Draft April 3, 2000

Evolution of the Canadian Input-Output Tables 1961 to date

by
Kishori Lal, Statistics Canada

Introduction

The work on Input-output (IO) tables in Canada started in the early 1960's with a small research staff of 6 to 7 persons under the kind, competent and visionary leadership of Terry Gigantes , a pioneer in input-output analysis. Right from the very beginning, it was decided that IO tables must fulfill several roles and provide: (a) an audit and management tool to improve economic statistics for their consistency, accuracy and comprehensiveness; (b) benchmarks for Gross Domestic Product (GDP), its income side and components, its expenditures side and components and GDP by industry estimates, both at current prices and constant prices and (c) a framework for structural analysis. About the same time, the 1968 international System of National Accounts, under the leadership of Professor Richard Stone of Cambridge University England, was being developed. It had incorporated an input-output framework as the foundation of the production accounts of the SNA. The Canadian IO framework turned out to be very similar to the one developed in the 1968 SNA. The first Canadian IO tables (use and make matrices), with industry by commodity dimensions, were published at the same time as the 1968 international SNA manual was released.

Given the roles assigned to the IO tables, it became obvious that they must also reflect industry information contained in the records of the producing establishments. In the real world, establishments produce both principal products and secondary products. Thus, it was decided to show both principal products and secondary products for industries in the make matrix. This was an innovation and, in our judgement, a very important one: it was no longer necessary to remove the output of secondary products and their inputs from the industry records in the IO framework. Further, there are many more commodities in the economy than industries. To reflect this information in a straight-forward manner, IO tables must be rectangular rather than square: that is, to have many more commodities (goods and services) than industries. This was another innovation of the Canadian tables and it closely resembled the thrust if the 1968 SNA

and also closely reflected the real world. However, showing the detail of commodity output for each industry resulted in large number of commodities becoming confidential under the Statistics Act of Canada. Thus it was decided to establish a separate unit in the IO division to be responsible for developing IO models based on our own parameters, as well as those provided by outside users for simulations on confidential data. The users would receive useful results, while we at Statistics Canada protected confidential data. For IO tables to become benchmarks for other SNA series, it became necessary to incorporate the SNA conventions on pricing, sectoring, classifications, etc. These early decisions have served us well and continue to be in effect.

Brief History

There have been four, not fully separable, phases in the evolution of the Canadian annual IO tables. We do not yet produce any sub-annual tables yet. The annual tables are: current price national IO tables, constant price national IO tables, current price provincial IO tables, and experimental projected IO tables. Statistics Canada started developing annual national IO tables for the year 1961 early in the 1960's, and it completed them in 1968. Soon thereafter, funding was provided to produce annual IO tables at current prices. Beginning with the reference year 1961, 37 annual national IO tables at current prices have been completed, the latest for the year 1997.

Early in the 1970's, funding was procured to produce constant price IO tables. So far, the following tables have been produced: annual tables for 1961 to 1971 at 1961 base prices; tables for 1971 to 1981 at 1971 prices; tables for 1981 to 1986 at 1981 prices; tables for 1986 to 1992 at 1986 prices and tables for 1992 to date at 1992 prices. We do not plan to produce constant price tables for the entire period since 1961 at one price base. We derive from the foregoing tables constant price value added by industry, using our preferred double deflation method. These value added series by industry are then chained or linked but retaining the growth rates produced during the respective periods. Thus, the growth rate of value added by industry for the 1961 to 1971 period remains the same, no matter which year we use to express the dollar values for the entire period since 1961. The 1993 SNA strongly recommends determining growth rates of GDP, or its components, or by industry, using chain indices.

