CREAP 2001 Version 3.0

Data Documentation and Methods[1]

Trien T. Nguyen
Department of Economics
University of Waterloo

Krista Phillips
Res. & Env. Management
Simon Fraser University

Karen Huff
Tracy Snoddon
Randall M. Wigle[2]
Department of Economics
Wilfrid Laurier University

Revised: September 5, 2008

[1]The authors acknowledge the excellent research assistance by Daniel Pham, Lillian Li, Leo Lui and Hongkun Wu. Many people kindly answered questions about creating this data set, especially Bill Chandler, Maxime Fougère, Ziyad Ghanem, Simon Harvey, Yu Lan, Brent Langen and Ronald Rioux. The authors alone are responsible for any errors.

[2]Correspondence author: Randall M. Wigle, Department of Economics, Wilfrid Laurier University, 75 University Avenue West, Waterloo, Ontario N2L 3C5, Canada. Email: rwigle@wlu.ca.

Contents

List of Tables

List of Figures

Chapter 1

Overview

The new CREAP 2001 Version 3.0 is a microconsistent data set for Canada's provinces built on S-level input-output (IO) tables made available by the Input-Output Division of Statistics Canada (StatCan) [1]. The data set is available for download from CREAP web site (http://creap.wlu.ca).

The plan of this documentation is as follows: chapter 2 describes the provincial energy and emissions data which have been added to the main IO data. Chapter 3 documents various direct and indirect taxes embedded in the data.[1] Additional technical details can be found in the appendices. For example, appendix A presents basic components of the data set such as aggregation commodities, sectors, demand classes, and regional abbreviations. Appendix B explains the process of disaggregating the manufacturing sector into 8 new sub-manufacturing sectors. Finally, appendix C briefly describes how the data set was built from various raw inputs. Additional information is available from the two companion documentations [3] [4].

1.1 Data Balancing

The original S-level data released by StatCan had a large number of cells suppressed due to confidentiality. As a result, the raw IO tables are not balanced. Many sectors earn non-zero profits or in some cases, have no data at all. Similarly, the total availability does not equal the total absorption for several goods. This problem is particularly acute for the Atlantic Provinces and Territories

[1] There are significant details on direct and indirect taxes including public accounts consistent with the data. However, at the time of this writing, limited use of the data is made in the base BMRT model.

The first step in building the CREAP data is to balance the S-level data. In many cases, we have an idea of which cells are missing and which have zero values. We use this information to reduce the number of elements over which to search. Values which are present in the data are assumed to be approximately correct (with a maximum adjustment of $50,000). Provincial data are balanced one-by-one for each individual province, with the trade data being balanced subsequently. Having identified elements which we believe are missing, and those which would seem to be very nearly correct, we then balance the tables using a minimum squared error (MSE) approach.

We understand that a large number of errors might still remain in our balanced S-level data. Usually the problems arise because of missing data from the published tables. We would appreciate any help identifying and fixing apparently inaccurate values in our data. Those who have questions, find errors or have suggestions are invited to contact the correspondence author at the address on the title page.

1.1.1 September 20008 Release (Version 3b)

In the process of using the CREAP Version 3.0 data, an error in the New Brunswick input-output data was discovered. In Statistics Canada's provincial IO data, some cells are suppressed due to confidentiality concerns. This is particularly true of the smaller provinces. There are many suppressed cells in the New Brunswick IO data because of concentration in a number of sectors. To create a balanced data set, values must be assigned to these entries.

The original CREAP (Version 3.0a) balanced data for the New Brunswick Mining Sector included output of $300 million worth of refined petroleum products (RPP) with no inputs of crude oil. An experiment involving an increase in world prices of crude and refined petroleum products caused New Brunswick's mining sector to expand sharply, as it was able to increase production of RPP at no extra cost. The revised data has moved all production of RPP to the Chemical Refining Sector, which has significant inputs of crude oil.

1.2 Manufacturing Disaggregation

A problem of the raw S-level IO data is that there is only one aggregate manufacturing sector in each province. We need to break up this single manufacturing sector into a number of smaller sub-manufacturing sectors. With the help from staff at Human Resources and Social Development Canada

(HRSDC), we were able to obtain additional Census data on sectoral employment and employment income by province to be used in the manufacturing disaggregation in each province [6]. Table A.2 (page 15) provides a list of all sectors in the balanced data set, including 8 disaggregated manufacturing sub-sectors (marked by †) created using knowledge of the pattern of input goods, the pattern of output goods and sectoral employment. Appendix B provides more details about the process of manufacturing disaggregation.

Chapter 2

Energy

This chapter briefly describes the provincial energy and emissions data which were added to the IO structure in the data set. We foresee further refinement of this data in coming releases. In particular, we hope to add greenhouse gas emissions related to combustion of fuels and the production of oil and gas.

