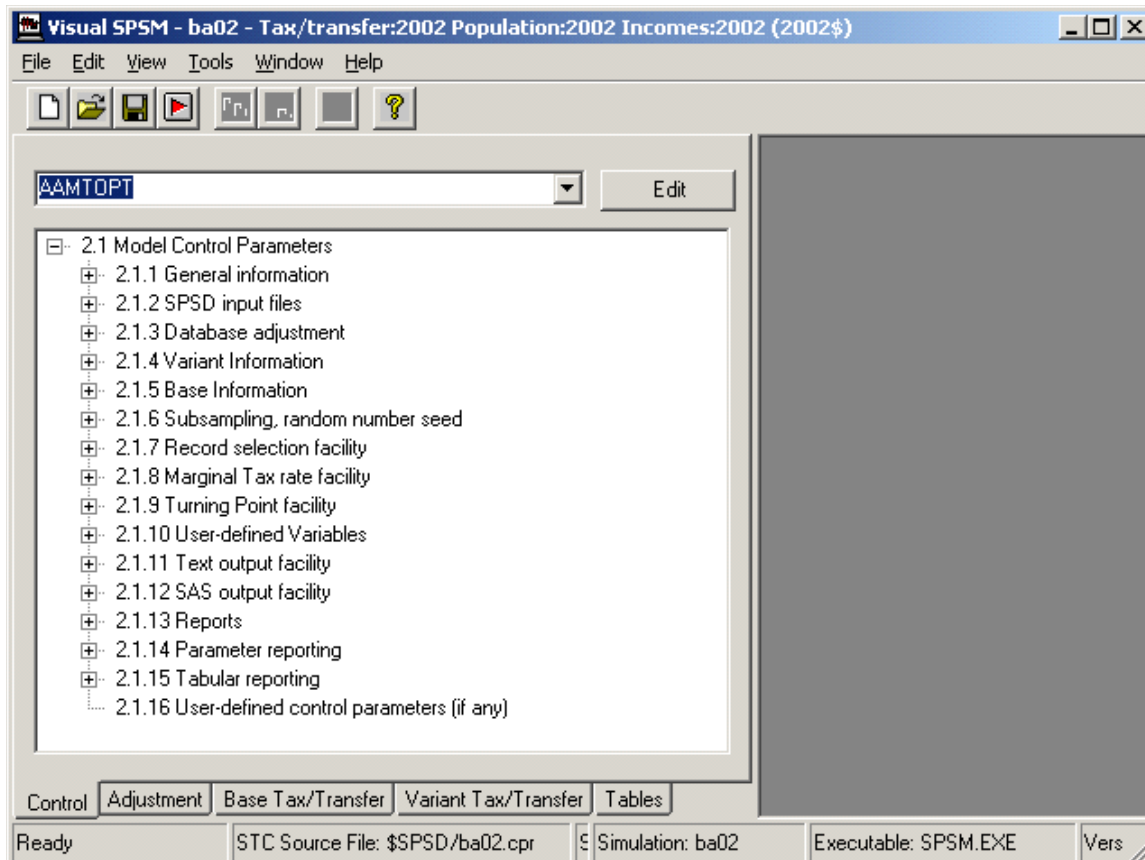
Each of the three types of parameter files are now discussed in turn.

Control Parameters

A control parameter file contains specifications for parameters used to control SPSM general operations. It gives information for such things as input and output file specifications, record selection requests, cross-tabulation and other user facilities. It does not give information on the parameters used to specify database growth, or parameters of the tax/transfer system. This information is provided by separate parameter files. The majority of control parameters are used to control the SPSM user facilities. The control parameter file also acts as documentation for a program run because it contains all the information necessary to make the run. Control files normally have a .cpr extension.

Database Adjustment Parameters

Database adjustment parameters allow the values of variables on the database to be changed in order to represent dollar values in calendar years other than 2003. For example, the dividend income of each individual on the database is multiplied by the database adjustment parameter GFIDIV. If in 2005 the value for Ontario is 1.201, all dividend income is 20.1 percent higher. Some of the database adjustment parameters are more structural in nature. For example, the parameter PTF specifies the low income cutoff thresholds for the default tables, and the parameter UITARGYRMAX establishes maximum insurable earnings levels for unemployment insurance. If any database adjustment parameters are modified from their initial values in the database adjustment parameter files during the user dialogue, then a

modified database adjustment file (with file extension of '.apr') will be generated.

The database adjustment parameters supplied with the full SPSD/M are based on differences between SPSD per capita base year (2003) figures and per capita national accounts data for available years. For years where projections must be made a simple extrapolation of the growth rate over the past two years is performed. Users may provide their own estimates if they wish to supply alternate growth rates.

Tax/Transfer Parameters

The final set of parameters is for controlling the algorithms which simulate taxes paid to and cash transfers received from government. Users provide values for the appropriate year of analysis or the particular tax/transfer scenario of interest. These parameters can switch entire programs on or off as well as specify benefit levels, reduction rates, turning points, tax tables, and so forth. If, for example, a user wished to examine the impact of increasing the maximum Old Age Security benefit to $6000, he or she would change the value of BOAS.

OUTPUT FACILITY

The SPSM writes all requested simulation results to output files. There are three main purposes for which the SPSM generates outputs.

- To document parameter settings which produce a program run
- To provide case by case information on simulation results
- To provide tabular aggregate information on simulation results

Outputs which document the program run are stored in the various specified parameter files. These parameter files are output in a self-documenting format, with parameters labelled and arranged in sections. These files may be viewed on the screen, or printed if desired. An output control parameter file is always produced. It contains, among other things, input and output files used in the run. The run can thus be duplicated using the output control parameter file in a subsequent run. Certain of the control parameters are informational and are not directly modifiable by the user. An example is ALGDESC, that displays which version of the tax/transfer algorithm was used for the standard and alternate tax/transfer algorithms. If the user has modified database adjustment parameters or tax/transfer parameters in the user dialogue an output parameter file of the appropriate type, including the specified changes, is produced.

Case by case results are useful for a number of reasons. They can be used as input to other software programs. If, for example, a user wished to perform a regression analysis on SPSD/M variables he or she would create a dataset containing only the variables of interest for each household (or any other level of analysis). That file would be read and processed by other analytical software with regression and sorting capabilities such as SAS, SPSS, GAUSS, etc. Case by case results can also be used as input to the SPSM to provide base result variable values. Finally, case by case output can be printed or viewed on the screen for debugging or other analytical purposes. For example, if a user wished to examine in great detail the five households who had lost the most disposable income in a comparative simulation, he or she would first isolate the household sequence numbers of the top five losers using the distributional analysis facility. A second run which would select only the

five households of concern, and print out all variables for those households could then be submitted.

Aggregate results take three distinct forms; built-in tables whose format is fixed by SPSM, user-specified cross-tabulations, or the report produced by the distributional analysis facility. Tables can be printed, analyzed directly, or used as input in other software packages. Spreadsheet packages can be particularly useful for manipulating the results from aggregate tables. Please see the *Tools User's Guide* for more information on the spreadsheet interface facilities supplied with the full SPSD/M.

DISPOSABLE AND CONSUMABLE INCOME

An important distinction to keep in mind when using the SPSD/M is the difference between disposable and consumable income. The analysis of federal and provincial taxes paid by individuals and families should include commodity and sales taxes, especially as these types of taxes play a larger role in the overall tax system. The tax\transfer parameters supplied with the SPSD/M support analyses which are concerned with both income and sales tax.

Disposable income is defined in the SPSD/M as total income minus total federal and provincial income taxes. It therefore represents the amount of income an individual or family has available for spending (e.g. shelter, food, savings). Consumable income, on the other hand, is defined as disposable income less commodity taxes embodied in household consumption. As such, it is not a post all tax concept because not all taxes are accounted for when determining consumable income. For example, municipal property taxes, and provincial land transfer taxes are not included in the calculation.

The SPSD/M built-in tables and standard reporting facilities define gainers and losers in terms of consumable income. If users wish to express their analyses in terms of disposable income two approaches are possible. First, the user can disable the calculation of commodity taxes and cause consumable income to be set equal to disposable income. Alternatively the user can specify custom tables (using the X-tab facility) which are defined in terms of disposable income.

SPSM FILES

The SPSD/M uses a great number of different types of files. Database files and supplied parameter files are found in the \spsd directory. The use of directory structures and strict naming conventions have been used in order to avoid confusion.

The following table gives the file extension naming conventions used for SPSM input and output files. Users of Visual SPSM may never see most of these files.

Database Inputs

	`.SPD`	SPSD binary database file
	`.FXV`	SHS household level expenditure data
	`.WGT`	SPSD file containing survey weights

Parameter Inputs
Complete and mandatory

	`.CPR`	Control parameter file
	`.APR`	Database adjustment parameter File
	`.MPR`	Tax/transfer parameter file

Partial and optional, created by user

	`.CPI`	Control parameter include file
	`.API`	Database adjustment parameter include file
	`.MPI`	Tax/transfer parameter include file

Partial, created by Visual SPSM

	`.CPD`	Control parameter difference file
	`.APD`	Database adjustment parameter difference file
	`.MPD`	Tax/transfer parameter difference file

SPSM Outputs
Printable

	`.TBL`	Output tables result file
	`.PRN`	Case output facility text results file
Input to further analysis	`.MRS`	Binary SPSM results file

Program Run Control

	`.BAT`	SPSM Dialogue commands
	`.CMD`	SPSM Dialogue commands
	`.CTL`	SPSM Dialogue commands

Visual SPSM state

	`.STA`	Visual SPSM state

SPSD/M File Naming Conventions

The database files are all named by Statistics Canada and should not be changed. There are many types of input files. The parameter files with `.CPR`, `.APR` and `.MPR` must be complete and valid SPSM parameter files and values for all three must be specified for any given program run. The `.CPI`, `.API`, and `.MPI` files are created by the user with a text editor and contain subsets of the corresponding `.CPR`, `.APR` and `.MPR` files which may be read in during the user dialogue or by using Visual SPSM. The `.CPD`, `.APD` and `.MPD` are similarly subsets of parameters but they are created automatically by Visual SPSM when parameters are changed.

A `.BAT` or `.CMD` file contains a sequence of responses to SPSM prompts corresponding to an actual user dialogue. These files are mainly used with Classic SPSM.

