



Growth and Validation Guide

This guide is designed to provide SPSPD/M users with information relating to the growth of the SPSPD/M to represent different years and the validation of SPSPD/M results.



Statistics
Canada

Statistique
Canada

Canada

Table of Contents

Introduction.....	1
Projection Parameters	2
Growth Factor Parameter Creation and Projection.....	2
Model Parameter Projection.....	2
SPSD/M Validation.....	3
Version 8.0 Base Year (1996) Validation Table.....	3
Overview.....	3
SCF and SPSM comparison with National Accounts Data	4
Market Income	4
Transfer Income	5
UI/EI Expenditures and Number of Beneficiaries	6
Income Taxes	7
Payroll Taxes.....	8
Commodity Taxes.....	9
Comparisons with Other Micro-data Sources.....	9
Comparisons with Aggregate Time Series Statistics	11
Comparisons with Other Micro-simulation Models	12
Adjusting SPSPD/M Results to Control Totals	13
Post Hoc Adjustment	13
Creating a New Weight File.....	13
Adjusting Growth Factors.....	15
Record Conversion.....	15
Raking Weights.....	16

Introduction

SPSD/M's base year (1997) demographic data comprises a probability sample from provincial populations representing individuals and their living arrangements (i.e., individuals grouped in families and households). The demographic structure of the database is subsequently "grown":

1. By adjusting household weights to bring provincial age-sex distributions into line with Statistics Canada's population estimates for succeeding years, and
2. By simultaneous adjustment of household weights to agree with Labour Force Survey data on annual average unemployment rates and employment to population ratios by age and sex and by province and sex for the years in which survey data are available.

In addition, database "growth" involves adjustment of dollar denominated components of income and tax to account for inflation, real growth and/or other trends affecting aggregate amounts. This is done in two ways:

1. Adjusting individual dollar denominated items in order to bring the aggregate dollar amounts into line with the National Accounts or Human Resources Development Canada data;
2. Adjusting individual dollar denominated items in order to bring the growth rates into line with growth rates in the National Accounts or published estimates.

Static demographic and economic "growth" of this form assumes that the population remains unchanged in relative terms. For example, provincial population growth due to migration is accounted for by adjusting the base year household weights. That is, the characteristics (e.g., family size, income, etc.) of persons already resident in a province are used to represent the characteristics of migrants to that province. Thus, some aspects of the relative distributional impacts of variant tax/transfer models will represent the base year regardless of the nominal year in which they are implemented or the growth parameters used.

Static database adjustments force agreement with aggregate control totals by minimal change (if any) to the underlying micro data. More ambitious attempts to project micro data would have to explicitly assign new characteristics to individuals, families and households. These characteristics would have to represent labour market activity, personal finances and demographic events that are likely in light of 1997 micro data. Moreover, the assigned characteristics have to make sense in combination: migrants often change jobs and income levels and are more likely to be young and unmarried than the general population. Each modification made to micro data will either weaken or distort correlations that should be present or introduce spurious correlations. Keeping microdata changes to a minimum guarantees that the correlations present in the data have an empirical basis.

Projection Parameters

Growth Factor Parameter Creation and Projection

For 1988 to 2000, the year to year change in each growth factor parameter is forced to reproduce the growth of the conceptually closest benchmark series in the System of National Accounts - all on a provincial basis. This technique assures that each database variable never falls out of alignment with the growth of its proxy SNA series. For example, if the SNA farming income series for Ontario grows by 1.2% from 1990 to 1991, then the corresponding database variable, *idsefm*, grows at exactly the same rate for all Ontario residents.

There are two exceptions to this method. CPP/QPP income is adjusted using average benefit by year of age. This data comes from the Department of Human Resources Development Canada. All FAMEX variables are grown using GDP which results in no distortion of the income/expenditure equality.

As SNA data is unavailable for subsequent and future years, we base the growth factors on projections done by the private sector provided to us by Finance Canada. The income parameters are grown by projected personal income. Expenditures are grown by projected GDP.