2.1 Energy Disaggregation

Since energy goods in the S-level aggregation are highly aggregated, we need to disaggregate S-level data in order to bring these energy goods out. For example, the following S-level goods were disaggregated into energy and non-energy components:

- Mineral fuels (`M_F`) were broken into three sub-components:
  - energy goods: *crude oil* (`CRU`)
  - energy goods: *coal products* (`COL`)
  - energy goods: *natural gas* (`GAS`)

  Since `M_F` has no non-energy components, it was completely replaced by these three newly-created energy goods.
- Petroleum and coal products (`P_C`) were broken into two sub-components:
  - energy goods: *refined petroleum fuels* (`RPP`) include all refined petroleum products plus a modest amount of other products such as tar and lubricating oils
  - non-energy goods: *petroleum and coal products* (`ORP`) include all remaining *non-fuel* petroleum and coal products
- Utilities (`UTL`) were broken into two sub-components:
  - energy goods: *fossil-based electricity* (`ELY`) include all electricity gener-

ation plus water services, sewer services, and electricity distribution
- non-energy goods: *other utilities* (UTL) include all remaining non-energy components of utilities

Table 2.1 summarizes the disaggregation of energy and non-energy goods. The numbers correspond to the orders of commodities in Table A.1 and the marker † denotes commodities which have been disaggregated from one or more existing S-level goods.

Table 2.1: Energy Goods Disaggregation

S-Level	Energy	Non-Energy	Description
M_F	6 CRU		Crude oil †
	7 COL		Coal products †
	8 GAS		Natural gas †
P_C		28 ORP	Petroleum and coal products
	29 RPP		Refined petroleum fuels †
UTL		37 UTL	Other utilities
	38 ELY		Electricity †

The disaggregation is based on number of sources. Production of fossil fuels was disaggregated using NRCan's Mineral and Mining Statistics [2]. Final and intermediate demand was disaggregated using the Comprehensive Energy Use Database [5] and the L-level IO table.

Chapter 3

Taxes

The data set has detailed allocations of both indirect and direct taxes to all transactions in the economy. There are 3 levels of government (see table 3.1)

Table 3.1: Levels of Government

1	F	Federal government
2	P	Provincial government
3	L	Local government

and 4 taxable primary factors (see table 3.2)

Table 3.2: Taxable Primary Factors

1	W_S	Wages and salaries
2	SLI	Supplementary labor income
3	MIX	Mixed income
4	PRF	Other operating surplus (profits)

The data set also has subsidies which were handled in a similar fashion (e.g., negative taxes).

3.1 Indirect Taxes

Indirect taxes are identified in the IO framework as those which apply to products (e.g., inputs or demands) and those which apply to production (e.g., levels of sector activity). The IO Division of StatCan provided additional details on the breakdown of indirect taxes on products, separately identifying GST, provincial sales taxes, gaming taxes and many other categories. These are allocated to the inputs and final demand entries to which they apply. For these taxes, it is also possible to allocate them among the three levels of government. Table 3.3 summarizes the categories of indirect taxes on products identified in the data. Note that the abbreviation codes have been changed for many tax categories.

Table 3.3: Indirect Tax Classes

			Fed	Pro	Loc
1	AL	Alcohol (gallonage) tax		AL_P	
2	AM	Amusement tax		AM_P	AM_L
3	AT	Air transportation tax	AT_F		
4	CD	Customs duties	CD_F		
5	ET	Excise tax	ET_F		
6	GA	Gasoline tax	GA_F	GA_P	
7	GS	Goods and services tax (GST)	GS_F		
8	HS	Harmonized sales tax (HST)		HS_P	
9	RS	Retail sales tax		RS_P	RS_L
10	TP	Trading profits (gaming, tobacco, liquor)	TP_F	TP_P	

3.2 Direct Taxes

Total revenues by province from income and corporate taxes (federal and provincial) were identified using the provincial economic accounts. Average income and corporate tax rates were calculated by dividing the total tax collections by the applicable base. These do not reflect any sector-specific characteristics of the corporate tax. It was further assumed that a small part of mixed income was corporate income. As a result, the corporate tax rates are correspondingly lower than for 'Other Operating Surplus' (PRF).

Table 3.4 shows the calculated tax rates for each province and territory (see table A.3 for the regional abbreviations).