The relationships of all these files to the structure of the SPSM are shown in the following diagram. File types with their extension are shown in ellipses.
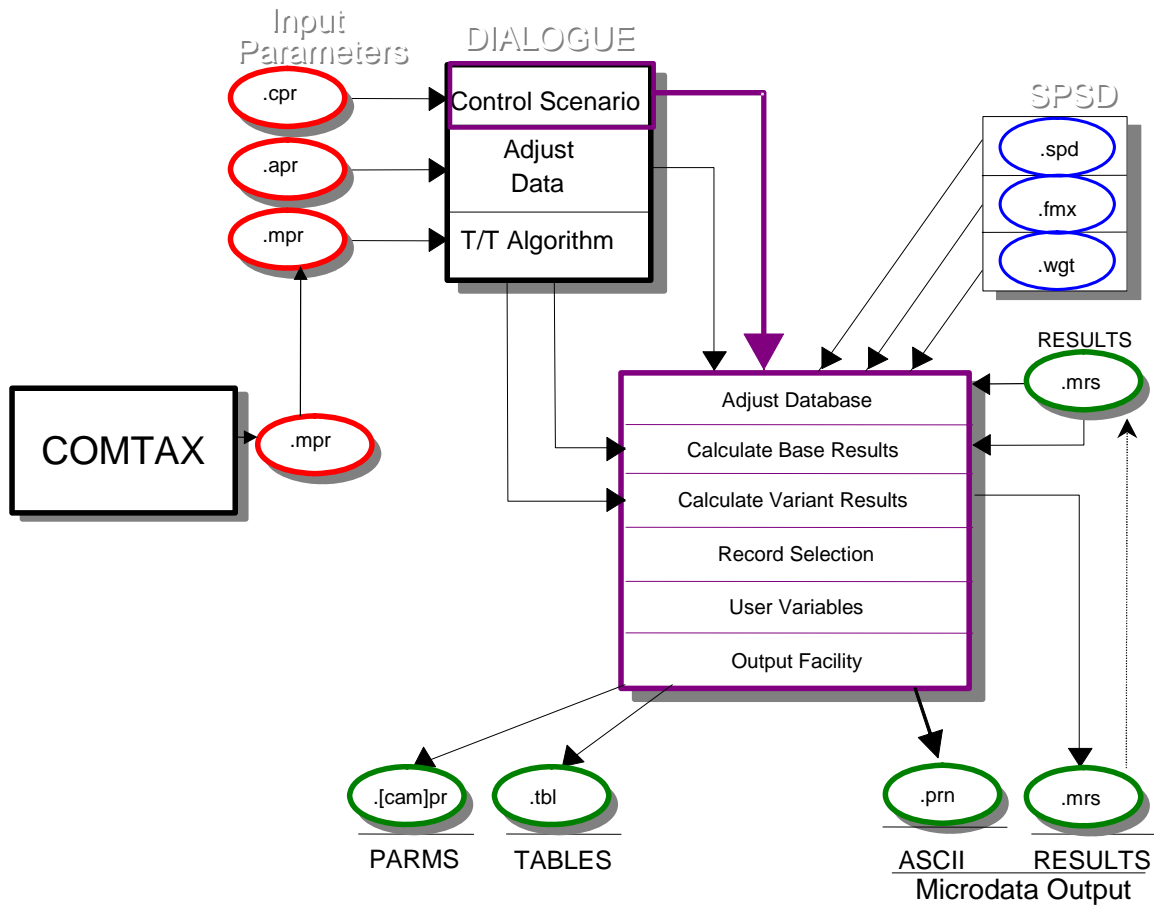
SPSD/M Files and Relationships

This diagram can be used as a handy reference while becoming familiar with the files and naming conventions. Note that two tax/transfer parameter files are input for base and variant results calculations. Also, the SPSM results file can only be used to give base results and can only be written using the variant results. Implicit in the diagram is the output of modified database adjustment and tax transfer parameter files during the dialogue.

SPSM User Facilities

The SPSM user facilities create a powerful and flexible environment for the analysis of various tax/transfer scenarios. All of the SPSM user facilities are controlled by parameters in the control parameter file. This section describes the purpose of each facility and provides an overview of its use.

USER-DEFINED VARIABLES

One of the powerful features of the SPSM is the ability to evaluate user specified expressions. These expressions allow users to create their own variables as logical and arithmetic manipulation of other SPSD/M variables. The expressions can be used in many of the SPSM user facilities described below. Expressions can allow users to simulate entirely

new tax and transfer programs totally within the black box mode. The following elements may be combined to form valid user expressions.

Variables	Any database, modeled, or user defined variables
Operators	Logical (and &&, or ‖, not !), comparative (<,<=,==,!=,>=,>), and arithmetic (+ - * /)
Constants	Any number
Parentheses	To control the order of evaluation of expressions
Level of analysis override	The level of analysis of a variable

For example, the following expression could be used to produce a variable which represents the ratio of average employment income to total income for individuals aged 21 to 64.

```
((idage >= 21) && (idage <= 64)) * idiemp/imitot
```

The term to the right of the multiplication sign (*) says that for any given individual take the total income and divide it by the employment income of that person. This produces the proportion of total income accounted for by employment income. Now consider the term to the left of the *. If an individual is over 20 years old and less than 65 years old then the first term will evaluate to 1 and the value of the expression becomes the value of the second term (the employment income ratio). If, however, the individual is under age 21 or over age 64 then the first term, and consequently the entire expression, will evaluate to 0.

Note that in this example the end result, if tabulated using the X-tab facility, becomes an average of ratios. If we had wanted to examine the ratio of averages, we would simply include the relevant terms of the expression in a custom table request.

PARAMETER EDITING

Parameter editing is usually performed using Visual SPSM. It can also be done with a text editor prior to opening Visual SPSM and read from a file.

Parameters types

The SPSM uses several types of parameters. Different types are checked for validity using different rules. The following is a description of the parameter types currently defined:

FLAG	A flag controls whether or not an algorithm (or program or calculation) is performed.
OPTION	An option parameter allows the choice of one of two or more options.
SCALAR	Scalar parameters take a single numeric value with or without a decimal point.
STRING	A string parameter is a single line or multiple lines of text.
VECTOR	A vector is a single column of numbers. The first value is the number of values which follow.
LOOKUP TABLE	A lookup table contains a single value followed by a set of three columns. The first column represents the input value, such as taxable income. The second column represents the output value corresponding to the input value in the same column. The third column represents the marginal change in the output value for the next increment (or tax bracket).
TABLE	A table is a numeric array with an arbitrary number of rows and columns. The array is preceded by a number indicating the number of rows which follow.

SCREENING RECORDS

A user will not always wish to simulate scenarios for every individual on the database, either because of the time required or because the simulation is applicable to only some sub-group of the population. In these instances the user will want to make use of the sub-sample facility or the selection facility. By using random number stream variables the user may employ these facilities to create take-up rates in black box mode.

Sub-Sampling

The SPSM runs quickly on modern computers, so subsampling is not usually necessary. However, it is still possible to run the SPSM with a partial sample.

In Visual SPSM, sub-sampling is controlled by setting a control parameter to the sub-sample desired. SPSM will then read and process the SPSD until the desired sub-sample has been attained. In Classic SPSM, users also have the option to press the CTRL-BRK key combination to interrupt the run after some fraction of SPSD has been processed. In either case, the sub-sample actually obtained is recorded in the control parameter file, and output tables are scaled by the reciprocal of this value.

Reading of the SPSD always proceeds sequentially, but a relatively unbiased sub-sample is nevertheless obtained because households in SPSD are by and large randomly ordered. The ordering is not completely random, however. In order to improve the accuracy of results obtained when using sub-sampling, SPSD has been arranged as a running stratified sample. Samples of set sizes (the first 5%, 25%, and 50% of the file) have been arranged to have representative provincial and household income distributions.

Selection Facility

The SPSM selection facility provides a means to restrict the set of individuals and families

which are processed by the SPSM output and reporting facilities. Selection satisfies two distinct needs. Firstly, it allows the analyst to focus attention on a particular sub-group of interest. Secondly, it can allow SPSM to execute faster by restricting the number of households processed to those of interest.

A string parameter contains an expression which is evaluated for each individual in the household. If the result of the evaluation is non-zero, the individual is considered to be selected. If an individual is selected, then **everyone** in the same family unit is also selected, irrespective of the evaluated value of the selection expression for other individuals. In other words, if at least one person in the family unit is selected, the entire family unit is selected.

To compute taxes and transfers correctly, SPSM always simulates entire households. Because of this, selection has no effect on the values of any variables, modeled or database, at the individual level. If, however, an examination of the database variables in the selection expression indicates that no individual in the household could possibly be selected, then SPSM skips immediately to the next household. This can result in a considerable decrease in SPSM execution time.

USER-DEFINED VARIABLE FACILITY

As already discussed, the SPSM has two distinct modes of use, termed 'black box' and 'glass box'. The 'glass box' mode provides the user with considerable flexibility to design new algorithms and create new variables, but it can be somewhat complicated to use, and requires some knowledge of programming. The user-defined variable facility (UVAR) allows the user to create new reporting variables in 'black box' mode. It allows the user to perform many analyses that would otherwise require programming changes to SPSM.

A new variable can be used just like any other SPSD/M variable. SPSD/M variables have built-in labels which are used by the SPSM reporting and output facilities. The user can supply a label to an analysis variable or it will use defaults. SPSD/M user variables also have an associated output precision, which gives the number of digits after the decimal point to use when formatting the variable in reports.

There are many applications for these variables. Counts of individuals or family units with specified characteristics, take-up rates, averages, percentages and even new benefit functions can easily be incorporated in black box mode.

CREATING TABLES

Built-in Tables

The SPSM is capable of producing ten different built-in tables. The tables have been designed to contain a large amount of data which can be processed to produce additional statistics. So, for example, the tables do not provide information on average OAS benefits. However, they do provide information on the total value of OAS benefits as well as the number of recipients of OAS which can then be used to determine average benefits.

The tables are grouped into five sets of pairs. The first table in each pair contains values in millions of dollars for over fifty selected variables. The second table in each pair contains

the number of persons, families or households (controlled by the user) which had some non-zero value for the variable reported. The five groups produce the following types of tables.

Table 0 and 0A   Canada totals
Table 1 and 1A   Totals by province
Table 2 and 2A   Totals by income group
Table 3 and 3A   Totals by family type
Table 4 and 4A   Totals by proportion of a specified income threshold

The user has only limited control over the format and contents of these tables. The user specifies a level of analysis which applies to all of the built-in tables. He or she can also specify the variable as well as the cut-points to be used in determining the columns of Table 2. Thus the user can easily create a table of results by age. A separate utility allows any of these built-in tables to be transformed into a format which is readable by Excel spreadsheets. An Excel spreadsheet which produces auxiliary statistics from these tables is also included.

User-specified X-tab Facility

The designers of the SPSM recognize that the built-in tables will not always meet the analyst's needs. Thus, a powerful facility that allows a user to create specific customized tables has been provided. A brief description can be found in the _User's Guide_ while the separate auxiliary guide, _XTab User's Guide_, presents detailed documentation on the X-tab facility.