Four occasional provincial -there are 10 provinces and since April 1999, three territories (previously two) in Canada- IO tables for the years 1974, 1979, 1984 and 1990 have

been produced. In addition to the make, use and final demand matrices, inter-provincial trade flows have also been produced. These tables have not been used as benchmarks for provincial SNA series and principally only for provincial IO models, mostly funded by cost recovery from outside users. This program, as of the reference year 1997, has now been funded to produce annual provincial tables. They will now play the same roles as national IO tables. In addition, the information produced in the provincial IO tables from 1997 onwards will be used by the Governments of Canada and three participating provinces for the distribution of goods and services tax. This tax in Canada is quite similar to the value added tax in Europe). The latest provincial IO tables, which are quite similar to the planned tables for 1997 and onwards, were completed for 1996 in November 1999. Tables for 1997 are expected to be released in November 2000. Every November after that date, we will have one more year of provincial IO tables.

Complete national IO are produced with a lag of 28 months from the reference year. For the current 28 months, Canadian SNA series, both quarterly and monthly, continue to be produced but their reconciliation with fully balanced IO tables and commodity flows must wait till IO tables are completed. Should we project the national IO tables to the current period, with a view to further improving the accuracy and consistency of the current period SNA series? Prompted by this question, attempts have been made during the last few years to project national IO tables, on an experimental basis, for the current period. By necessity, only partial information could be used as the more comprehensive information is not available yet. Starting with the latest fully balanced IO tables, one can use such information as the timely shipments, sales and employment data, and many other series normally used to produce the SNA series for the current period. Note that in Canada, we produce a full set of income and expenditure accounts at both annual and quarterly frequencies, as well as an economy-wide monthly GDP at constant prices by industry. This program is experimental and no funding has yet been requested.

Completion schedule

Statistics Canada’s computer facilities were relatively limited in the 1960's. Consequently, it took us about six years to finish the first IO table for 1961 in 1968. The dimensions in both industry space and commodity space have increased since the 1960's, but the interval between the reference year and the year of completion has been steadily reduced. Today, the current price national IO tables are completed after

28 months of the reference year; the constant price IO tables are completed after 32 months of the reference year; and the annual provincial IO tables are completed 34 months after the reference year. These are full set of tables with complete accounting of commodity flows with 680 commodities, 250 industries and 140 categories of final demand. We started with about 6 to 7 persons in the 1960's and the staff now is more than 50 persons, including a half-dozen of computer specialists.

Computing Environment

The first attempts at programming for the 1961 tables were done with MATOP, a mainframe software package. MATOP was a rigid spreadsheet-like language in which variables could not be referred to by name and the dimensions of a matrix were severely limited. The first inverse was calculated for the 1961 tables using a Fortran program, which took much longer to calculate than inverses calculated by today's computers.

During the 1970's, contracts for computing were entered into with outside firms (Computel, Systems Dimensions Limited (SDL), and Canada Systems Group (CSG)). For the first outside contract, it was necessary to go to Computel's offices in order to submit jobs; all programs, JCL (Job Control Language) and data were on punch cards. Any jobs for submission to the Statistics Canada mainframe computer had to be taken to the Statistics Canada main computer job submission area, printouts would be available for pickup 24 hours later. The division had several cardpunch machines for staff to use for their programs. Then, an IBM 2741 typewriter terminal was obtained, and from it, jobs could be submitted and printouts received. Later, access to the CSG computer was by means of a REMCOM card reader-printer. In the late 1970's, a Digital PDP 11/34 computer was obtained for the division, along with terminals and a line printer. This allowed jobs to be created and stored online and then submitted in card image format to the Statistics Canada mainframe computer or to the CSG mainframe computer; job results were printed on the line printer. As a result, card punch machines and card readers became obsolete.