Table 3.4: Direct Tax Rates

Direct Tax Rates (%)	NF	W_S	SLI	MIX	PRF
Income tax	Federal	7.93	7.93	7.93	7.93
Income tax	Provincial	5.17	5.17	5.17	5.17
Capital (income) tax	Federal	0.00	0.00	1.55	3.87
Capital (income) tax	Provincial	0.00	0.00	0.64	1.59

Direct Tax Rates (%)	PE	W_S	SLI	MIX	PRF
Income tax	Federal	8.66	8.66	8.66	8.66
Income tax	Provincial	5.11	5.11	5.11	5.11
Capital (income) tax	Federal	0.00	0.00	2.00	5.01
Capital (income) tax	Provincial	0.00	0.00	0.78	1.95

Direct Tax Rates (%)	NS	W_S	SLI	MIX	PRF
Income tax	Federal	9.60	9.60	9.60	9.60
Income tax	Provincial	5.81	5.81	5.81	5.81
Capital (income) tax	Federal	0.00	0.00	2.34	5.86
Capital (income) tax	Provincial	0.00	0.00	0.90	2.24

Direct Tax Rates (%)	NB	W_S	SLI	MIX	PRF
Income tax	Federal	8.70	8.70	8.70	8.70
Income tax	Provincial	5.09	5.09	5.09	5.09
Capital (income) tax	Federal	0.00	0.00	1.88	4.70
Capital (income) tax	Provincial	0.00	0.00	0.89	2.22

Direct Tax Rates (%)	QC	W_S	SLI	MIX	PRF
Income tax	Federal	8.16	8.16	8.16	8.16
Income tax	Provincial	8.78	8.78	8.78	8.78
Capital (income) tax	Federal	0.00	0.00	2.97	7.44
Capital (income) tax	Provincial	0.00	0.00	1.41	3.52

Direct Tax Rates (%)	ON	W_S	SLI	MIX	PRF
Income tax	Federal	10.99	10.99	10.99	10.99
Income tax	Provincial	4.85	4.85	4.85	4.85
Capital (income) tax	Federal	0.00	0.00	3.26	8.14
Capital (income) tax	Provincial	0.00	0.00	1.73	4.33

Direct Tax Rates (%)	MB	W_S	SLI	MIX	PRF
Income tax	Federal	8.67	8.67	8.67	8.67
Income tax	Provincial	5.44	5.44	5.44	5.44
Capital (income) tax	Federal	0.00	0.00	1.79	4.47
Capital (income) tax	Provincial	0.00	0.00	0.87	2.18

Direct Tax Rates (%)	SK	W_S	SLI	MIX	PRF
Income tax	Federal	7.77	7.77	7.77	7.77
Income tax	Provincial	4.48	4.48	4.48	4.48
Capital (income) tax	Federal	0.00	0.00	1.46	3.65
Capital (income) tax	Provincial	0.00	0.00	0.66	1.65

Direct Tax Rates (%)	AB	W_S	SLI	MIX	PRF
Income tax	Federal	8.15	8.15	8.15	8.15
Income tax	Provincial	3.03	3.03	3.03	3.03
Capital (income) tax	Federal	0.00	0.00	1.96	4.89
Capital (income) tax	Provincial	0.00	0.00	1.10	2.75

Direct Tax Rates (%)	BC	W_S	SLI	MIX	PRF
Income tax	Federal	9.68	9.68	9.68	9.68
Income tax	Provincial	4.37	4.37	4.37	4.37
Capital (income) tax	Federal	0.00	0.00	2.19	5.46
Capital (income) tax	Provincial	0.00	0.00	0.83	2.07

Direct Tax Rates (%)	YT	W_S	SLI	MIX	PRF
Income tax	Federal	6.48	6.48	6.48	6.48
Income tax	Provincial	2.81	2.81	2.81	2.81
Capital (income) tax	Federal	0.00	0.00	1.95	4.88
Capital (income) tax	Provincial	0.00	0.00	0.74	1.86

Direct Tax Rates (%)	NW	W_S	SLI	MIX	PRF
Income tax	Federal	4.93	4.93	4.93	4.93
Income tax	Provincial	2.09	2.09	2.09	2.09
Capital (income) tax	Federal	0.00	0.00	3.89	9.72
Capital (income) tax	Provincial	0.00	0.00	1.56	3.89

Direct Tax Rates (%)	NU	W_S	SLI	MIX	PRF
Income tax	Federal	7.21	7.21	7.21	7.21
Income tax	Provincial	3.01	3.01	3.01	3.01
Capital (income) tax	Federal	0.00	0.00	1.51	3.78
Capital (income) tax	Provincial	0.00	0.00	1.34	3.36

Bibliography

[1] Input-Output Division (2005). *The Input-Output Structure of the Canadian Economy 2000-2001*. Technical Report (Catalogue Number 15-201-XIE), Statistics Canada.

[2] Minerals and Mining Statistics Division (2007). *Mineral Production of Canada, by Province and Territory*. Online statistics accessed October 22, 2007 http://mmsd1.mms.nrcan.gc.ca/mmsd/production/production_e.asp, Natural Resources Canada.

[3] Nguyen, Trien T., Tracy Snoddon, and Randall M. Wigle (2007). "BMRT 2001 Verion 3.0: Model Documentation." Technical Report (October 2007), University of Waterloo and Wilfrid Laurier University.