Model Parameter Projection

Projection of a model parameter only occurs when no official estimate is available. For example, the year 2004 model parameters governing the unemployment insurance benefits have not yet been released by the federal government. Therefore, we must project these parameters for 2004 from the last year in which we have an official estimate.

Finance Canada provided us with projected Average Weekly Industrial Wage, Personal Income, GDP, and CPI.

Users can determine how each model parameter is projected by examining the parameter's accompanying UPDATE statement in the .mpr files. For example, a FACTOR=CPIM3 statement means that the parameter will be projected at the CPI value minus three. Users can examine the *factors.txt* file in the *SPSD* directory for more information on the implicit growth underlying these three series.

SPSD/M Validation

Model validation has always been a critical component in the construction of the SPSPD/M. The validation of results occurs for both the base year and for subsequent years in which the base year data is grown. Base year validation involves a reconciliation with many different sources of detailed data which are readily available. The underlying population is not being subjected to growth methodologies and many conceptual differences have been accounted for.

Validation becomes more difficult in subsequent projection years due to the increasing lack of availability of reference sources as well as the shortcomings of the static aging methodology to account for more variables and interactions. When dealing with simulation in future years, projections can only be validated with other projections. The further out in time one goes, the less a handful of aggregate growth factors applied can accurately account for complex shifting multivariate distributions in a population.

Nevertheless, there is a need for these projections and a need for some agreement on the "accuracy" of resulting estimates. There are three classes of reference standards which can be used for validation in either the database year or subsequent years; Micro-Data, Aggregate Time-series data, and other simulation model results. The first includes other survey and administrative micro-data sources (eg. Census, T4 administrative file). The System of National Accounts is an example of time-series data while the Department of Finances Tax/Transfer model is another microsimulation model.

In any validation exercise using any of the reference classes, one must be sure that the conceptual differences are accounted for (eg. the employment income of the comparison data is conceptually equivalent to the employment income concept on the SPSPD/M). In addition, any ways in which the population underlying the comparison data source differs from the population underlying the SPSPD/M must be accounted for. The next section shows some validation tables for the SPSPD/M 9.0 while the final sections give examples of conceptual and coverage problems by discussing the three reference classes.

Version 9.0 Base Year (1997) Validation Table

Overview

The base year (1997) of Version 9.0 has been validated against a variety of other sources. The results are presented here.

The first table shows how both the Survey of Consumer Finances and the SPSPD/M compare to National Accounts Data. In general, the SPSPM's estimates are closer to the National Account estimates due to adjustments made in the database creation process.

The following tables show how the SPSPD/M compares to other published aggregate sources. Estimates come from the following sources: the Survey of Consumer Finances, the System of National Accounts, the Statistics Relating to Income Security in Canada, Social Security Statistics, and Taxation Statistics. When more than one estimate was available, all estimates are shown. The first two columns

of these tables show the percentage and dollar difference between the SPSD/M and the most comparable data available (these are underlined in the table). There are conceptual differences between the sources.

SCF and SPSM comparison with National Accounts Data

This table shows how both the Survey of Consumer Finance and the SPSD/M compare to the System of National Accounts. It serves to demonstrate the results of database enhancements made to the SPSD/M.

Percent of National Accounts Estimates

		SCF	SPSM
	Market Income		
1	Wages and Salaries	105.9	100.2
2	Farm Self-Employment Income	137.5	97.8
3	Non-Farm Self-employment Income	104.2	101.3
4	Total Earnings	106.0	100.3
5	Investment Income	49.4	107.4
	Transfer income		
6	Government Transfers to Persons	93.4	104.7
8	OAS and GIS	98.7	98.3
11	Child Tax Benefit	98.2	97.3
15	UI benefits	86.4	101.1
16	CPP/QPP income	94.6	99.7
17	GST Tax Credit	96.1	104.5

Notes:

Rows:

5. SPSD/M investment income is higher than the SCF due to the imputations of interest income.

Market Income

This table shows the different estimates for market income. The difference between the SPSD/M and the estimate that most closely matches the SPSD/M concept is shown in the first two columns. The estimates that are compared are underlined in the table.