An individual tabulation request consists of multiple components. It typically begins with the specification of the relevant unit of analysis, i.e. individual (IN:), nuclear family (NF:), census family (CF:), economic family (EF:) or household (HH:). If the tabulation request does not include a unit specification, a default unit of individual (IN:) applies. The remainder of the tabulation request consists of one or more "levels" separated by asterisks. One of the levels, the tabulation level, indicates the item or items to be tabulated. Any remaining levels are classification levels. They specify the particular pattern of categorization desired in the table.

Naturally, there are some restrictions on just what the X-tab facility can tabulate, but the design of the SPSM makes it quite flexible. For starters, the analyst can tabulate any analysis variable (any non-categorical variable) available in the SPSD or calculated in the SPSM. Complete lists of these variables appear in the _Variable Guide_. As well, the analyst can tabulate any of the user-defined variables. Further, the analyst can define "on the fly" the desired tabulation expressions, constructing them from any of the preceding types of variables and appropriate mathematical operators.

An example of a table request is:

```
EF: hdprov+ * {units};
CF: {immtot, units, immtot/units} * hdtenur+;
```

This request would generate two tables. The first would contain values for the number of economic families in each province as well as a total row. The second table (to the right of the semi-colon) would tabulate total income accruing to census families, the numbers of such families, and the average income per census family, all broken out by tenure categories

(rented, owned with mortgage, etc.) as the columns.

DISTRIBUTIONAL ANALYSIS FACILITY

The distributional analysis facility calculates a set of univariate distributional statistics for a sample of SPSD cases. If the distributional facility is activated a user specified random sample of up to 200,000 cases is stored and sorted in memory and distributional statistics are calculated on this sample. The statistics provide the following information.

- Number of zero and non-zero observations
- Descriptive Statistics (sum of weights, sums, sum of squares, weighted mean)
- Extreme values (bottom five values and top five values)
- Selected quantiles (quintiles, quartiles, etc.)
- Histogram plot

The output from the distributional facility can be found under the **Tables** tab in Visual SPSM. In addition, all output from the distributional facility is written to the .TBL file for the program run. An example using the distributional facility is shown in tutorial session 3. The following is an example of output from the distributional facility that has been produced for the variable *imccea* (Child Care Expenses Allowed). It was created using a previous version of the SPSM.

```
                       SPSD/M (Database 10.00)
                   Base Description: No base results
                Variant Description: Current values for 1992
        [Driver: Version 10.0 : 88-09, File: $SPSD/ba92.mpr]
                        Sample: 1.0000  AGENAME='Standard'
        Distribution report: Child care expenses allowed for Individuals

            Total observations = 234659
            Zero observations  = 225632
            The following statistics are based on 9027 non-zero observations.

            Descriptive Statistics:

            Sum of weights         =         872953
            Weighted Sum           =      1908380904
            Weighted Sum of Squares = 7555806526246
            Weighted Mean          =           2186

            Extreme Values (with associated household numbers):

               Minima  hdseqhh               Maxima  hdseqhh

                    1    81136                 12000      484
                    2    41280                 12000      485
                    6    30241                 12000      488
                    6    30242                 12000    28664
                    7    40293                 12000    71555


            The following statistics are based on a sample of 3000.


            Selected Quantiles:

            Q1 = 569   P1=  78   P90=5038   P20= 444   P60=2032
            Med=1678   P2=  88   P95=6308   P30= 701   P70=2786
            Q3 =3181   P5= 147   P98=8000   P40=1062   P80=3729
                       P10= 221   P99=8000
```
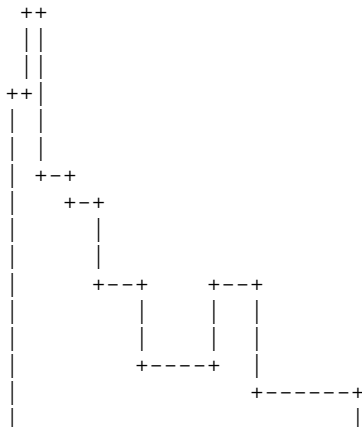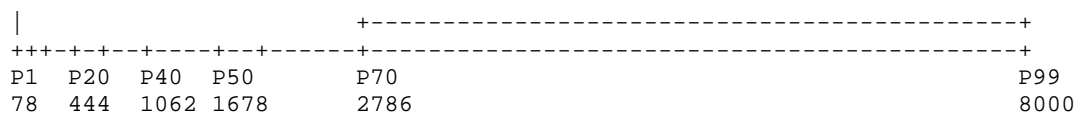
Note that the maximum value on the database is equal to a full deduction for three children. If extreme values are not as anticipated, the user can then produce a detailed report on the relevant households by using the selection facility and the print file output facility.

```
  Histogram Plot:


    ++
    ||
    ||
  ++|
  | |
  | |
  | +-+
  |   +-+
  |     |
  |     |
  |     +--+     +--+
  |        |     |  |
  |        |     |  |
  |        +----+  |
  |             +------+
  |                    |
  |                    |
```

```
|                             +--------------------------------------------+
+++-+-+--+----+--+------+------+--------------------------------------------+
P1  P20 P40  P50     P70                                                  P99
78  444  1062 1678    2786                                                8000
```

The distribution report has been produced for all non-zero values of *imccea* showing the skewed, truncated shape of the density function for the tax expenditure. The bi-modal distribution reflects the maximum claim for families with different numbers of children.

MARGINAL TAX RATE FACILITY

The marginal tax rate is defined as the proportion of an extra dollar of income that is taxed. It is a useful concept because it measures the extent to which incentives to obtain additional income are reduced by the tax/transfer system. SPSM provides a facility to compute marginal tax rates. When the facility is activated, the tax/transfer system is applied twice to each household: once to the original incomes and once to the incremented incomes, and the resulting change in consumable income is noted.

The definition of marginal tax rate given above is not complete. To calculate a marginal tax rate, one must in addition specify the source of income being incremented, the amount of the increment to income, and which individuals are to receive the increment. The change in consumable income at the individual level can then be aggregated to produce marginal tax rates at different family levels of analysis.

Marginal tax rates differ depending on the family level of analysis. Consider a married couple, where one person has zero income. If $500 is given to the individual with no income, that person's marginal tax rate would be zero. The consumable income of the person's spouse would decrease however, due to a reduction in the married tax credit/deduction. Hence the marginal tax rate of the spouse would compute to an infinite value, since the spouse's taxes have increased, even though he/she received no additional income. Considered as a family, however, the marginal tax rate would evaluate to a reasonable value.

CASE OUTPUT FILES

In addition to tabular data the SPSM can write out information from the database and model to a data file in a number of different formats. These files contain the specified sub-sample of an individual SPSM program run and can be used for subsequent analysis using either the SPSM, a text editor, or another software product. The cases written can be at any level of analysis and may include any database or modeled variables. For example an output file could be created containing the age, sex, census family type and incomes of all individuals aged 18 to 64. Alternately a file containing the urbanization class, province, residential tenure, number of children and government cash transfers of all census families in Alberta could be produced.

Text Output Facility `(.prn)`

The SPSM text output facility creates a character file containing case by case micro-data information. One can view it using Visual SPSM, or one can use it as input into some other

computer program, such as a spreadsheet, database system, or other SPSM facility.

The variables whose values are to be displayed are given in a string parameter (ASCVARS). Each case that is output corresponds to the level of analysis given in another control parameter (ASCUNIT). The selection facility is generally used in conjunction with this facility, since otherwise very large files could result. The facility is especially useful for testing new scenarios. For example, if a number of households losing disposable income are noticed in a run that should only result in gainers, the relevant information on an entire losing household can be printed out and examined.

A number of styles of output are supported. An ASCSTYLE value of 1 produces a report designed to be easily human-readable. One household is output per page, and one variable is output per line as shown below.

Text output facility style #1

```
<page break>
hdseqhh   Household sequence number ............    61
idefseq   Economic family sub-sequence number ..     0     0     0
idcfseq   Census family sub-sequence number ....     0     0     0
hdprov    Province ............................     6
idage     Age .................................    28    24     0
idsex     Sex .................................     0     1     1
idcfrh    Relationship to census family head ...     0     1     2
immmkt    Market income ....................... 27953 17350     0
immtran   All transfer income .................   194   371     0
<page break>
hdseqhh   Household sequence number ............    62
idefseq   Economic family sub-sequence number ..     0     0     0
idcfseq   Census family sub-sequence number ....     0     0     0
hdprov    Province ............................     6
idage     Age .................................    53    50    18
idsex     Sex .................................     0     1     0
idcfrh    Relationship to census family head ...     0     1     2
immmkt    Market income ....................... 36457  3750  4274
immtran   All transfer income .................     0  1222     0
```

An ASCSTYLE value of 2 produces a report designed to be read using a spreadsheet import function while 3 produces a report designed to be read using a spreadsheet or a database system. An ASCSTYLE value of 5 produces a report designed to be converted into a compressed format that can be read by the SPSM. This feature allows users to modify SPSD database files.

Below is an example of ASCSTYLE=5 for one household (not all variables values are shown).

```
000001 47 217 2 1 2 1 3 4 1
0 0 0 0 0 0 0 42 0 3 11 13 99 1 2 3 0 52 1 0 0 0 0 0 0 0 0 0 1 0
0 0 ...
```

SPSM results file (.MRS)

The SPSM can produce a special file known as an SPSM results file. This file is written using the same data compression techniques as the SPSD files. Because of this, SPSM results files occupy a modest amount of disk space but they can only be read by the SPSM program. The results file is written using the variant results, typically so that it can be read in

as base results in a subsequent run. If base results are generated by reading in an existing results file, the SPSM operates about twice as fast as if results had been calculated "on-the-fly".

The SPSM results file is always written at the individual level of analysis and may contain any number of modeled variables.

TURNING POINT FACILITY

The SPSM Turning Point Facility provides a means to analyze the points at which an individual household's marginal tax rate changes as its income increases. The marginal tax rate is the proportion of tax that is paid on one additional dollar of income. The turning points represent changes in the slope of the entire tax/transfer function (the marginal tax rate). The facility computes the various income levels where an individual household, given its characteristics, would experience changes in its marginal tax rate and then outputs information on the current value of a number of variables at each of these levels of income. The facility is useful for determining, for example, if a proposed means tested social assistance program will result in a disincentive to look for employment.

GOAL-SEEKING FACILITY

This tool is useful for answering "what if" type questions. For example, let's assume that you wish to get rid of the goods and services tax (GST) and replace the lost revenues by an increase in the income tax rates. So, what would the income tax rates have to be? It would be possible to answer this question by first running a base case to observe the level of federal tax revenue obtained when the GST is collected, and then set a variant case which sets the GST to zero.

The facility operates by iteratively executing a black box version of the SPSM and manipulating the parameter files between runs based on the values found in the default tables. See the _Tools User's Guide_ for more information.