[4] Nguyen, Trien T. and Randall M. Wigle (2007). "BMRT 2001 Verion 3.0: User Guide." Technical Report (October 2007), University of Waterloo and Wilfrid Laurier University.

[5] Office of Energy Efficiency (2007). *Comprehensive Energy Use Database*. Online statistics accessed May–August, 2007. http://www.oee.nrcan.gc.ca/corporate/statistics/neud/dpa/home.cfm?attr=0, Natural Resources Canada.

[6] Statistics Canada (2007). *Employment and Employment Income by Province and Sector: Custom Tables from the 2001 Census Provided to Human Resources and Skills Development Canada*. Technical Report, Statistics Canada.

Appendix A

Reference Tables

A.1 Commodities

Table A.1 lists 57 commodities with abbreviations and descriptions (see \creap\defines\sets01.gms and \creap\defines\sets-cb-01.gms). Two commodities of the previous version become obsolete and have been dropped, namely, non-competing imports (NCI) and unallocated imports and exports (UMX). The marker † denotes commodities which have been disaggregated from one or more existing S-level goods. In particular, the following new commodities are worth-noting:

- Mineral fuels (M_F) were broken and replaced by three individual components, namely, crude oil (CRU), coal products (COL), and natural gas (GAS).
- Petroleum and coal products (ORP) now contains only *non-fuel* components as *fuel* components become a new commodity, namely, refined petroleum fuels (RPP).
- Fossil-based electricity now becomes a separate commodity (ELY).

Table A.1: List of Commodities

1	GRA	Grains
2	OAG	Other agricultural products
3	FRS	Forestry products
4	F_T	Fish, seafood and trapping products
5	ORE	Metal ores and concentrates
6	CRU	Crude oil †

List of Commodities (cont.)

7	COL	Coal products †
8	GAS	Natural gas †
9	NMM	Non-metallic minerals
10	S_M	Services incidental to mining
11	MFD	Meat, fish and dairy products
12	FVF	Fruit, vegetables and other food products, feeds
13	BEV	Soft drinks and alcoholic beverages
14	TOB	Tobacco and tobacco products
15	LRP	Leather, rubber and plastic products
16	TEX	Textile products
17	CLO	Hosiery, clothing and accessories
18	LUM	Lumber and wood products
19	F_F	Furniture and fixtures
20	P_P	Wood pulp, paper and paper products
21	PRP	Printing and publishing
22	MET	Primary metal products
23	OMP	Other metal products
24	MEQ	Machinery and equipment
25	MVP	Motor vehicles, other transport equipment and parts
26	ELE	Electrical, electronic and communications products
27	NMP	Non-metallic mineral products
28	P_C	Non-fuel petroleum and coal products
29	RPP	Refined petroleum fuels †
30	CPP	Chemicals pharmaceuticals and chemical products
31	MPO	Other manufactured products
32	RES	Residential construction
33	NRC	Non-residential construction
34	REP	Repair construction
35	TRS	Transportation and storage
36	COM	Communications services
37	UTL	Other utilities

List of Commodities (cont.)

38	ELY	Electricity (fossil) †
39	W_M	Wholesaling margins
40	R_M	Retailing margins
41	REN	Gross imputed rent
42	FIR	Other finance, insurance and real estate services
43	BSV	Business and computer services
44	PED	Private education services
45	HSS	Health and social services
46	A_M	Accommodation services and meals
47	O_S	Other services
48	TRM	Transportation margins
49	OPS	Operating, office, cafeteria and laboratory supplies
50	TEA	Travel, entertainment, advertising and promotion
51	NPI	Non-profit institutions serving households
52	GOV	Government sector services
53	OGS	Sales of other government services
54	W_S	Wages and salaries
55	SLI	Supplementary labor income
56	MIX	Mixed income
57	PRF	Other operating surplus

A.2 Sectors

Table A.2 lists 33 aggregated sectors with their abbreviations and descriptions. Details are in the GAMS source files \creap\defines\sets01.gms and \creap\defines\sets-cb-01.gms.

There are 25 original sectors (known as S1, ..., S25 in the old version) plus 8 new manufacturing sub-sectors (marked by †) created by breaking up the original manufacturing sector S8. Data for these 8 manufacturing sub-sectors were calculated using a minimum-squared-error disaggregation procedure (see appendix B), rather than obtained directly from raw data sources (which were non-existent or unavailable). The residuals of the original manufacturing sector S8 are now used for the territories only.