Comparison of Market Income, 1997 (millions of dollars)

				A	B	C	E	F
		<i>Difference</i>		Source of Estimate				
		%	\$	SCF	SPSM	SNA	T1/T4	Greenbook
1	Wages and Salaries	100.2	933	412,363	390,324	389,391	391,591	384,434
2	Farm Self-employment Income	97.8	-66	4,045	2,876	2,942	2,825	1,873
3	Non-Farm Self-employment Income	101.3	442	35,037	34,052	33,610	37,570	30,088
4	Total Earnings	100.3	1,310	451,444	427,253	425,943		416,395
5	Investment Income	107.4	2,832	18,918	41,094	38,262		45,766

Notes:

Columns

- A The SCF figures are from the 1997 Publication reconciliations.
- B The SPSM figures are for the SPSD/M version 9.0 Release.
- C The Italicized SNA figures are from the 1997 Publication reconciliations and have been adjusted to match the SCF universe and concepts which is used by the SPSD/M.
- E Special Tabulation of T4 files (Row 1) and T1 files (Rows 2-3).
- F Taxation Statistics, Revenue Canada Taxation.

Rows

- 5. SPSD/M investment income is higher than the SCF due to the imputations of interest income.

Transfer Income

This table shows the different estimates for transfer income. The difference between the SPSD/M and the estimate which most closely matches the SPSD/M concept is shown in the first two columns. The estimates that are compared are underlined in the table.

Comparison of Transfer Income, 1997 (millions of dollars)

				A	B	C	D	E	F
		<i>Difference</i>		Source of Estimate					
		%	\$	SCF	SPSM	SNA	Redbook	SSS	Greenbook
	Transfer Income								
6	Government Transfers to Persons	104.7	3,815	75,440	84,547	<i>80,731</i>			
8	OAS and GIS	98.3	-371	21,100	21,284	<i>21,375</i>	21,654		
9	Social Assistance	100.7	83	9,203	12,147	<i>11,433</i>	12,064	14,016	5,844
	Federal Transfers								
11	Child Tax Benefit	97.3	-143	4,997	5,085	<i>5,228</i>	5,134	4,986	
12	OAS	99.9	-24		16,409		16,433		14,607
13	GIS	99.0	-46		4,590		4,636		
14	SPA	98.8	-5		393		398		
15	UI Benefits	105.8	755	11,192	13,763	<i>12,272</i>		13,008	11,724
16	CPP/QPP Income	99.8	-47	19,345	20,914	<i>20,961</i>	21,334		20,158
17	Sales Tax Credit	108.2	228	2,845	3,001	<i>2,625</i>			2,773
	Provincial Transfers								
18	Family Programs	93.2	-69		944	<i>1,013</i>	1,069		
19	Elderly Programs	90.7	-24		235		259		
20	Provincial Tax Credits	103.1	148	862	1,833	<i>1,777</i>			1,684

Notes:

Columns:

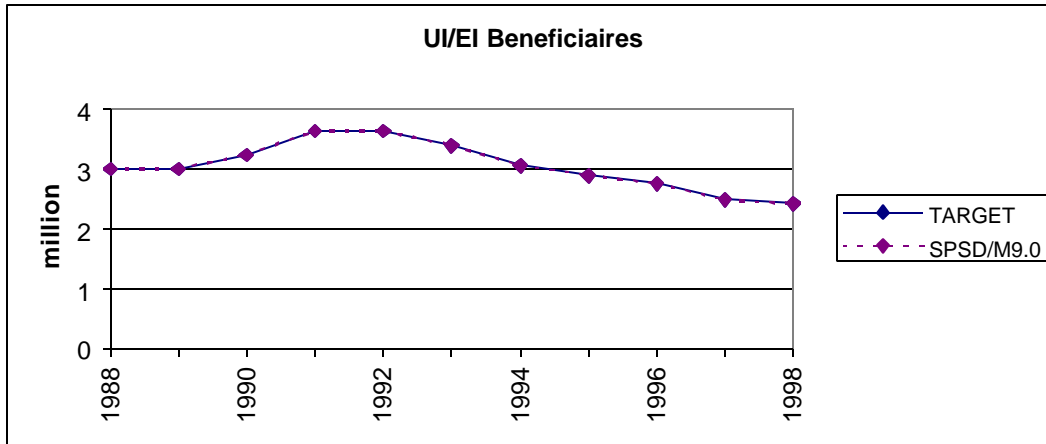
- A The SCF figures are from the 1997 Publication reconciliations.
- B The SPSM figures are for the SPSD/M version 9.0 Release.
- C The Italicized SNA figures are from the 1997 Publication reconciliations and have been adjusted to match the SCF universe and concepts which is used by the SPSD/M.
- D Statistics Related to Income Security Programs, Human Resources Development Canada.
- E From RCT (Row 11) and from CANSIM (Row 15)
- F Taxation Statistics, Revenue Canada Taxation.

