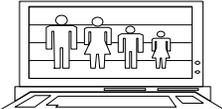




SPSD/M 

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.

June 15, 1995



Canada

Table of Contents

Introduction	2
Projection Parameters.....	3
Growth Factor Parameter Creation and Projection	3
Model Parameter Projection.....	3
SPSD/M Validation.....	4
Beta 6.0 Validation Table	4
Overview	4
SCF and SPSM comparison with National Accounts Data	5
Market Income	7
Transfer Income	8
Income Taxes	10
Payroll Taxes	12
Commodity Taxes	12
Comparisons with Other Micro-data Sources.....	13
Comparisons with Aggregate Time Series Statistics.....	16
Comparisons with Other Micro-simulation Models.....	17
Adjusting SPSPD/M Results to Control Totals	18
Post Hoc Adjustment.....	18
Creating a New Weight File.....	18
Adjusting Growth Factors.....	20
Record Conversion	20
Raking Weights	22

Introduction

SPSD/M's base year (1992) 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 published Revenue 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 Conference Board 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 1992 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 1992 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 changes to micro data to a minimum guarantees that the correlations present in the data have an empirical basis.

Projection Parameters

Growth Factor Parameter Creation and Projection

The methodology underlying the projection of the growth factors found in the .Apr files has been totally revised for this release.

For 1984 to 1996 series, 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.

The sole exception to this method applies to the database employment income variable *idiemp*. In this instance the database variable is forced to equal the SNA series for employment income in each province for the 1984 to 1996 series.

As SNA data is unavailable for subsequent and future years, we base the change in the growth factors on the latest Conference Board forecasts. As these forecasts are not provided at the provincial level, we must settle for movement in series at the Canada level to grow the database variables. All income and tax database parameters are grown by the movement in gross wages and salaries. The expenditure variables are projected on the growth of total personal expenditures. Finally, savings variables are grown based on the movement of aggregate savings in Canada. For these years, the *idiemp* has not been forced to conform with the Conference Board estimate.

Model Parameter Projection

Projection of a model parameter only occurs when no official estimate is available. For example, the 1999 model parameters governing the unemployment insurance benefits have not yet been released by the federal government. Therefore, we must project these parameters for 1999 from the last year in which we have an official estimate.

Three series, for which forecasts are available from the Conference Board, are employed to project model parameters. They are: the nominal Gross Domestic Product, the Consumer Price Index, and the Average Weekly Industrial Wage. 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.prn* 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 Beta 6.0 while the final sections give examples of conceptual and coverage problems by discussing the three reference classes.

Version 6.0 Base Year (1992) Validation Table

Overview

The base year (1992) of Version 6.0 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 SPSPD/M's estimates are closer to the National Account estimates due to adjustments made in the database creation process.

The following tables show how the SPSD/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	99.5	99.8
2	Farm Self-Employment Income	82.2	78.8
3	Non-Farm Self-employment Income	88.0	98.8
4	Total Earnings	98.6	99.5
5	Investment Income	50.4	90.3
	Transfer income		
6	Government Transfers to Persons	82.6	101.3
7	Family allowances	96.3	97.5
8	OAS and GIS	100.9	106.4
11	Child Tax Credit	76.2	91.4
15	UI benefits	81.9	101.1
16	CPP/QPP income	83.5	101.5
17	Sales Tax Credit	89.9	109.4

Notes:

Rows:

- 1 Wages and Salaries are higher than the SCF primarily due to the imputation of high income Canadians. As such the survey tends to over-report wages and salaries. The final release will scale these income back to meet appropriately adjusted SNA totals.
- 2 Farm Self employment income is lower than the SCF because the SPSD takes the value for Farming net income from the Greenbook for High Income filers and replaces the original SCF value.
- 5 SPSD/M investment income is higher than the SCF due to the imputations of interest income
- 8 SPSD/M OAS/GIS income is higher than the SCF because it is modelled on the SPSD/M and not reported as on the SCF. Moreover, the SNA

control totals are lower than the administrative totals which are the conceptually more appropriate comparison.