Table A.2: List of Sectors

1	S1	AGR	Crop and animal production
2	S2	FLG	Forestry and logging
3	S3	FHT	Fishing, hunting and trapping
4	S4	SAF	Support activities for agriculture and forestry
5	S5	MIN	Mining and oil and gas extraction
6	S6	UTL	Utilities
7	S7	CON	Construction
8	S8	MFG	Aggregated manufacturing (for territories only)
9		FDBEV	Food beverages and feeds †
10		CLTEX	Clothing footwear, textiles †
11		WPPUB	Wood, paper, printing, publishing †
12		CHREF	Chemicals and refining †
13		PRMIN	Plastic rubber mineral prod †
14		METPR	Metals and metal products †
15		MEQVP	Machinery equipment vehicles and parts †
16		FFMIS	Furniture fixtures and miscellaneous †
17	S9	WHL	Wholesale trade
18	S10	RET	Retail trade
19	S11	TRW	Transportation and warehousing
20	S12	ICI	Information and cultural industries
21	S13	FIR	Finance, insurance, real estate and renting and leasing
22	S14	PST	Professional, scientific and technical services
23	S15	ASW	Administrative and support, waste management
24	S16	EDN	Education services
25	S17	HSA	Health care and social assistance
26	S18	AER	Arts, entertainment and recreation
27	S19	AFS	Accommodation and food services
28	S20	OSV	Other services (except public administration)
29	S21	MOS	Operating, office, cafeteria and laboratory supplies
30	S22	MTA	F2 travel, entertainment, advertising and promotion
31	S23	MTR	F3 transportation margins

List of Sectors (cont.)

32	S24	NPI	Non-profit institutions serving households
33	S25	GOV	Government sector

A.3 Regions

Table A.3 lists 15 regions in the data set with their abbreviations and descriptions (see file \creap\defines\sets01.gms). There are the usual 10 provinces, 3 territories, and two aggregate entities, namely, Canada as a whole and the rest of the world (ROW). The following changes are worth-noting:

- The combined region NN of Northwest Territories and Nunavut in the old version becomes obsolete and is replaced by two separate regions, namely, NW and NU.
- Canada is abbreviated as CD (instead of CANADA in the old version).

Table A.3: List of Regions

1	NF	Newfoundland
2	PE	Prince Edward Island
3	NS	Nova Scotia
4	NB	New Brunswick
5	QC	Québec
6	ON	Ontario
7	MB	Manitoba
8	SK	Saskatchewan
9	AB	Alberta
10	BC	British Columbia
11	YT	Yukon Territories
12	NW	Northwest Territories
13	NU	Nunavut
14	CD	Canada
15	W	Rest of World

A.4 Demand Classes

Table A.4 lists 4 aggregated final demand classes in the data set with their abbreviations and descriptions. Compared to the previous version, the number of final demand classes is now smaller and their abbreviation codes have been changed.

Table A.4: List of Demand Classes

1	`PEX`	Personal expenditures
2	`INV`	Investment expenditures
3	`GOV`	Government expenditures
4	`GIC`	Government (public) investment

Appendix B

Manufacturing Disaggregation

This chapter outlines the process of manufacturing disaggregation used in this version of the data set. As mentioned in section 1.2, the need for this procedure arises from the fact that currently the S-level IO data, while useful in giving provincial breakdowns, allow only one aggregate manufacturing sector in each province. Therefore, we need to have a way to break up this aggregate manufacturing sector into 8 sub-sectors for each province.

B.1 Overview

The general idea is to use additional information from the L-level IO data (national) and occupation/employment data (provincial) to make appropriate scalings based on patterns of inputs, outputs and sectoral employment while still maintaining consistency requirements such as zero profits and add-up constraints. Penalty functions are defined in terms of minimum squared errors (MSE) and the process of manufacturing disaggregation is formulated as an NLP problem. An optimal solution to this problem provides the best sub-sectoral allocation within acceptable tolerances.[1]

[1] In the future, it might be possible to use the energy information as well as the employment income to generate a more accurate split of the manufacturing sector.

B.2 Disaggregation Procedure

The source file \creap\build\creap-base-disagg.gms contains details of the implementation of the manufacturing disaggregation procedure. This section provides a roadmap of the procedure for the benefits of the interested reader. The section can be safely skipped by those not interested in reading technical details and the GAMS code.

1. **Read S-Data (Provincial)**
   - Index sets

index	description	dimension
$p \in$ P	S-data provinces	10 provinces
$sg \in$ SG	S-data goods	54 goods
$ss \in$ SS	S-data sectors (S8: manufacturing)	25 sectors
$ns \in$ NS	S-data new manufacturing sectors	8 sectors

   - Read \creap\data\io-2001\IO-2001-TDa-S.gdx for S-level IO data on intermediate inputs and sectoral outputs *for each province*.

S-data	description	dimension
$\texttt{INS}_{(p,sg,ss)}$	S-data intermediate inputs	$p \times sg \times ss$
$\texttt{OUS}_{(p,sg,ss)}$	S-data sectoral outputs	$p \times sg \times ss$

2. **Read L-Data (National)**
   - Index sets

index	description	dimension
$lg \in$ LG	L-data goods	472 goods + total
$ls \in$ LS	L-data sectors	119 sectors + total
$lms \in$ LMS	L-data *manufacturing* sectors	45 sectors

   - Read \creap\data\io-2001\io-2001-BAL-L.gdx for L-level IO data on intermediate inputs and sectoral outputs.