Rows:

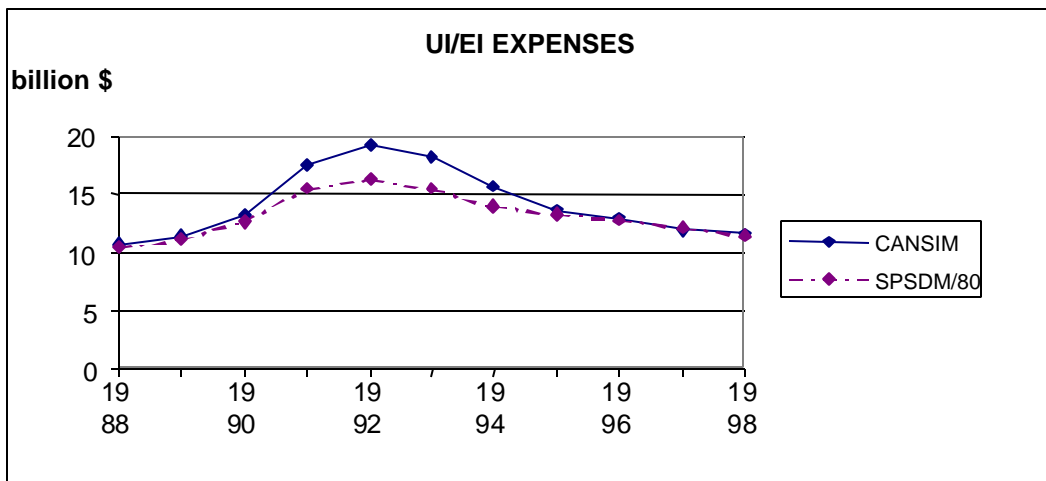
- 9 SNA is direct relief provincial and local.

UI/EI Expenditures and Number of Beneficiaries

In the Figure on UI/EI Beneficiaries, we observe that this release could accurately reproduce the number of UI beneficiaries in all the years. The adjustment factor UITARGET helps us to reach the right number of UI beneficiaries.



The next Figure shows how this release performs in the estimation of UI/EI expenses relative to CANSIM estimates. In this release the 1991 to 1994 expenditures are highly underestimated, mostly because the characteristics of the UI beneficiaries in the 1997 base year are significantly different from those in the 1992 recession.



Income Taxes

This table shows the different estimates for income taxes. The difference between the SPSPD/M and the estimate which most closely matches the SPSPD/M concept is shown in the first two columns. The estimates which are compared are underlined in the table.

Comparison of Income taxes, 1997 (millions of dollars)

				B	C	F
		<i>Difference</i>		<i>Source of Estimate</i>		
		%	\$	SPSM	SNA	Greenbook
	<i>Federal</i>					
21	Total Income assessed	99.9	-560	529,599		530,159
22	Total Deductions	88.9	-5,812	46,404		52,216
23	Taxable Income	101.8	8,630	486,390		477,759
24	Basic Federal Tax	101.2	881	72,206		71,325
25	Federal Individual Surtax	100.0	1	2,819		2,818
26	Federal Income Tax	98.0	-1,456	72,267	<i>67,403</i>	73,722
27	Number of Filers (000)	97.1	-610	20,424		21,034
28	Number of Taxable Filers (000)	103.6	514	14,881		14,367
	<i>Provincial</i>					
29	Québec provincial income taxes	100.5	73	14,384	<i>14,564</i>	14,311
30	Provincial income tax, excl. Québec	97.9	-597	28,435	<i>33,214</i>	29,031

Notes:

Columns:

- B The SPSM figures are for the SPSD/M version 9.0 Release.
- C The Italicized SNA figures are from the 1997 Publication reconciliations and have been adjusted to match the SCF universe and concepts which is used by the SPSD/M.
- F Taxation Statistics, Revenue Canada Taxation.