REFERENCE VARIABLE FACILITY

Values of any number of modeled variables produced in an SPSM run can be accessed in a subsequent run by using SPSM reference variable facility. The facility permits the simultaneous comparison of any number of simulation scenarios. The most common use of the Reference Value Facility is to access user variables produced in a previous run. The reference value facility can also be used to replace the values of SPSD variables. The user may use this feature to, for example, model a behavioural response of persons responding to a shortfall of UI benefits by increasing their employment income.

Example Simulations

This section contains several worked examples of SPSD/M use. They use the Visual SPSM interface. They should be performed by the user at his or her computer in the sequence presented here. The examples provide guidance on using some of the many SPSD/M features.

The text and results obtained with your computer may not exactly match the results printed in this guide. The results on your screen, if the example steps are followed exactly, should be interpreted as correct. This is due to the fact that the documentation sometimes lags the most recent version of the software. Please call Statistics Canada at (613) 951-3774 should you have any questions.

**SESSION 1: CHANGE TAX/TRANSFER PARAMETERS**

This example creates a simple simulation which changes the values for the National Child Benefit Supplement in 2003. The maximum supplement will be increased by $200 per child.

First, open Visual SPSM. Create a new simulation by clicking on **File**, **New**. The following box appears, scroll down and pick ba03.cpr to create a simulation in 2003 with the 2003 tax and transfer system.
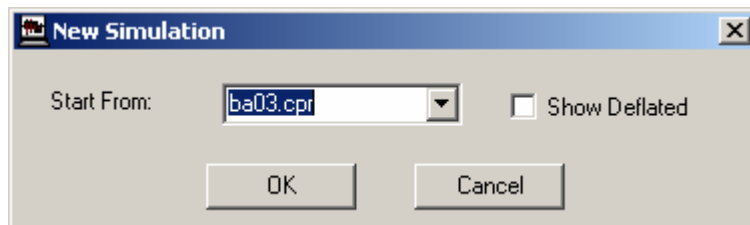

The goal of this simulation is to increase the Child Benefit Supplement. The first step is to find the parameters which need to be edited. One way of doing this is to use the SPSD/M help file. Click on **Help, Contents** to open the Help file. You can either search the help file or you could navigate the tree. Let's start by navigating the tree. Open the parameter guide. You want to find the parameters for the supplement, so you click on Government Transfers and Personal Income Taxes. Then click on Government Transfers, then finally on the Federal Child Tax Benefit. You can read about what the various parameters do, but you will find that the three parameters that you need to change are FCBEIS1, FCBEIS2, and FCBEIS3 which represent the value of the benefit for the first, second, and third+ child.

You then need to close the help file and go back to Visual SPSM. We want to increase the value of these parameters by $200. Given the fact that you know the names of these parameters, you could simply type them in the Edit box at the top left hand of the screen, and then press Enter.
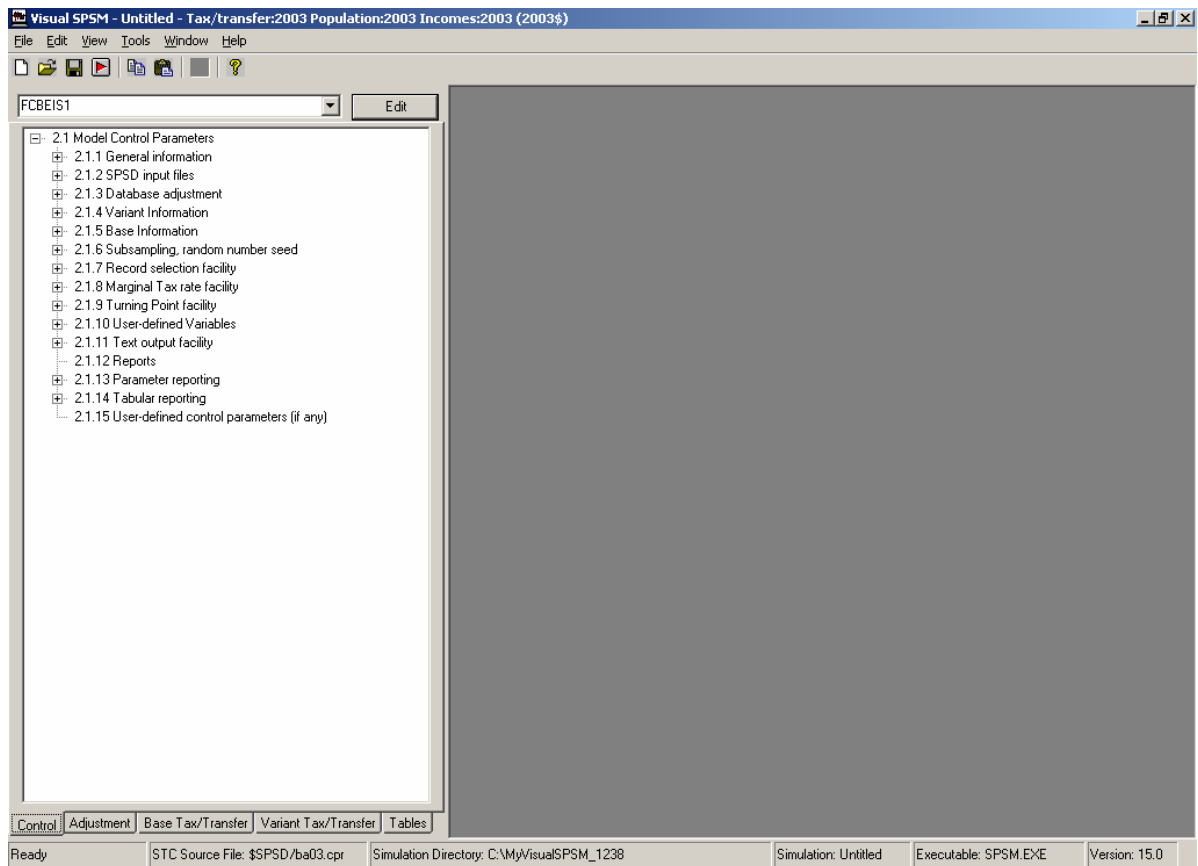
But instead, let's navigate to the tree again. We do not want to run a base scenario, so click on the Variant Tax/Transfer tab at the bottom of the screen. Then open the tree until you find the Federal Child Tax Benefit parameters.
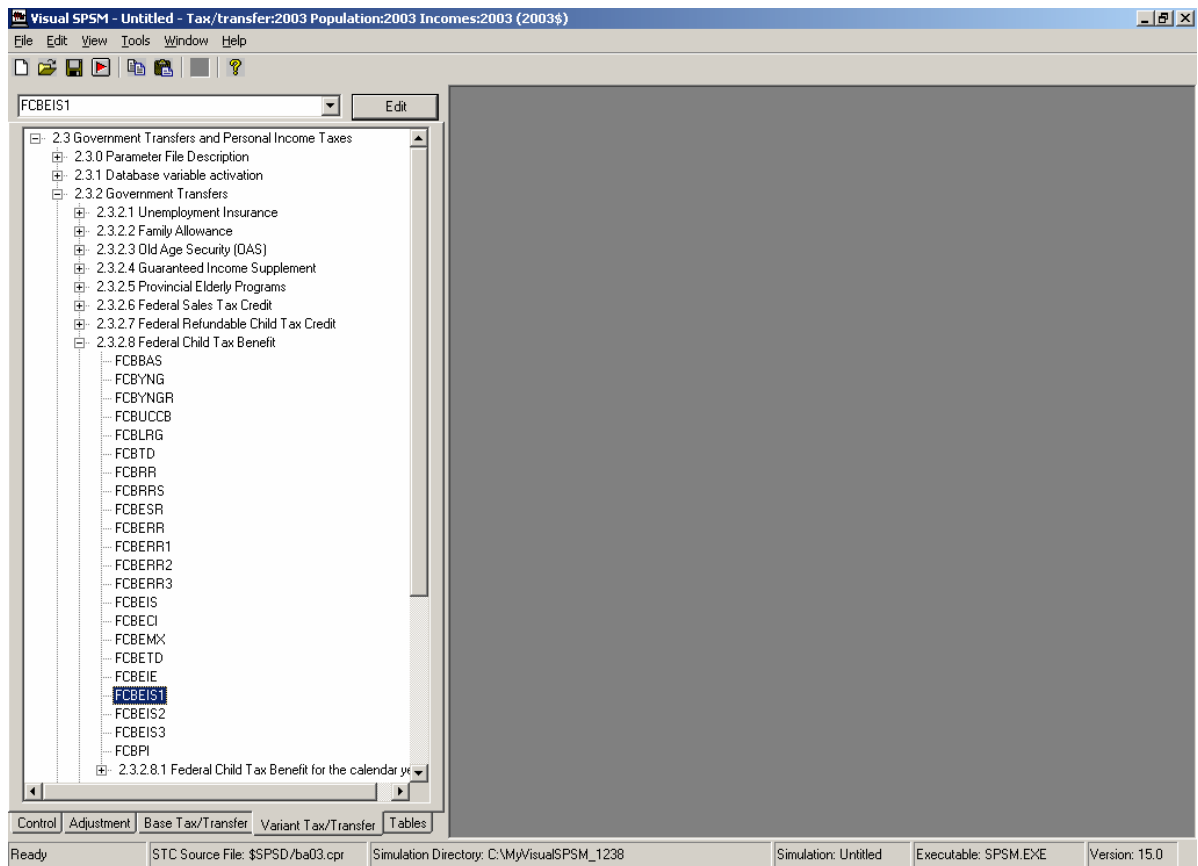
Double click on FCBEIS1 to open the Parameter window. You see the following screen.
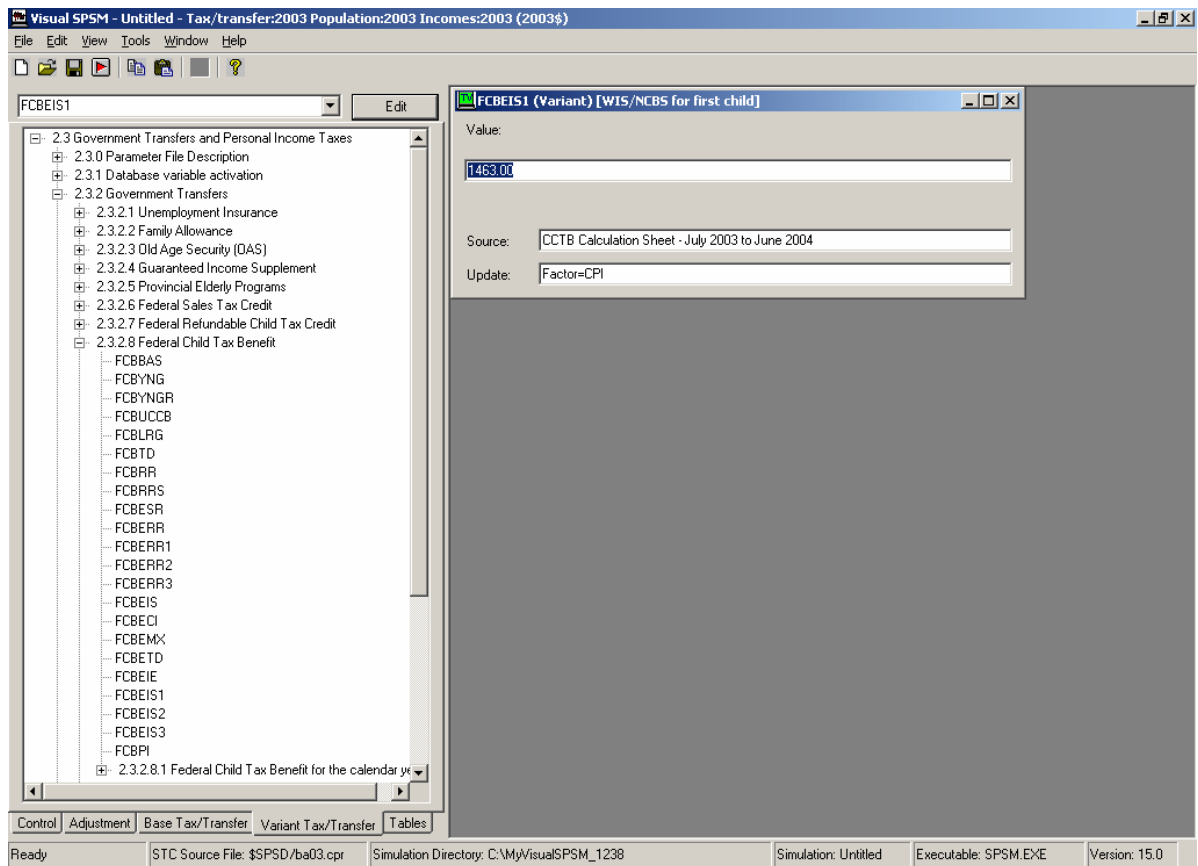
You then want to replace the value of the parameter of $1463 by $1663, increasing the benefit by $200. You may also wish to change the Source, indicating that this is a new value.
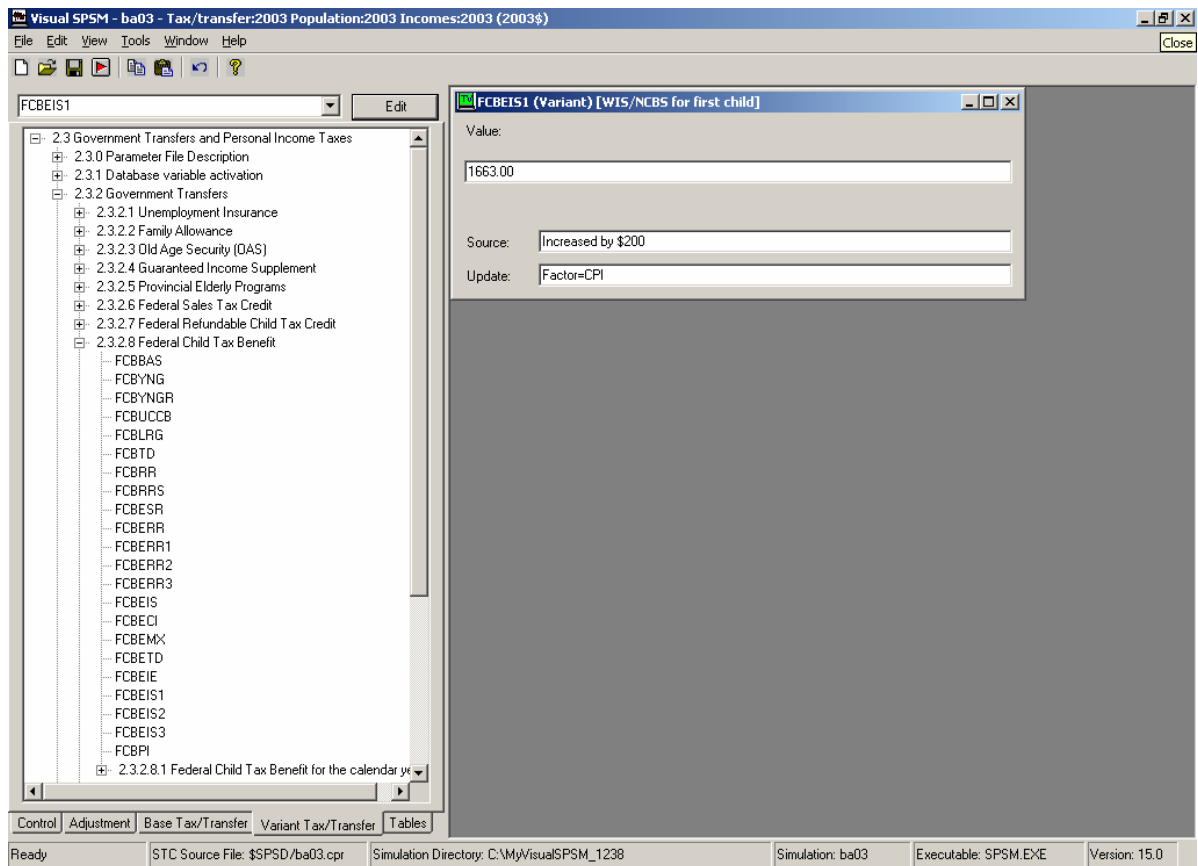
Increase FCBEIS2 and FCBEIS3 in the same manner. The only other thing that we will change is the description of this scenario. Scroll to the top of the tree and open the Parameter File Description tab. Then click on MPRDESC. This parameter allows you to label this variant scenario. Change the text to read "Increase NCBS by $200 per child".