L-data	description	dimension
$\texttt{INL}_{(lg,ls)}$	L-data intermediate inputs	$lg \times ls$
$\texttt{OUL}_{(lg,ls)}$	L-data sectoral outputs	$lg \times ls$

   - Restrict attention to L-data manufacturing sectors only

$$\texttt{INLMFG}_{(lg,lms)} = \texttt{INL}_{(lg,lms)} \tag{B.1}$$

$$\texttt{OULMFG}_{(lg,lms)} = \texttt{OUL}_{(lg,lms)} \tag{B.2}$$

3. **Aggregate L-Data to S-Data**

   - Run aggregation utility \creap\inclib\aggr.gms with
     · target sets \creap\defines\target1.gms
     · mappings \creap\defines\map-ls-01.gms
     to convert L-data (size 472×45) into S-data (size 54×8).

$$\texttt{INLMFG}_{(lg,lms)} \xrightarrow{\texttt{aggr.gms}} \texttt{INPTN}_{(sg,ns)}$$
$$\texttt{OULMFG}_{(lg,lms)} \xrightarrow{\texttt{aggr.gms}} \texttt{OUPTN}_{(sg,ns)}$$

S-data	description	dimension
$\texttt{INPTN}_{(sg,ns)}$	S-data disaggregated input patterns	$sg \times ns$
$\texttt{OUPTN}_{(sg,ns)}$	S-data disaggregated output patterns	$sg \times ns$

4. **Read Occupation/Employment Data (Provincial)**

   - Index sets

index	description	dimension
$o \in \texttt{O}$	occupational categories	26 categories + total
$lab \in \texttt{LAB} \subseteq \texttt{SG}$	primary labor factors	2 factors (W_S, SLI)

   - Read \creap\data\io-2001\occn-cb-01.gdx for *provincial* occupation and employment data.

S-data	description	dimension
$\texttt{OCCN}_{(p,o,ns)}$	employment by occupation	$p \times o \times ns$
$\texttt{EINC}_{(p,o,ns)}$	employment income by occupation	$p \times o \times ns$

   - Calculate shares **eshr** of employment incomes in manufacturing sector totals and shares **wshr** of labor incomes in total costs

$$\texttt{eshr}_{(p,ns)} = \frac{\sum_{o} \texttt{EINC}_{(p,o,ns)}}{\sum_{ns}\sum_{o} \texttt{EINC}_{(p,o,ns)}} \in [0,1] \tag{B.3}$$

$$\texttt{wshr}_{ns} = \frac{\sum_{lab} \texttt{INPTN}_{(lab,ns)}}{\sum_{ns} \texttt{INPTN}_{(sg,ns)}} \in [0,1] \tag{B.4}$$

   and target scale parameters **scaletgt** in manufacturing sector

$$\texttt{scaletgt}_{(p,ns)} = \left(\frac{\texttt{eshr}_{(p,ns)}}{\texttt{wshr}_{ns}}\right) \sum_{lab} \texttt{INS}_{(p,lab,\texttt{S8})} \tag{B.5}$$

- Calculate targets `in_tgts` for intermediate inputs and `ou_tgts` for sectoral outputs in manufacturing sector

$$\texttt{in\_tgts}_{(p,sg,ns)} = \texttt{scaletgt}_{(p,ns)} \left(\frac{\texttt{INPTN}_{(sg,ns)}}{\sum_{sg} \texttt{INPTN}_{(sg,ns)}} \right) \tag{B.6}$$

$$\texttt{ou\_tgts}_{(p,sg,ns)} = \texttt{scaletgt}_{(p,ns)} \left(\frac{\texttt{OUPTN}_{(sg,ns)}}{\sum_{sg} \texttt{OUPTN}_{(sg,ns)}} \right) \tag{B.7}$$

5. **Objective Function For Each Province**

- Set up a big loop running through all provinces $p \in \texttt{P}$

variable	description	dimension
$\texttt{IN}_{(sg,ns)}$	disaggregated input matrix (solution)	$sg \times ns$
$\texttt{OU}_{(sg,ns)}$	disaggregated output matrix (solution)	$sg \times ns$

- For each province, calculate penalty functions `penaltyI` for intermediate inputs and `penaltyO` for sectoral outputs as sums of squared errors (SSE) between the solution matrices `IN`, `OU` and their corresponding targets `tgti`, `tgto` (which are exactly the same as `in_tgts`, `ou_tgts` defined in equations B.6, B.7)

$$\texttt{penaltyI} = \sum_{ns} \sum_{sg} \texttt{ai}_{sg} \left(\frac{\texttt{IN}_{(sg,ns)} - \texttt{tgti}_{(sg,ns)}}{\texttt{tgti}_{(sg,ns)}} \right)^2 \tag{B.8}$$

$$\texttt{penaltyO} = \sum_{ns} \sum_{sg} \texttt{ao}_{sg} \left(\frac{\texttt{OU}_{(sg,ns)} - \texttt{tgto}_{(sg,ns)}}{\texttt{tgto}_{(sg,ns)}} \right)^2 \tag{B.9}$$

  where `ai`, `ao` are some pre-determined weights (set to unity in the `GAMS` code).