Rows:

- 21 Both SPSM and Greenbook are restricted to Taxable Filers Only. The Greenbook figure is net of Tax Exempt Income (Social Assistance, Workers Compensation and GIS/SPA).
- 22 Both SPSM and Greenbook are restricted to Taxable Filers Only. The Greenbook figure is net of Tax Exempt Income (Social Assistance, Workers Compensation and GIS/SPA).
- 23 Taxable Filers Only.
- 24 GB has non-provincial netted out (all returns).
- 26 GB does not include Québec non-refundable tax abatement, matches SPSM (imtxf net of imqtaa).
- 29 Québec figures from the Portrait de la fiscalité des particuliers au Québec
- 30 GB does not include Québec taxes.

Payroll Taxes

This table shows the different estimates for payroll taxes. The difference between the SPSD/M and the estimate which most closely matches the SPSD/M concept is shown in the first two columns. The estimates which are compared are underlined in the table.

Comparison of Payroll Taxes, 1997 (millions of dollars)

				B	C	D	F
		<i>Difference</i>		Source of Estimate			
		%	\$	SPSM	SNA	UI stats	Greenbook
31	UIC Contributions	104.8	408.5	8,911	7,808	8,503	8,005
32	CPP/QPP Contributions	109.5	1,275.1	8,620	7,345		7,874

Notes:

Columns:

- B The SPSM figures are for the SPSD/M version 9.0 Release.
- C The Italicized SNA figures are from the 1997 Publication reconciliations and have been adjusted to match the SCF universe and concepts which is used by the SPSD/M.
- D Statistics Related to Income Security Programs, Human Resources Development Canada.
- F Taxation Statistics, Revenue Canada Taxation.

Commodity Taxes

This table shows the different estimates for commodity taxes. The difference between the SPSD/M and the estimate which most closely matches the SPSD/M concept is shown in the first two columns. The estimates which are compared are underlined in the table.

Comparison of Commodity Taxes, 1997 (millions of dollars)

				B	C
		<i>Difference</i>		Source of Estimate	
		%	\$	SPSM	SNA
	<i>Federal</i>				
33	Goods and Services Tax	96.4	-604	16,329	17,632
34	Federal Indirect Taxes	96.6	-290	8,139	8,310
	<i>Provincial</i>				
35	Sales Taxes	96.2	-585	14,674	16,208
36	Other Provincial Commodity Taxes	96.8	-310	9,262	9,392

Notes:

Columns:

- B The SPSM figures are for the SPSD/M version 9.0
- C SNA input-output table derived to match the SPSD/M concept.

Comparisons with Other Micro-data Sources

The main advantage of comparing the SPSD/M against other micro-data sources lies in the ability to examine various demographic and economic distributions in great detail. For example, the Dept. of Revenue's "Greenbook" file, which is a stratified sample of T1 returns, could be used to compare the number of individuals receiving the child tax benefit by income class.

However, the Greenbook, as well as most other micro-data sources, suffer from both a lack of timeliness and conceptual differences in the population on which it provides estimates. The lack of timeliness means that validation exercises can only be performed for historical years while a great deal of policy analysis focuses on the future. If, for instance, one is constructing a new benefit system for the year 1994, most people would prefer to simulate in 1994 (complete with heroic growth assumptions if necessary) and not use, say, 1988 as a proxy. A further problem for non-database years is that the Greenbook provides distributional information for a particular population which is conceptually different from either the SPSM's base year or projected year.

The following example is provided to demonstrate how the conceptual differences in coverage can affect the estimates. By comparing the Wages & Salaries estimates, by province, in 1990 in the Greenbook with the Statistics Canada estimates from Labour Income Division some substantial differences arise. These are outlined in the table below.