This is all the parameters we need to change. In this scenario, we are going to use the built-in tables, so there are no other changes necessary. Save the simulation in a new directory by clicking on File, Save Simulation As... Save the simulation as Ex1.cpr in an appropriately named directory. Then run the simulation by clicking on File, Execute Simulation.

The Log Window opens if there are errors. If the simulation executed successfully you may still scan the Log Window (as you did in the Quick Start examples) by selecting **View, Log** from the toolbar. Close the window.

Click on the **Tables** tab. Open Table 0, results by Census family. If you scroll down, you'll see that CTC/Child Benefits are now $7,557.70. Since we didn't run a base scenario, we can't compare the results. But if you open up the simulation you ran in Quick Start which did a simulation for 2003 with no tax and transfer changes, you will find that the benefit used to be worth $7,135.20 so the change amounted to $422.5 million in increased benefits. You

cannot, though, find out what was the average increase per family. To do that, you should create a base scenario and also create new tables and new user variables. The next example will show you how to do that.

SESSION 2: BASE SCENARIO, CUSTOM TABLES, USER ANALYSIS VARIABLES.

In this example, we will keep the same tax and transfer change that we included in Session 1, namely we will increase the maximum National Child Benefit Supplement (NCBS) per child by $200. However, in this example we will compare the results with a base scenario where the change did not occur, we will create user variables, and we will create user tables.

The first thing to do is to open the example simulation you created in Session 1. So go to File, Open Simulation, and then navigate to the Ex1.cpr and open it. This scenario includes in its variant scenario all the changes in the tax and transfer system that we need.

The first thing that we wish to do is to create a base scenario. In order to do this, type in BASMETH in the Edit parameter box and click Edit. The following box opens up.
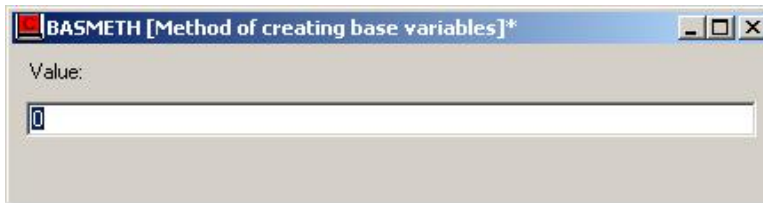
By default, BASMETH is set to zero. To find out what zero means, open the SPSD/M help file (by clicking on Help, Contents) and type in BASMETH in the index. You see that when BASMETH is set to zero, no base results are used. We will want to set BASMETH to 2, in order to create base results using the default tax and transfer algorithm.

The next thing to do is to create some new variables. To do this, go to User Defined Variables in the tree under Control Parameters. Two parameters can be edited. Open them both. The first is UVARFLAG which turns on the user variable facility. Check the box so that user variables are created.

The second is UVAR. This opens a blank box which can be edited. The *User's Guide* gives you the syntax required to create user variables. We want variables which will represent the distribution of the difference between the base and the variant amounts of NCBS, as well as variables indicating whether a person received any NCBS at all. The first step is to find out the name of the variable of interest, the variable which contains the value of the NCBS. Using the help file, we find that the variable is called `imfcbene`. To get the base amount for a variable, you need to put an underscore before the name. The following shows what ought to be entered in the UVAR box in order to create the user variables for this example:

```
diff = imfcbene-_imfcbene;
spdiff = split(diff,0, 200, 400);
label(spdiff)="Increase in benefits";
gotvar = imfcbene>0;
label(gotvar) = "Received variant benefits";
```

Three variables have been created and we have labeled each. `diff` is the difference between the variant benefits and the base benefits. `spdiff` uses the split function to create a classification variable which splits the difference between the variant and the base benefits, split into four groups, Min-0, 1 – 200, 201-400, and 400+. `gotvar` will be equal to 1 when a person received variant benefits and zero otherwise.