- For each province, the objective function can now be defined as the sum of both input and output components

$$\texttt{penalty} = \texttt{penaltyI} + \texttt{penaltyO} \tag{B.10}$$

6. **Constraints For Each Province**

- For each province, the following consistency constraints must be satisfied:

(a) *zero profit constraints*: for each newly disaggregated manufacturing sector $ns \in \texttt{NS}$, the total value of outputs must equal the total value of inputs

$$\sum_{sg} \texttt{OU}_{(sg,ns)} - \sum_{sg} \texttt{IN}_{(sg,ns)} = 0 \tag{B.11}$$

(b) *add-up constraints*: for each goods $sg \in \texttt{SG}$, the total value of inputs/outputs summed over all newly disaggregated manufacturing sectors must equal the value of aggregated inputs/outputs in the original matrix before the break-up

$$\sum_{ns} \texttt{IN}_{(sg,ns)} = \texttt{INS}_{(p,sg,\texttt{S8})} \tag{B.12}$$

$$\sum_{ns} \texttt{OU}_{(sg,ns)} = \texttt{OUS}_{(p,sg,\texttt{S8})} \tag{B.13}$$

7. **Optimization Problem For Each Province**

- For each province, the manufacturing disaggregation procedure can be formulated as an NLP problem of finding matrices `IN` of disaggregated intermediate inputs and matrices `OU` of disaggregated sectoral outputs such that the objective function (B.10) is minimized subject to consistency constraints (B.11, B.12, B.13).

Appendix C

Data Assembly Process

This chapter describes the files and programs used to assemble this version of the CREAP data set. Section C.1 discusses some practical challenges in converting numerous raw data into proper formats required for the data assembly process. Section C.2 provides brief explanations on the location of files and utility programs.

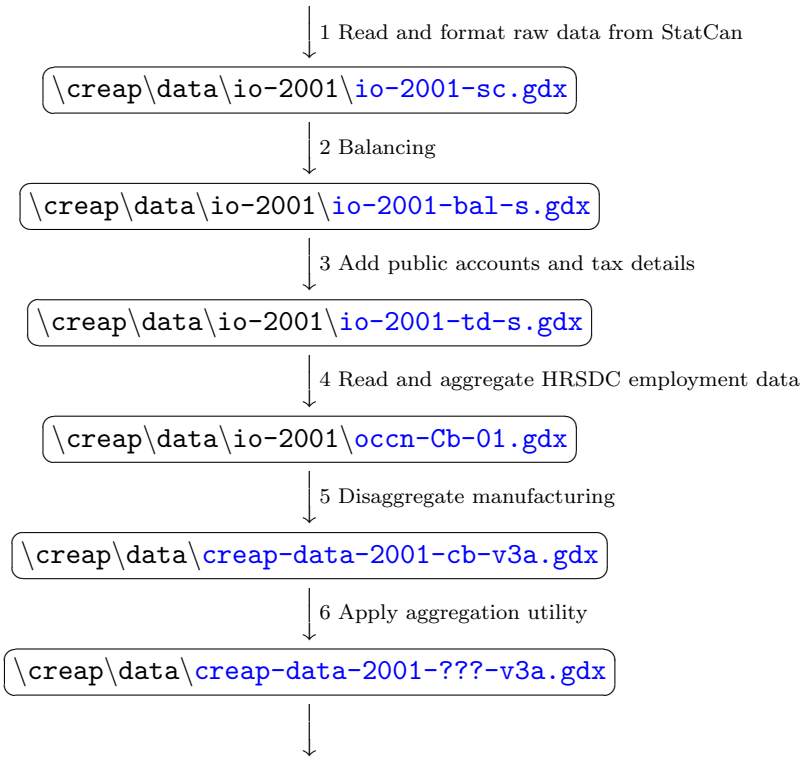
Not all of these files are included in the standard distribution package. In particular, the GAMS programs and source data mentioned in items 4 and 5 of section C.2 (page 25) are *not* included with the normal data distribution.

C.1 Data Input & Formatting

The provincial IO data poses a number of practical challenges in terms of entering it into a GAMS readable format. Although key data of intermediate inputs (IN), sectoral outputs (OU), final demands (FD) and trade flows (TR) are available in IVT files, the large number of tables to be processed make it a daunting task. Altogether, there are 174 tables in total: 39 tables *each* for IN, OU, and FD, plus 57 tables for TR.