The first versions of in-house matrix software (FINDEM and UTILE) were developed, in the early 1970's, to create the input, output and final demand matrices for one year. The input matrix contained values at purchaser prices, the margins (there are seven margins-transportation, pipeline, storage, wholesale trade, retail trade, and tax on

products) and producer values of all the commodities in the system. The first limited simulations were also done at this time. Most of the calculations to prepare the tables were done on worksheets, with the aid of calculators. To obtain data from other divisions, staff would have to go to the divisions to copy the information on worksheets. The first industry and commodity balancing at purchaser prices was done on cards. Generalized job submission routines were developed to allow analysts to perform calculations using the in-house matrix software, but without having to know JCL. FINDEM was a Fortran routine that performed rudimentary matrix operations on a restricted number of variables. UTILE, an extension of FINDEM, was the first attempt to introduce operations on sparse matrices; the number of matrix operations was limited as well as the sizes of the matrices.

The successor to FINDEM and UTILE was MOP (Matrix Operations Package), which performed a broad range of matrix operations on sparse matrices. It was the first attempt to integrate a random-access directory containing characteristics of matrices to allow error scanning, and it eliminated the need for definition statements for each variable referenced. Some of the matrix operations which could be performed included addition, subtraction, multiplication, inner products, transposition, normalization, printing, aggregations. This allowed the calculation of inverses, impact and multiplier tables. JCL was still required to run these programs.

In 1984, Input-Output Division's computing was done on an in-house Digital VAX 780 mid-range computer (later a VAX 8600), other divisions in Statistics Canada also had access to it. In 1990, the division acquired its own Digital VAX 3900, later replaced by a VAX 4200. Staff accessed these computers by means of terminals, with a large line printer for printouts. This was a major advance in computing for the division, as the staff gained direct access to the matrix software with instant results shown on the screen. With these computers came the first access to spreadsheet software, 20/20 and LOTUS, as well as to the word processor WordPerfect. Interfaces were developed to transfer data between the matrix software and the spreadsheet software. In 1994, the division acquired a Digital ALPHA 3000-600 S computer (followed by an ALPHA 4100), personal computers connected to the ALPHA by a Local Area Network (Digital Pathworks), PC software Lotus and WordPerfect, later Microsoft Excel and Microsoft Word, Laser Jet printers. These computers operate 24 hours a day, seven days a week. Each computer acquisition has resulted in a marked increase in processing speed, computer memory and disk storage. The processing speed of the 4100 is

considerably more than 10 times faster than the 300-600 S computer. These computers currently handle 50 and more users with no effect on processing speeds.

In moving to the Digital VAX computers, a new interactive matrix software package, TERF (Terminal Entry Review Facility), was developed. TERF is a fully interpretive, interactive language which allows multiple dimensions, sparse matrix operations, looping, run-time variables for indexing, error checking, and inheritance of characteristics for resulting variables. Confidentiality routines were also developed, allowing confidentiality checks on the data for the standard aggregation levels, as well as for the special aggregations demanded by customers. Special programs were developed for transferring the published data to the Statistics Canada mainframe CANSIM database. TERF is constantly being refined and enhanced to meet current data processing needs. The earlier matrix software was limited to two dimensions (e.g. industries and commodities), this progressed to three dimensions (e.g. years, industries and commodities) and, now, to four dimensions (e.g. years, provinces, industries and commodities).

Until the mid 1990's, the industry and commodity updating/balancing at purchaser prices was done with printouts and hand updates. The analysts would have printouts of their industries and commodities and would mark their changes on the printouts. In the 1970's, these updates were punched on cards prior to entering into the master files. In the 1980's and early 1990's, the changes were entered using a screen-entry updating program via the terminals; this was very time consuming as all updates were proof-read before entering into the master files. For the 1993-1995 tables, divisional systems staff developed a multi-user Graphical User Interface (GUI) for the industry/commodity updating/balancing for the national tables, using Visual Basic. This program accesses directly and instantly the master files for the inputs, outputs and final demand, as well as the margins and the inter-provincial trade flows, and the user can then enter updates and see immediately whether or not the industry or commodity is balanced. Some analytical tools are also included in this software package. The updates can be processed in a few minutes and then stored in the master files, and this can be done several times in one day. For the 1995 final and 1996 preliminary tables, the updating/balancing program was extended to include the provincial dimension for 1996. For the 1996-1997 tables, extra flexibility was added for processing two years of national and provincial data.