Finally, we now want to create new output tables. We will need to edit two more control parameters, XTFLAG and XTSPEC. The former turns on the user-tabulation facility, while the latter will define the tables.
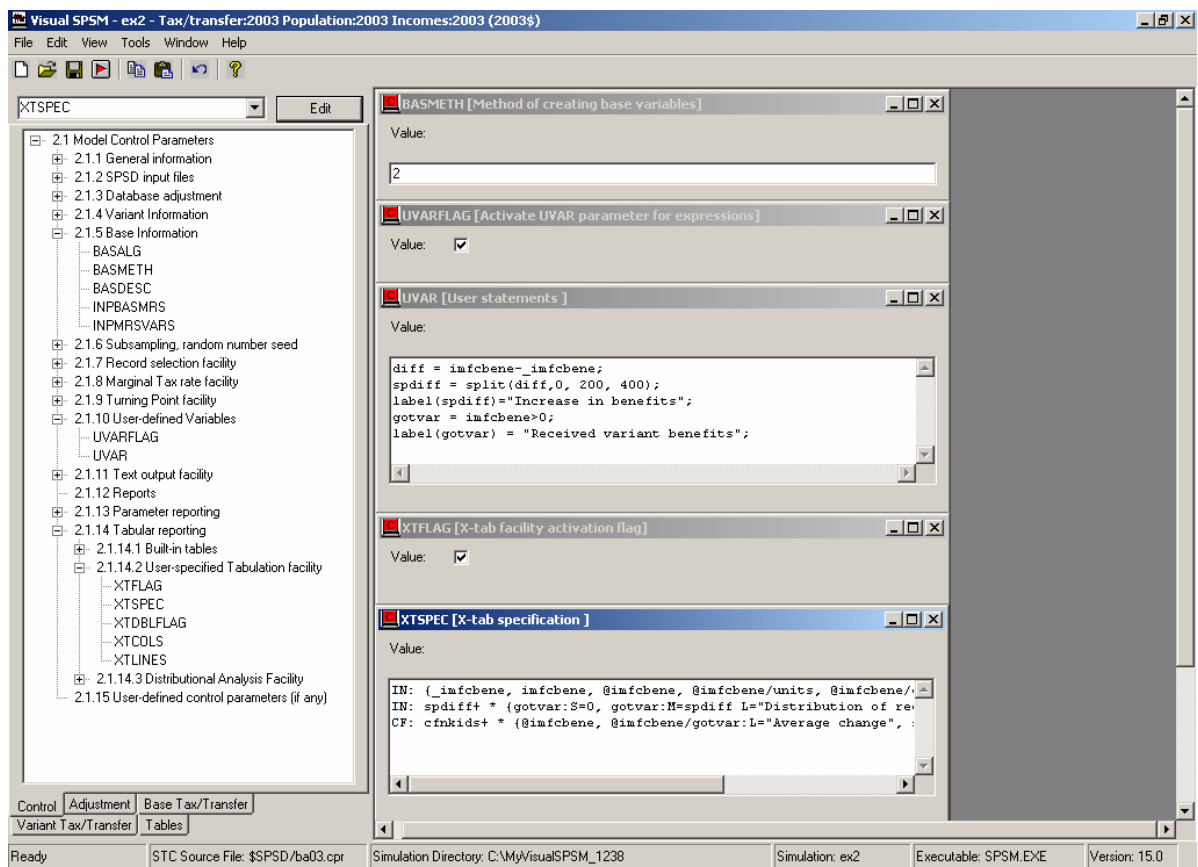
Open the two parameters, using the edit box. Tick on XTFLAG to produce tables and then enter the following in XTSPEC.

```
IN: {_imfcbene, imfcbene, @imfcbene, @imfcbene/units,
@imfcbene/gotvar:L="Average change in NCBS"};
IN: spdiff+ * {gotvar:S=0, gotvar:M=spdiff L="Distribution of
recipients"};
CF: cfnkids+ * {@imfcbene, @imfcbene/gotvar:L="Average change", spsdrecs};
```

This will produce three user tables. Each table request ends with a semi-colon, and each
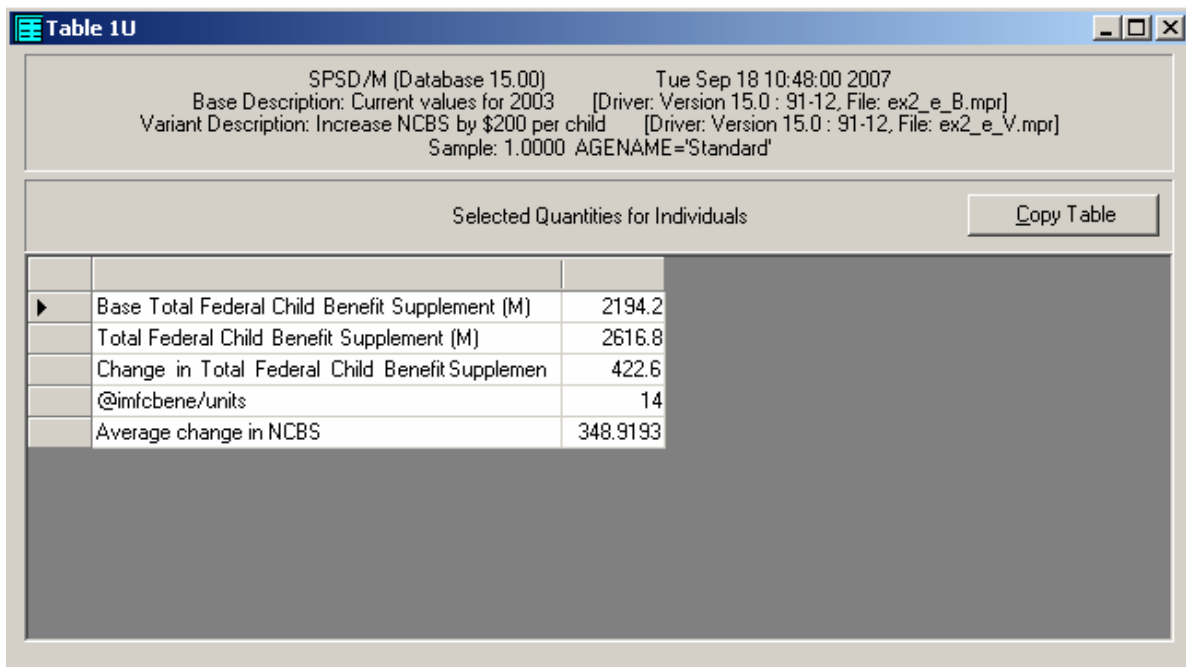
starts with a prefix indicating which family level is being tabulated. So the first table is a table of individuals while the third is a census family table. See the *User's Guide* or *XTab User's Guide,* for more information about creating tables. At this point save the simulation and run the model. You can then open the three user tables. We will go through them one by one.

Table 1: Totals, changes, and averages

Here is the syntax again of the first table.

```
IN: {_imfcbene, imfcbene, @imfcbene, @imfcbene/units,
@imfcbene/gotvar:L="Average change in NCBS"};
```

It is a table of individuals which will contain five columns. The first two will be the base and variant values of NCBS, with the base variable being distinguished by the leading underscore. When a variable is prefixed by @, it means that the difference between the variant and the base will be produced, which is what is output in the third column. Finally, the fourth column represents the average change in benefits for all individuals whereas the fifth is the average change for people who received benefits. The units keyword represents individuals in an Individual level table. The last two columns had their labels overwritten using the "L" qualifier. The results of the table can be seen below. This change cost $423 million with an average change of $14 / person or $423 / person who received benefits.
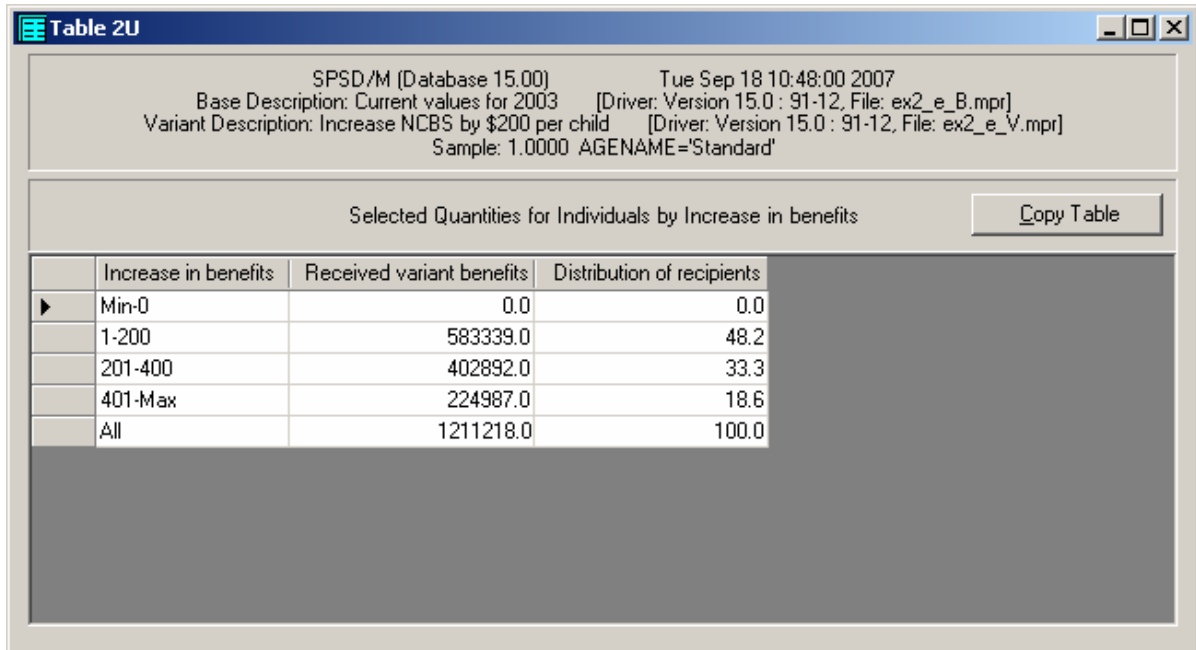

Table 2: Classification variables and distributions

Here is the syntax of the second table.

```
IN: spdiff+ * {gotvar:S=0, gotvar:M=spdiff L="Distribution of
recipients"};
```

The second table gives the distribution of people who receive the new benefit by the amount received. It is a two dimensional table with the addition of a classification dimension

(spdiff). The same variable, gotvar, is tabulated twice -> the difference is that the second variable uses the M qualifier to normalize the variable. The results show that thought the average change in benefits $423 (as seen in table 1) 48.2% of people received less than $200.

Table 2U

SPSD/M (Database 15.00)   Tue Sep 18 10:48:00 2007
Base Description: Current values for 2003   [Driver: Version 15.0 : 91-12, File: ex2_e_B.mpr]
Variant Description: Increase NCBS by $200 per child   [Driver: Version 15.0 : 91-12, File: ex2_e_V.mpr]
Sample: 1.0000 AGENAME='Standard'

Selected Quantities for Individuals by Increase in benefits   Copy Table

Increase in benefits	Received variant benefits	Distribution of recipients
Min-0	0.0	0.0
1-200	583339.0	48.2
201-400	402892.0	33.3
401-Max	224987.0	18.6
All	1211218.0	100.0

Table 3: Family tables and records

Here is the syntax of the third table.

```
CF: cfnkids+ * {@imfcbene, @imfcbene/gotvar:L="Average change", spsdrecs};
```

The final table changes the unit of tabulation. Instead of tabulating individuals, it tabulates census families. Only one member of a census family can receive the NCBS, so the same variables as for the individual tables can be used safely. It gives the total and average change of NCBS. You can see that the average change in all groups but families with six children is less than the maximum. This is because the benefit is clawed back by income. This table also shows the number of SPSD records used to create the table. It indicates that great caution ought to be used for families with 6 children as there were only 6 records used and the statistical variability will be large and the results will not be reliable.