Comparison of Wages and Salaries, by Province, 1990 (millions of dollars)

Province	StatCan	Greenbook	Difference	Percentage Difference
Nfld.	4,541	4,530	-11	-0.2
P.E.I.	1,040	1,009	-31	-3.0
N.S.	8,417	8,504	87	1.0
N.B.	6,610	6,460	-150	-2.3
Que.	75,708	72,051	-3,657	-4.8
Ont.	141,780	130,080	-11,700	-8.3
Man.	11,014	10,586	-428	-3.9
Sask.	8,463	8,303	-160	-1.9
Alta.	31,844	30,191	-1,653	-5.2
B.C.	39,813	37,214	-2,599	-6.5
Terr.	1,853	1,201	-652	-35.2
Canada	331,083	310,129	20,954	-6.3

It is clear from this table that the Greenbook does not provide good wage and salary estimates for most provinces. For Canada as a whole, the Greenbook underestimates the Statistics Canada figures by almost \$21 billion or 6.3%. On a provincial basis, the Greenbook estimates for Ontario, Alberta, and British Columbia are off by more than 5% while the estimates for Newfoundland and Nova Scotia are quite accurate. Since there exists no conceptual differences in the definition of wages and salaries in each data source one must conclude that the differences in the estimates stem from the coverage of the Greenbook sample. Therefore, to compare the wages and salaries figures from the Greenbook file to those in the SPSD/M would not be appropriate.

There exist other features of the Greenbook coverage which make comparisons with the SPSD/M difficult. For example, the Greenbook file includes the tax returns of the deceased provided that the estate filed a return. As the SPSD/M contains no deceased it becomes difficult to compare certain estate-sensitive income components such as the income from capital gains.

In conclusion, all micro-data sources suffer from the same kind of limitations that have been outlined above for the Greenbook. These limitations show the need for caution in interpreting results from any validation exercise.

Comparisons with Aggregate Time Series Statistics

An alternative approach to validation is to compare the SPSD/M results with aggregate time series statistics such as those published by Statistics Canada in the System of National Accounts. Although these estimates are much more timely than micro-data sources, there are often serious conceptual and coverage differences in the series. Consider the example described in Table 2 that compares provincial income tax revenues from the System of National Accounts with the Greenbook file.

Table 2 - Comparison of Provincial Income Tax, by Province, 1990

Province	SNA (\$ millions)	Greenbook (\$ millions)	Difference (\$ millions)	Percentage Difference
Nfld.	498	423	-75	-15.1
P.E.I.	116	96	-20	-17.2
N.S.	994	866	-128	-12.9
N.B.	730	623	-107	-14.7
Que.	14,544	n/a	n/a	n/a
Ont.	16,150	13,253	-2,897	-17.9
Man.	1,046	1,140	94	9.0
Sask.	1,084	948	-136	-12.5
Alta.	2,764	2,694	-70	-2.5
B.C.	4,118	3,725	-393	-9.5

The table above shows major differences in provincial income tax estimated from these two data sources. For Ontario alone, the Greenbook estimate is almost \$2.9 billion, or 18% lower than the SNA estimate. In addition, the estimates for every province, with the exception of Alberta, would appear to be poor.

There are, however, conceptual differences in the treatment of refundable tax credits between the two data sources. The SNA income taxes figures come from the government revenue table and are gross of most refundable tax credits. These credits are netted out in the calculation of the provincial balance through a compensating entry on the expenditure side through a transfers payment to persons. The Greenbook has already eliminated the provincial refundable tax credits from its estimate.

There exist two profound conceptual differences between the SNA estimates and the conceptual basis of the SPSD/M. First, the SPSD/M models household cash incomes and cash transactions between households and governments while the SNA estimates only forms of income accrued from production or transfers payments from businesses and governments. Therefore, total personal sector income includes many imputed (non-cash) items such as the employer contributions to pension plans, rental income accruing from owner-occupied dwellings, and imputed interest stemming from the System of National Accounts' treatment of the financial industry. It would not include income from capital gains, income from trusted pension plans, gross income from insurance claims, or alimony income.