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, 1992 (millions of dollars)

				A	B	C	F
		<i>Difference</i>		Source of Estimate			
		%	\$	SCF	SPSM	SNA	Greenbo
1	Wages and Salaries	99.8	-628	340,684	<u>341,725</u>	<u>342,353</u>	319,0
2	Farm Self-Employment Income	78.8	-898	3,475	<u>3,331</u>	<u>4,229</u>	1,0
3	Non-Farm Self-employment Income	98.8	-295	21,230	<u>23,828</u>	<u>24,123</u>	20,0
4	Total Earnings	99.5	-1,833	365,389	<u>368,872</u>	<u>370,705</u>	340,0
5	Investment Income	90.3	-4,469	23,138	<u>41,431</u>	<u>45,900</u>	

Notes:

Columns

- A The SCF figures are from the 1992 Publication reconciliations. The figures are based on original weights unadjusted for the 1996 census.
- B The SPSM figures are for the SPSD/M version B6.0, Beta Release and are subject to change in the final release
- C The Italicized SNA figures are from the 1992 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

- 1 Wages and Salaries are higher than the SCF primarily due to the imputation of high income Canadians. As such the survey tends to over-report wages and salaries. The final release will scale these income back to meet appropriately adjusted SNA totals.
- 2 Farm Self employment income is lower than the SCF because the SPSD

5 takes the value for Farming net income from the Greenbook for High Income filers and replaces the original SCF value.
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 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 is underlined in the table.

Comparison of Transfer Income, 1992 (millions of dollars)

				A	B	C	D	E
		<i>Difference</i>		<i>Source of Estimate</i>				
		%	\$	SCF	SPSM	SNA	Redbook	SSS
Transfer Income								
6	Government Transfers to Persons	101.1	926	67,308	<u>82,383</u>	<u>81,457</u>		
7	Family allowances	97.5	-87	3,334	<u>3,375</u>	<u>3,462</u>		
8	OAS and GIS	98.5	-284	17,657	<u>18,673</u>	<u>17,495</u>	<u>18,957</u>	
9	Social Assistance	101.6	207		<u>12,867</u>	<u>12,660</u>	12,864	
Federal Transfers								
10	Federal Family Allowance	98.2	-52		<u>2,855</u>		<u>2,907</u>	
11	Child Tax Credit	92.5	-185	1,884	<u>2,287</u>			
12	OAS	100.4	62		<u>14,354</u>		<u>14,292</u>	
13	GIS	102.2	92		<u>4,319</u>		<u>4,227</u>	
14	SPA	129.1	128		<u>565</u>		<u>438</u>	
15	UI Benefits	98.7	-247	15,142	<u>18,789</u>	18,486		<u>19,036</u>
16	CPP/QPP Income	101.5	236	13,569	<u>16,493</u>	<u>16,257</u>	16,661	
17	Sales Tax Credit	105.9	150	2,282	<u>2,688</u>	<u>2,538</u>		
Provincial transfers								
18	Family Programs				<u>520</u>			
19	Elderly Programs				<u>612</u>			
20	Provincial Tax Credits				<u>1,333</u>			

Notes:

Columns:

A The SCF figures are from the 1992 Publication reconciliations. The

- figures are based on original weights unadjusted for the 1996 census.
- B The SPSM figures are for the SPSD/M version B6.0, Beta Release and are subject to change in the final release
 - C The Italicized SNA figures are from the 1992 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, March 1996, Human Resources Development Canada.
 - E Social Security Statistics, Canada and the Provinces, 1970-71 to 1994-95, Human Resources Development Canada.
 - F Taxation Statistics, Revenue Canada Taxation
- Rows:**
- 9 SNA is direct relief provincial and local

Income Taxes

This table shows the different estimates for income 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 is underlined in the table.