Table 3U			

SPSD/M (Database 15.00)   Tue Sep 18 10:48:00 2007
Base Description: Current values for 2003   [Driver: Version 15.0 : 91-12, File: ex2_e_B.mpr]
Variant Description: Increase NCBS by $200 per child   [Driver: Version 15.0 : 91-12, File: ex2_e_V.mpr]
Sample: 1.0000 AGENAME='Standard'

Selected Quantities for Census Families by Number of children in census family   [Copy Table]

Number of children in census family	Change in Total Federal Child Benefit Supplemen	Average change	SPSD Records
0	0.0	0.0000	63038
1	110.8	194.5047	12849
2	161.6	387.6283	13355
3	89.3	584.9701	3899
4	37.7	777.6419	971
5	19.5	951.4706	462
6	3.8	1200.0000	6
7	0.0	0.0000	0
8	0.0	0.0000	0
9	0.0	0.0000	0
10	0.0	0.0000	0
11	0.0	0.0000	0
12	0.0	0.0000	0
13	0.0	0.0000	0
14	0.0	0.0000	0
15	0.0	0.0000	0
16	0.0	0.0000	0
17	0.0	0.0000	0
18	0.0	0.0000	0
19	0.0	0.0000	0
20	0.0	0.0000	0
21	0.0	0.0000	0
22	0.0	0.0000	0
23	0.0	0.0000	0
24	0.0	0.0000	0

SESSION 3: NEW EARNED INCOME TAX CREDIT SIMULATION

This final tutorial session shows a more advanced use of the SPSD/M in black box mode. It is not necessary to proceed with this example before conducting research with the SPSD/M. Those readers who are comfortable with the use of the model thus far are invited to try this last session.

In this session we will be testing the fiscal impacts on households and governments of introducing an earned income tax credit (EITC) program. The testing of a new program would normally be a glass box application requiring re-compilation of the model. However for certain types of programs the black box mode is adequate. Specifically, these cases are where the new program has no implications for personal taxes and are not extremely complicated. The technique makes use of the user-defined variables and selection facility.

The earned income tax credit for this example has the following specifications. These specifications are not meant to represent a realistic earned income tax credit, but they are sufficiently close for the purposes of this example.

Eligibility: Individuals aged 21 years or over are individually eligible if they reside

	in census families in which children under the age of 21 years are present.
Maximum Benefit:	Maximum benefits are $1,200 for each eligible person.
Income Test:	Income testing will be based on the employment income of an individual plus the employment income of their spouse if present.
Turning Points:	Benefits begin with the first dollar of earned income. At $8,000 dollars of employment income the benefits reach the maximum and continue at that level until $12,000 of earned income where they begin to be reduced.
Reduction Rates:	15% of earned income is payable as a benefit up to a maximum of $1,200. After $12,000 of earned family income maximum benefits are reduced by 10¢ for each additional dollar of earned income.

The above specifications can be represented by the following graph.
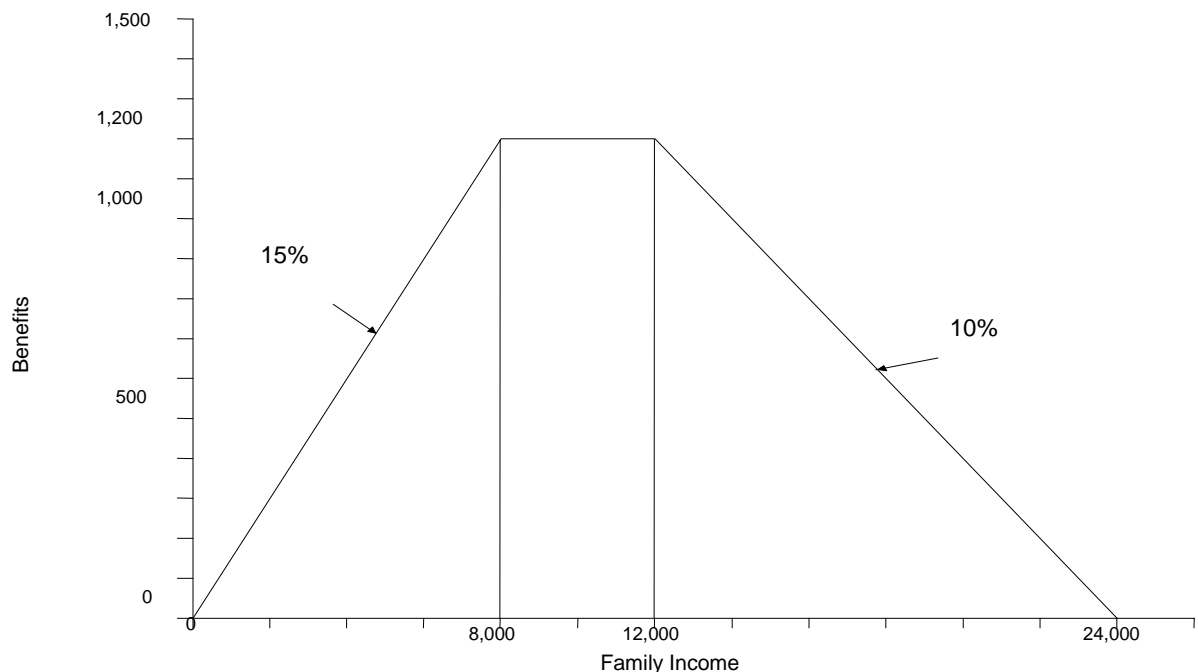
Figure. Earned Income Tax Credit Benefit Structure

This example makes use of a partial include file which contains the parameters that we wish to modify. To create this simulation, you need to open a new simulation in Visual SPSM for 2003, and then go to **File**, **Read Parameter File**, and then read the file named `eitc_EN.cpi` which can be found in the spsm\example directory in the location where you installed spsm. These partial include files are text files and can be examined using a program like NotePad.

The four figures presented on the following pages contain the content of `eitc_EN.cpi`. Numbered comments are used to explain the file's contents.

```
[1]     ####
        ##eitc.cpi - Include file for Tutorial Session 3
        ##
        ##$Id$
        ##
        ##Purpose: This file contains the necessary control parameters to simulate
        ##      a new earned income tax credit.  The description of this new
        ##      program and explanation of parameters is contained in the
        ##      Introduction and Overview Guide under the heading of Tutorial
        ##      Session 3.
        ##
        ################################################################# ###################

        ###
[2]     ## 2.1.7 Record selection facility
        ###

        SELFLAG            1                 # Selection facility activation flag
        SELUNIT            2                 # Selection facility family level
        SELSPEC                              # Selection specification
                    (idage < 21) && (idcfrh == 2)
```

1. The comment block at the beginning of the file provides information on the purpose of the file contents as well as the name of the file. These comment blocks are useful in parameter include files but should not be used in the parameter files themselves, because the SPSM will strip these block comments (lines starting with ##) from the output parameter file.

2. The selection facility will be used to perform a portion of the eligibility test. The selection level of analysis is set to census families (SELUNIT=2). The selection specification checks for individuals under the age of twenty-one whose relationship to the head of the census family is that of a child. As a result, only those census families with children under 21 will be selected. More precisely, if SELSPEC evaluates to zero for every member of a specific SELUNIT (a census family in this case) then that family will be rejected.

```
        ###
[3]     ## 2.1.9 User-defined Analysis Variables
        ###
        UVARFLAG    1
        UVAR
[4]     nfempinc=idiemp + SP:idiemp;
[5]     if (idage>=21) {
[6]     if (nfempinc<8000) {
            eitc = .15*nfempinc;
        }
[7]     else if (nfempinc<=12000) {
            eitc = 1200;
        }
[8]     else if (nfempinc<24000) {
            eitc = 1200-((nfempinc-12000)*.10);
        }
        }
        else {
           eitc=0;
        }
[9]     label(eitc)="Earned income tax credit";
```

3. The user-defined variable facility is used to simulate the EITC program. Keep in mind that each user-defined variable is evaluated sequentially at the individual level of analysis. Consequently expressions should not make reference to user variables that have not yet been

defined.

4. *nfempinc* is the combined employment income from wages and salaries of the head and the spouse of the census family which will be subsequently used for income testing. This is accomplished by specifying the employment income of the individual being processed (*idiemp),* plus the employment income of the spouse (SP:*idiemp),* using the modifier SP: to indicate the value of the variable should be read from the spouse.

5. Benefits will be restricted to individuals over the age of 20. This is accomplished by establishing a conditional group of statements with in the braces{}. See the *User's Guide* for more information on statements.

6. The EITC benefit payable to eligible individuals having a combined head/spouse employment income of less than $8,000 is then calculated as 15% of head plus spouse employment income. This is the first slope of the function. Note that zero employment income will result in zero benefits.

7. The maximum EITC benefit of $1,200 is then payable to eligible individuals having a combined head/spouse employment income of between $8,000 and $12,000.

8. The EITC benefit payable to eligible individuals having a combined head/spouse employment income of between 12,000 and 24,000 is then calculated as maximum benefits less 10% of employment income in excess of 12,000. This is the third and final slope of the function.

9. The variable *eitc* is then labelled.

```
[10]    dispinc=immdisp+eitc;
        label(dispinc)="Disposable Income";
[11]    gainer=eitc>0;
        label(gainer)="Received EITC Flag (Gainer)";
[12]    Nochange=(eitc==0);
        label(nochange)="Unaffected by EITC Flag";
[13]    agegrp=split(idage,20,64);
        label(agegrp)="Age";
[14]    Empigrp=split(idiemp,0,8000,12000,24000);
[15]    Dispgrp=split(immdisp,5000,10000,15000,20000,25000,
                            30000,35000,40000,45000);
        label (dispgrp)="Base disposable income group";
```

10. Because the definition of *immdisp* (disposable income) is not modifiable in black box mode we will need to create and label a new variable *(dispinc)* which shows disposable income after the new tax credit.

11. *gainer* is true (1) if the individual gained from the EITC. Otherwise the value is zero

12 nochange is true (1) if the individual was unaffected by the EITC. Otherwise the value is 0.

13. We want to examine the results across three different classification dimensions. These new variables allow us to validate the SPSM implementation of the new EITC program and to measure the distributive effects of that program. The *agegrp* variable breaks the sample

into three groups according to age.   These groups are use to produce custom Table 1U.