The Input-Output Division collects and processes large amounts of data from many sources, from both inside Statistics Canada and outside. The majority of the data is available in machine-readable form, in a wide variety of formats, and is transferred to the divisional computer and stored in matrix format or spreadsheet software. In the mid 1970's, the annual Census of Manufacturing data were put into matrix format and stored on the divisional computer for processing by the analysts; this was the first of the mechanized data available to the analysts. A single dimension of this matrix now contains approximately 50,000 establishments for a single year.

Over the years, the working levels for industries and commodities have increased. For example, the number of industry working levels has increased from 191 to 216 to the current level of 243; it will increase further with the implementation of NAICS (North American Industry Classification System) for the 1997 IO tables and onwards. The number of working level commodities has increased from 602 to 627 to the current level of 679, and will increase further because of the implementation of NAICS. The IO matrices have changed from single year matrices, to time series, to time series with the provincial dimension. The first in-house matrix software had only a handful of matrices whose number of rows or columns exceeded 1,000 elements; now there are some matrices with approximately 1,000,000 elements in one dimension. The international trade data are processed at the Harmonized System (HS) of commodities (over 21,000) and concorded to the IO working level of commodities.

The national input-output model was first developed in the early 1970's on cards using the FINDEM software. The first versions of the national model used the inverse method; this was changed to the iterative method. The inter-provincial model was first developed for 1966 provincial data, again using cards and the FINDEM software. For both models, calculations are done at the worksheet level, and the results aggregated to the level requested by a client. A GUI front-end has been developed for the National Model; the calculations are performed by TERF and the results are available in seconds, in a spreadsheet, or printout.

The publication environment has also changed dramatically over the years. For example, for the 1961 tables, MATOP was used at Computel, the data were on cards and the printouts were used for the publication. In later years, the text was typed separately and the data had to be specially formatted for sending to publication by outside firms. Now, with the desktop publishing software, the whole publication can be

prepared by IO staff, data can be prepared, transferred into Excel and then into the desktop publishing software.

Integration

Current price estimates of GDP as well as the underlying detail such as labour income, mixed income, taxes and subsidies on products and other surplus are identical in the IO tables and other SNA series produced at Statistics Canada. Similarly, expenditures on GDP and the underlying detail such as household expenditures, expenditures by the governments. Gross fixed capital formation, exports and imports are all identical throughout the Canadian SNA. In addition, the classification of categories of final demand, say for households, (there are more than 40 categories based on the international Classification of Individual Consumption by purpose- COICOP) is identical. This permits users to navigate from one set of statistics to another. This, as noted above, was thought very important right from the beginning of IO work at Statistics Canada.

Concluding Remarks

The IO program in Canada today is vastly different from the one we started in the 1960's. Its speed of delivery has increased even when its dimensions both in industry space and commodity space have greatly expanded. Its coverage is now many-fold, from one national table to 12 provincial/territorial tables. It has developed a proud history of annual tables from 1961, a rare statistical feat anywhere in the world. Its constant price version provides an excellent framework for structural analysis. We would not have been able to achieve this without the availability and ever increasing computer capacity. All this marvellous expansion occurred because of the importance that we have attached, right from the very beginning, to the role of IO tables in the framework of economic production statistics in the Canadian SNA program. We have been very conscious of acquiring and retaining a highly dedicated and professional staff, without which such an expansion could not be possible.

Note:

I am grateful to many colleagues in the System of National Accounts branch at Statistics Canada, particularly Yusuf Siddiqi and Abe Tarasofsky for their very useful comments and suggestions in the preparation of this paper. A special acknowledgement is due to Marilyn Constantineau of the IO Division in the preparation of the computer environment section of this paper.