Secondly, the personal sector, as defined by the SNA, does not correspond with the household universe underlying the SPSD/M. Charitable, churches and other religious organizations, universities, and trusted pension funds are all components of the SNA personal sector which are not found in the SPSD/M population.

In conclusion, the above discussion makes clear many of the pitfalls of verifying the SPSD/M results with aggregate time series data.

Comparisons with Other Micro-simulation Models

Unfortunately, comparisons with other micro-simulation models of the same nature of the SPSD/M are time-consuming tasks and rarely performed. The high number of assumptions to be reconciled and their interrelated impacts have prevented an exhaustive comparison of two models. The models are under constant development and would change substantially over the time frame of a validation exercise. Moreover, there are only a handful micro-simulation models of this kind with which comparisons could be performed. Ultimately a reconciliation of two models would be very useful especially in future years.

Adjusting SPSD/M Results to Control Totals

There are a number of ways in which users can adjust SPSD/M results to control totals coming from other sources.

Post Hoc Adjustment

In this method, SPSD/M is run using off-the-shelf parameter files, either using files pertaining to the base year or to a projected year. Tables are produced (using UVAR and XTSPEC) and the results turned into proportions. These proportions are then applied to corresponding control totals to give distributional results that are consistent with the imposed control totals.

Consider a scenario that gives the number of gainers and losers by income group. If the total number of families in each income group differs between the SPSM run and a desired benchmark source, then the gainers and losers can be scaled by the proportion of total SPSM counts to total benchmark counts. This is an operation that can be performed most simply using a spreadsheet. Please see the section on the import facility in the *Tools User's Guide* for more information on facilitating the import of data from SPSM output tables into a spreadsheet.

This method can be adapted even if exactly corresponding totals between SPSM and the benchmark data source are not available, by using corresponding indicator variables. For example, if for a particular scenario SPSM gives a certain result for provincial tax, but the provincial employment income is 10% lower compared to a benchmark source, the SPSM provincial tax result can be scaled up by 10%, since *ceteris paribus* taxes follow employment income fairly closely.

Creating a New Weight File

A new weight file can be created if the user wishes to adjust the distribution of the population in SPSD to reflect a structural change. For example, if the Social Assistance population has increased by 20% compared to SPSD/M default results, the user can create a new weight file where the desired population has been increased, and the remainder of the population decreased to maintain a constant total population. This method involves three steps:

1. Perform an SPSD/M run that gives the number of households classified by the variable of interest. In our simple example, we can create a classificatory variable that partitions the population into households that receive Social Assistance and those that do not. The following control parameter settings will produce a table similar to the one illustrated below.

UVAR

```
label(sastat) = "Household SA status";
levels(sastat) = "No SA", "Received SA";
if (HH:imisa > 0) sastat = 1;
else sastat = 0;
```

XTSPEC

HH: sastat+ * {units}

Table 1U: Unit Count (000) for Households by Household SA status

Household SA status	Unit Count (000)
No SA	8862.0
Received SA	1744.0
Both	10606.0

In our example, we wish to scale up the Social Assistance population by 20%. If we wish to maintain the total number of households constant, we must scale down the non-SA population by a factor of 0.803. This factor is derived in the formula below using numbers from the above table.

$$1 - ((1744.0 * 1.20) / 10606.0) = 0.803$$

- Next we must produce an output file that scales the weight of each household by the desired factor and outputs the results to an ASCII output file. The following control parameter settings will accomplish this.

UVAR

```
label(sastat) = "Household SA status";
levels(sastat) = "No SA", "Received SA";
label(newwgt) = "Adjusted household weight";
if (HH:imisa > 0) sastat = 1;
else sastat = 0;
if (idhhrh == 0) {
if (sastat == 1) newwgt = 1.200 * hdwgthh;
else newwgt = 0.803 * hdwgthh;
}
```

```
ASCFLAG      1
OUTASC       saadj.prn
ASCUNIT      4
ASCSTYLE     4
ASCDELIM
ASCEXTPRC   3
ASCVARS      newwgt
```

```
XTFLAG      1
XTSPEC
HH: sastat+ * {units}
```

It is necessary to ensure that newwgt is set for only one member of the household, since the ASCII output facility automatically aggregates analysis variables to the level (in this case household) given by

ASCUNIT. The setting of ASCEXTPRC ensures that the newwgt will retain its fractional information rather than being rounded to an integer.