14. The *empigrp* variable has five different values depending on the level of employment income.

15. The *dispgrp* variable has ten different values depending on the level of disposable income.

```
[16]    ###
        ## 2.1.15.1 Built-in tables
        ###

        T0FLAG              0               # Table 0 request flag


        ###
[17]    ## 2.1.15.2 User-specified Tabulation facility
        ###

        XTFLAG              1               # X-tab facility activation flag
        XTSPEC                              # X-tab specification
[18]    IN:{eitc, gainer:S=3, nochange:S=3} * agegrp+;
        IN:empigrp+ * {eitc, eitc/gainer:L="Average Benefits",
                    gainer:S=3, nochange:S=3};
        IN:dispgrp+ * {eitc, gainer:S=3, nochange:S=3, scfrecs};


        ###
[19]    ## 2.1.15.3 Distributional Analysis Facility
        ###

        DISTFLAG            1       # Distribution facility activation flag
        DISTUNIT            0       # Distribution facility family level
        DISTVAR         eitc        # Distribution facility variable
        DISTSAMP        200000      # Distribution facility sample size
        DISTZERO            0       # Distribution facility zero inclusion flag
```

16. The first built-in table, which is activated as a default is turned off here.

17. The X-tab facility is used to evaluate the impacts of the new program.

18. The specification shown here produces three tables. Note that the tables have been forced to be tabulated at the individual level of analysis (IN:). In this type of application where benefits are being calculated on an individual basis the user **may** specify a table request at a higher family level and all analysis variables will be rolled up to that level. **However**, the classification variables must be calculated for the desired family level of table. Thus for a NF: table the user would group NF:immdisp rather than immdisp. Also, the age would refer to the head of the nuclear family.

19. The distributional analysis facility is used to explore the shape of the distribution of new EITC benefits. The level of analysis is set to individual benefits (DISTUNIT=0). Because DISTZERO is set to 0, individuals receiving no EITC benefits are excluded from the distributional report.

This file can be found in the \SPSM\EXAMPLE directory. Create a new simulation for 2003, include this parameter file by going to **File**, **Read Parameter File**, in the **Files of Type** box

change the type to **Control Parameter Include Files (*.cpi)**, and then find the file and click Open. The parameters will be read. Save the simulation and execute it. We will now briefly examine the outputs of the run which can be found under the tables tab. They have been reformatted to appear in Classic SPSM format for ease of reading.

Table 1U: Selected Quantities for Individuals by Age

```
+--------------------------------+--------+--------+--------+--------+
|Quantity                        | Min-20 | 21-64  | 65-Max |  All   |
+--------------------------------+--------+--------+--------+--------+
|Earned income tax credit (M)    |    0.0| 1033.1|     3.8|  1036.9|
|Received EITC Flag (Gainer) (000)|   0.0| 1495.8|     7.1|  1502.9|
|Unaffected by EITC Flag (000)   |  7917.9| 6631.1|    18.6| 14567.7|
+--------------------------------+--------+--------+--------+--------+
```

Table 1U seems to indicate that the program has been correctly implemented as specified. No benefits are paid to children aged 20 and under. Total benefits of $1 billion accrue to 1.5 million individuals. There are only 18,600 individuals over the age of 64. This is because the selection facility restricts the sample to census families having children under age 21. In total, only an estimated 16,070,600 (1502.9 + 14567.7) Canadians of any age live in such families in 2003.

Table 2U: Selected Quantities for Individuals by Wages & salaries Group

```
+----------+------------+------------+------------+------------+
|Wages &   |Earned income|  Average   |Received EITC|Unaffected by|
|salaries  | tax credit |  Benefits  |Flag (Gainer)|  EITC Flag  |
|Group     |    (M)     |            |   (000)    |   (000)    |
+----------+------------+------------+------------+------------+
|Min-0     |      209.1|   668.3960|      312.9|     8012.4|
|1-8000    |      339.0|   630.4710|      537.7|     1356.2|
|8001-12000|      207.2|  1070.2387|      193.6|      388.8|
|12001-24000|     281.6|   613.8137|      458.7|      825.8|
|24001-Max |        0.0|     0.0000|        0.0|     3984.4|
+----------+------------+------------+------------+------------+
|All       |     1036.9|   689.9343|     1502.9|    14567.7|
+----------+------------+------------+------------+------------+
```

Table 2U shows the same results as Table 1U except they are now broken out by individual employment earnings level. As we would expect, no benefits are paid to individuals with over $24,000 of earned income. The $209.1 million in EITC benefits paid to individuals having no employment income appears to be an error. Recall, however, that the program as specified is based on family employment income. This means that if an employed person earning between $1 and $24,000 had a spouse with zero earnings, that spouse would be eligible for the benefit as eligibility is based on family income, but the benefit is assigned to each individual.

Notice that the average benefits for the 8001-12000 employment income group is $1,070. Some readers may have expected the number to be $1,200, or maximum benefits. This would be the case if the table were restricted to unattached individuals. Again, this is not the case as the table is based on the income of individuals but the benefits are based on family employment income. The $1,070 is short of the maximum due to married couples who have more than one earner and whose combined income is less than 24,000. In these cases the income of the second earner moves the family employment income into the range where benefits are reduced. If a check of whether the benefit was following the rules set out, then the classification variable ought to have been defined using head and spouse earnings (*nfempinc*).

Table 3U: Selected Quantities for Individuals by Base disposable income group

Base disposable income group	Earned income tax credit (M)	Received EITC Flag (Gainer) (000)	Unaffected by EITC Flag (000)	SLID Records
Min-5000	94.4	175.5	7778.3	18355
5001-10000	193.1	244.5	885.7	2646
10001-15000	236.9	302.8	686.9	2338
15001-20000	218.2	328.5	682.7	2403
20001-25000	136.5	207.4	678.3	2167
25001-30000	70.4	114.7	723.9	2090
30001-35000	31.8	44.8	697.1	1804
35001-40000	20.3	28.1	607.6	1485
40001-45000	5.2	8.1	393.6	979
45001-Max	30.0	48.5	1434.3	3141
All	1036.9	1502.9	14567.7	37407

Table 3U shows the distribution of EITC benefits over individuals by disposable income group. Benefits are paid to persons with high incomes because sources of income other than from employment are not being taken into account in determining benefits.

```
Distribution report: Earned income tax credit for Individuals

   Total observations = 137000
   Zero observations  = 124302
   The following statistics are based on 12698 non-zero observations.

   Descriptive Statistics:

   Sum of weights            =         1502919
   Weighted Sum              =      1036915375
   Weighted Sum of Squares   =    942952714864
   Weighted Mean             =             690

   Extreme Values (with associated household numbers):
```

```
       Minima   hdseqhh                              Maxima   hdseqhh

           0      45501                      1200       157
           0      45501                      1200       157
           0      18203                      1200       158
           0      18203                      1200       158
           0      23942                      1200       159

    Selected Quantiles:

    Q1 = 365           P1=    8    P90=1200    P20= 267    P60= 862
    Med= 717   P2=  15  P95=1200    P30= 434    P70= 982
    Q3 =1065   P5=  48  P98=1200    P40= 574    P80=1122
               P10= 113  P99=1200


Histogram Plot:
```
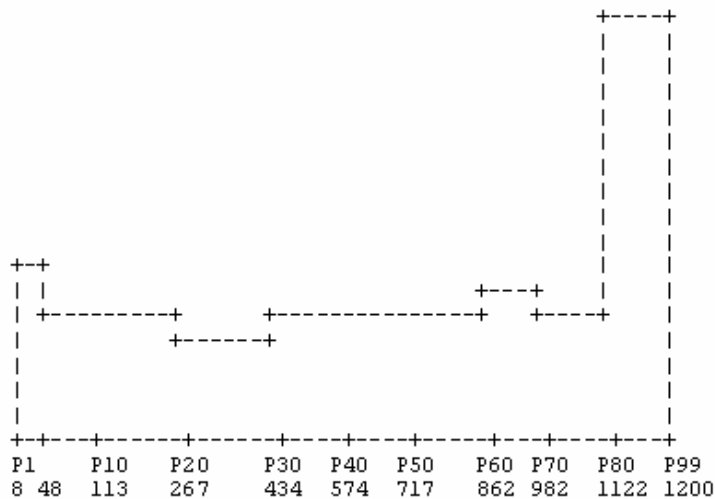


```
P1    P10    P20    P30 P40 P50   P60 P70  P80  P99
8  48  113    267    434 574 717   862 982  1122 1200
```

The distribution report indicates that the program has been implemented correctly. The maximum individual benefit is $1,200 as expected.

Guide to SPSD/M Documentation

SPSD/M INTRODUCTORY MANUAL

The SPSD/M is a complex and sophisticated tool for the analysis of the intricacies of Canada's tax/transfer system. The Introductory Manual is designed to get the user up and running in a minimum amount of time while providing an overview of the elements of the SPSD/M which must be understood to conduct valid research and analysis. The *SPSD/M Introductory Manual* consists of three documents:

How to Run the SPSM

Describes how to run the SPSM using both Visual SPSM and Classic SPSM.

Introduction and Overview

This is the current guide, which is designed to give an overview and introduction to the SPSD/M. It describes the construction of an SPSM simulation run, and also contains several worked examples of model use. Much of the material presented here is also presented in the other guides in a more comprehensive fashion.

Addendum

This document outlines the modifications that the SPSD/M has undergone subsequent to the previous release of the SPSD/M. Users will note that the chapter headings in this Addendum correspond to guides in the manuals.

SPSD/M USER'S MANUAL

The SPSD/M has been designed and created by senior analysts who actively perform simulations using the tool. It contains powerful features which allow a broad range of analysis options. Comprehensive descriptions of these features are contained in the five guides in this manual.

User's Guide

This guide describes how to run the SPSM in black box mode (i.e. without making any changes to the C++ language source code). The SPSM user-facilities are each described in a complete fashion.

User X-tab Facility

This auxiliary guide contains an expanded, more pedagogical treatment of the user-defined X-tab facility which is described in a complete, though terse fashion in the SPSD/M User's Guide.

Programmer's Guide

Intended for use by the glass box user, this guide describes how to make programming changes to the SPSM model.

Commodity Tax Model User's Guide

This guide describes the operation of the Commodity Tax (COMTAX) model. It also contains the mathematical description of the underlying Input-Output model.

Tools User's Guide

This guide describes various tools and utilities designed for use with SPSD/M, such as the spreadsheet interface and parameter file comparison utilities. The currently provided auxiliary tools take the form of other executable programs.

SPSD/M REFERENCE MANUAL

There are three main ingredients to the SPSD/M which are manipulated in various ways to create various tax/transfer scenarios: the database, the algorithms and parameters. This manual provides detailed information on the components of each of these three ingredients.

Algorithm Guide

The Algorithm Guide is designed primarily to provide SPSD/M users with an understanding of the specific personal tax and cash transfer programs which are simulated together with specific information on how each program was implemented. An attempt is also made where possible and practical to guide and caution the user, where necessary, as to the interpretation of the results. The guide is intended for both black box and glass box users.

Parameter Guide

This document provides an explanation of all parameters, including control parameters, that are part of the SPSM. This includes a detailed description of how the parameters are used, their values for the various parameter files provided, and wherever possible, the published source where the values were obtained.

Variable Guide

This document contains reference information on each SPSD/M variable. The first part of this document is organized by subject area. The second part consists of an encyclopedic reference to each SPSD/M variable, and is organized alphabetically. The reference section is for determining the detailed definition, characteristics, and use of a particular variable.

Database Creation Guide

This guide describes the general process, step by step, of constructing the micro-statistically representative SPSD.