Each of these tables has to be first manually saved as a spreadsheet and then have the GAMS row and column identifiers added. The latter process can not be easily automated because of inconsistencies in some of the elements. Also, some of the cells include invalid entries including various non-printable characters. The situation for the indirect tax details is even much worse due to the sheer number of individual tables.

Figure C.1: Visual Overview of the Data Assembly Process



C.2 File Locations

Figure C.1 provides a visual overview of the data assembly process. Data-related files generally reside in one of the following subfolders in the main directory tree of the distribution package:

- `\creap\data` pre-compiled data files in `gdx` or `xls` format
- `\creap\build` programs to rebuild pre-compiled data files
- `\creap\defines` source files for definitions, sets, and maps
- `\creap\inclib` source files belonging to include library

Note that the listed files might not be all available in the distribution package. Since this is still an on-going process, some of the steps are given cursory discussions below.

1. **Read and Format Basic IO Data**

- creap\data\io-2001\read-prov-01.bat
  - call various programs for reading and formatting basic IO data
  - produce output file \creap\data\io-2001\io-2001-sc.gdx

2. **Balancing**
   - creap\build\balance-s-01.gms
     - GAMS program to perform the data balancing
     - produce output file \creap\data\io-2001\io-2001-bal-s.gdx
   - creap\build\build-v3.mak
     - MAKE script to drive other programs and perform housekeeping chores

3. **Add Public Accounts and Tax Details**
   - creap\build\tax-public-01.gms
     - read balanced S-data \creap\data\io-2001\io-2001-bal-s.gdx
     - add public accounts and tax details
     - produce output file \creap\data\io-2001\io-2001-td-s.gdx

4. **Read and Aggregate Census/HRSDC Employment Data**
   The location is the subfolder creap\data\io-2001\occn. Most of the files read StatCan employment data and aggregate them to our IO classifications. They are accompanied by flexible aggregation utilities to allow arbitrary aggregations of the goods and sectors.
   - creap\data\io-2001\occn\census-hrsdc-2000.xls
     - read by gdxxrw
     - produce census-hrsdc-2000.gdx below
     - see census-hrsdc-2000.txt
   - creap\data\io-2001\occn\census-hrsdc-2000.gdx
     - read by census-hrsdc.gms
     - aggregate NAICS sectors to CREAP-base sectors
     - produce occn-Cb-01.gdx below
   - creap\data\io-2001\occn-Cb-01.gdx
     - provide employment and employment income broken down by province, CREAP-base sector, and occupation class
     - essential in the disaggregation of manufacturing (see chapter B)

5. **Disaggregate Manufacturing**
   - creap\build\creap-base-disagg.bat
     - batch file to run GAMS and generate a LaTeX listing
   - creap\build\creap-base-disagg.gms

- read S-data, L-data, and employment income data from HRSDC
- use Tom Rutherford's aggregation utility `creap\inclib\aggr.gms` to aggregate L-level data to S-level commodities and `CREAP`-base sectors (i.e., S-level sectors with manufacturing broken in to 8 subsectors)
- produce `creap\data\creap-data-2001-cb-v3a.gdx`
- create LaTeX files in the `creap\doc` folder

6. **Apply Aggregation Utility**
   This program aggregates a given data set based on pre-specified target sets and map files. It can however handle only simple aggregation rules in which a given source good/sector is allocated to just one target good/sector.

   - `creap\defines\???.set`
     - `GAMS` files to specify *target sets* (e.g., provinces, goods, sectors) drawn from `CREAP`-base data (or possibly `TD-S` data)
   - `creap\defines\???.map`
     - `GAMS` files to specify mappings from `CREAP`-base sets to target sets
   - `creap\inclib\aggr.gms`
     - Tom Rutherford's aggregation utility (required)
   - `creap\inclib\chktarget.gms`
     - debug routine for Tom Rutherford's aggregation utility (required)
   - `creap\inclib\checkset.gms`
     - debug routine for Tom Rutherford's aggregation utility (required)
   - `creap\build\creap-agg-v3.gms`
     - main program to perform the aggregation based on given target sets and map files
     - produce output file `creap\data\creap-data-2001-???-v3a.gdx` the suffix a after the version identifier indicates that the data set includes an index set for labor (`SET LAB`)
     - produce spreadsheets for `IN`, `OU`, `FD`, `TR` at the same time

7. **Disaggregate Skills**
   Once we have a set and mapping file specifying the target goods and sectors (see item 6 above), we can add a specification of the disaggregated classes of labor so that a data set with disaggregated labor inputs can be created.

   The procedure is however a bit indirect: first aggregate the data in the usual way. Then run `creap\build\skills-agg-v3.gms` to disaggregate the labor inputs (index set `LAB` must have already been defined).

It uses the *same* set and map file that was used to do the original data aggregation. Indeed, it reads the data file produced by the previous step.