3. The next step involves converting the ASCII file into an SPSM weight file. The utility bldwgt.exe will perform this operation for us if we issue the following DOS command:

```
c:\> bldwgt 0 saadj.prn saadj.wgt
```

The bldwgt.exe utility is described in more detail in the *SPSM Tools User's Manual*.

The weight file saadj.wgt can now be used in SPSD/M runs using the INPWGT control parameter.

Adjusting Growth Factors

The SPSD is constructed by combining data from a number of sources for a base year. In order to apply this information to subsequent years, dollar-denominated quantities (such as employment income) generally have to be increased to reflect per capita increases due to inflation and productivity changes. Through the database adjustment parameter files (.apr files) users can change these growth factors to force quantities to hit desired benchmark totals. Every dollar-denominated database variable in SPSD/M can be manipulated in this way, and the growth factors can be varied by province.

Note that this technique increases mean amounts per individual or per family. If the changes responsible for the discrepancy between SPSD/M results and the target source result from distributional differences, this approach may lead to false inferences. For example, if total Social Assistance payments need to be adjusted by 20% because the SA population has increased by 20%, this technique would instead increase mean SA benefits by 20% and keep the number of recipients unchanged.

Record Conversion

It is possible to change SPSD data as it is being read in. This technique can be used to selectively modify records that meet certain criteria. For example, a desired proportion of records of low-income families not receiving Social Assistance could be 'converted' into SA recipients by setting a non-zero value to the idisa variable. There are two very different ways of doing this. One way is to use the SPSM 'glass box' mode, which allows the user to add or modify the C language routines that read in data. A special routine, named adju.c, has been expressly provided for this purpose. Please see the [Programmer's Guide](#) for more information on this approach.

Another approach is to use SPSM's reference variable facility. This facility lets the user replace database variables by variables contained in an SPSM results (.mrs). This is a two step procedure. First, SPSM is run to produce a results file containing a modified value of a variable for each person on the database. The following example set of control parameters produce a new variable named newsa that 'converts' 20% of non-SA families with family disposable income less than \$15,000 by giving them \$5,000 in SA income. For all other families, newsa retains its previous value. Note that the \$5,000 benefit is given to the head of the Census Family. This ensures that the benefit is given only once to each qualifying Census Family. The newsa variable is written to a conventionally created

SPSM results file named newsa.mrs.

```
SEED          4
              0
              1
              2
              3
UVARFLAG      1
UVAR
    if (CF:idisa==0 && idrand4<0.20 && idcfrh==0 &&
CF:immdisp<15000) newsa = 5000;
    else newsa = idisa;
OUTMRSFLAG    1
OUTVARMRS     newsa.mrs
OUTMRSVARS    newsa
```

The next step involves using the SPSM reference variable facility to replace the database variable `idisa` with the `newsa` variable from the previously produce results file `newsa.mrs`. The following control parameters accomplish this:

```
REFFLAG      1
INPREF       newsa.mrs
REFVARS      idisa=newsa
```

Any SPSM run that includes these control parameters will use the 'converted' social assistance that was written to the `newsa.mrs` results file. The `REFVARS` control parameter tells SPSM to obtain the `newsa` variable from the file named by the `INPREF` parameter and rename it to `idisa` before performing any tax/transfer calculations. Please see the [User's Guide](#) for more information on the SPSM reference value facility.

Raking Weights

All of the weight files included as part of SPSD/M (with the exception of the base year) have been produced using a generalized version of a technique known variously as 'raking' or 'iterative proportional adjustment'. This technique modifies the weight of each household in the database so as to match desired arbitrary control totals. It is a computationally-intensive technique, and uses additional software programs constructed by the SPSD/M development team that require some time to master. If you wish to explore this option further, please contact an SPSD/M team member.