Comparison of Income taxes, 1992 (millions of dollars)

				B	C	F
		<i>Difference</i>		Source of Estimate		
		%	\$	SPSM	SNA	Greenbook
	<i>Federal</i>					
21	Total Income assessed	102.7	12,214	<u>457,647</u>		<u>445,433</u>
22	Total Deductions	97.9	-811	<u>37,832</u>		<u>38,643</u>
23	Taxable Income	103.4	13,801	<u>420,040</u>		<u>406,239</u>
24	Basic Federal Tax	104.6	2,607	<u>59,017</u>		<u>56,410</u>
25	Federal Individual Surtax	104.7	136	<u>3,022</u>		<u>2,887</u>
26	Federal Income Tax	104.6	2,725	<u>61,790</u>	59,622	<u>59,065</u>
27	Number of Filers (000)	97.5	-478	<u>18,866</u>		<u>19,344</u>
28	Number of Taxable Filers (000)	102.9	391	<u>13,877</u>		<u>13,486</u>
	<i>Provincial</i>					
29	Québec provincial income taxes	108.5	968	<u>12,356</u>		<u>11,387</u>
30	Provincial income tax, excl. Québec	104.5	1,104	<u>25,384</u>		<u>24,279</u>

Notes:

Columns:

B The SPSM figures are for the SPSPD/M version B6.0, Beta Release and are subject to change in the final release

C The Italicized SNA figures are from the 1992 Publication reconciliations and have been adjusted to match the SCF universe and concepts which is used by the SPSPD/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:
1992
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 is underlined in the table.

Comparison of Payroll Taxes, 1992 (millions of dollars)

				B	C	D	F
		<i>Difference</i>		Source of Estimate			
		%	\$	SPSM	SNA	Redbook	Greenbook
31	UIC Contributions	107.3	555.6	<u>8,173</u>	<u>7,618</u>		7,271
32	CPP/QPP Contributions	104.4	255.3	<u>6,068</u>	<u>5,813</u>	8,886	5,482

Notes:

Columns:

- B The SPSM figures are for the SPSD/M version B6.0, Beta Release and are subject to change in the final release
- C The Italicized SNA figures are from the 1992 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, March 1996, 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 SPSPD/M concept is shown in the first two columns. The estimates which are compared is underlined in the table.

Comparison of Commodity Taxes, 1992 (millions of dollars)

				B	C
		<i>Difference</i>		Source of Estimate	
		%	\$	SPSM	SNA
	<i>Federal</i>				
33	Goods and Services Tax	75.2	-4,447	<u>13,503</u>	<u>17,950</u>
34	Federal Indirect Taxes	30.4	-21642	<u>9,453</u>	<u>31,095</u>
	<i>Provincial</i>				
35	Sales Taxes	55.9	-9,254	<u>11,741</u>	<u>20,995</u>
36	Other Provincial Commodity Taxes	22.1	-30167	<u>8,548</u>	<u>38,715</u>

Notes:

Columns:

B The SPSM figures are for the SPSPD/M version B6.0, Beta Release and are subject to change in the final release

C The Italicized SNA figures are from the 1992 Publication reconciliations and have been adjusted to match the SCF universe and concepts which is used by the SPSPD/M

Rows:

33 There are no benchmark data conceptually equivalent to GST levied on the household sector. See the Commodity tax users guide for more details

34 Indirect Taxes on The SNA is not equivalent to the SPSPD/M Concept

35 There are no benchmark data conceptually equivalent to sales taxes levied on the household sector. See the Commodity tax users guide for more details

36 Indirect Taxes on The SNA is not equivalent to the SPSPD/M Concept

Comparisons with Other Micro-data Sources

The main advantage of comparing the SPSPD/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 credit by income class and by province.

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 under-estimates 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 SPSPD/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 which 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	na	na	na
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 exists 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 trustee 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 trustee 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 SPSPD/M Results to Control Totals

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

Post Hoc Adjustment

In this method, SPSPD/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 which 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 *SPSPD/M Tools Users 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% low 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 SPSPD to reflect a structural change. For example, if the Social Assistance population has increased by 20% compared to SPSPD/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 SPSPD/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 which 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	8940.0
Received SA	959.2
Both	9899.2

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 0.884. This factor is derived in the formula below using numbers from the above table.

$$1 - ((959.2 * 1.20) / 9899.2) = 0.884$$

- 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 = 0.884 * 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 bldwgt50.exe will perform this operation for us if we issue the following DOS command:

```
c:\> bldwgt50 0 saadj.prn saadj.wgt
```

The bldwgt50.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 *SPSM 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 *SPSM*

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.