Methodology for back-casting revisions to the 2007 and 2008 input-output tables

Introduction

The publication of the 2009 input-output (IO) tables introduced conceptual, classification, and statistical breaks from the previously published time series due to the 2012 comprehensive revision of the Canadian System of Macroeconomic Accounts (CSMA)[1]. While many other data products were revised historically to preserve time series continuity, the provincial and territorial IO tables were not recompiled prior to 2009 due to feasibility constraints. Because of this, the pre-2009 tables lost comparability to other revised CSMA products, such as the quarterly national economic accounts, the provincial economic accounts, estimates of GDP by industry and labour and multifactor productivity.

The labour-intensive nature of the IO compilation process precluded the possibility of recompiling the pre-2009 tables from the original source data according to the new concepts, classifications and statistical methods introduced with the 2012 CSMA historical revision. However, a compromise solution was to undertake a back-casting exercise to revise the current price provincial and territorial tables to attenuate the impact of time series breaks on analytical input-output products, such as the IO impact models and the symmetric industry-by-industry IO tables. The release of the 2007 and 2008 tables is the first installment in a series of releases which will eventually extend back to 1997.

The main purpose of the back-casted tables is to provide continuity for the analytical uses of the IO tables and not to serve as benchmarks in the historical period for the integrated MEA program as they do in current compilation. Since approximate modelling techniques are applied in their compilation, their overall quality is not equivalent to that attained via an extensive bottom-up compilation exercise at a detailed level.

The back-casted tables are aligned with most aggregates of the Provincial Economic Accounts (PEA) published in November of 2013. Due to statistical limitations, they are generated at a slightly more aggregate level (the 'Link 1997' aggregation) than the full detail of the new classifications from 2009 forward.

While quality differences cannot be quantified, the back-casted tables are suitable for macroeconomic analytical uses that do not require complete integration with other products (such as the historical series of industry volume measures or the multifactor productivity estimates).

Back-casting methodology

[1] Statistics Canada. 2013. "Modernization of the Input-Output tables". Industry Accounts Division. Statistics Canada.

The methodology for back-casting the IO tables can be roughly decomposed into the following steps, each of which will be further explained below:

a) Conversion of benchmark tables to new commodity classification
b) Adjustments for other classification changes
c) Adjustments for conceptual changes
d) Adjustments to align with PEA aggregates
e) Balancing the tables

a) Conversion of benchmark tables to new commodity classification

The new IO commodity classification introduced in 2009 provided more relevant and detailed definitions of the IO commodities but also represented a substantial break from the previous structure. The main approach has therefore been to reallocate, wherever possible, the benchmark estimates to the new commodities based on allocation weights from the primary source data.

The return to the use of source data is motivated by two main factors. The first is that in some cases source data provide information more aligned with the new classifications; such as the product information from the Annual Survey of Manufacturers. Secondly, the high level of detail in some source data can be used to reconstruct estimates with more precision. The best example is provided by international trade, where the high level of detail from the Harmonized System can be used to reconstruct an approximation of trade estimates under the new classifications.

However, the source data-based approach could not be generalized to all the new commodities. Due to statistical difficulties, aggregation was still required for 15 commodities, such as for crude oil and bitumen or the wholesale margins and commissions.

An important drawback of returning to the original source data for reallocating the value of commodity estimates is that it reintroduces incoherencies that had previously been resolved during the data confrontation (or balancing) phase of the compilation of the IO tables. Nonetheless, since this was an exercise in the reclassification of “known” commodity values, basic statistics such as industry and category totals or other subtotals such as GDP or total intermediate inputs by industry were not impacted.

A two step approach is used to resolve the contradictory objectives of re-estimating values based on the source data while maintaining all basic information from the existing tables. The first step estimates the values under the new commodity classification while the second step re-establishes additivity of the data in the supply-use dimension (i.e., rebalances the tables).

In the first phase, the source data are simultaneously mapped to both the old and new commodity classifications for each industry and final demand category in the old IO tables. These data provide allocation weights used to apportion an appropriate amount from each IO value from the old to the new commodities. While this method generates estimates of the new commodity values that respect the published industry estimates, it inevitably generates estimates that lack coherence in their supply and demand dimensions.

The second phase re-establishes coherence from a supply-demand perspective through an algorithmic approach. A constrained optimization model is applied that minimizes variations from the estimates determined in the first phase subject to a set of constraints that enforce the supply-demand identities in basic prices[2]. An additional constraint imposed on the model is that for each industry, the subset of new commodity values that compose a previously published commodity value must sum to the initial value of the latter. This latter constraint ensures that the industry and other basic structural identities in the published IO tables are not modified during the commodity balancing process. The model is solved in the provincial dimension and national estimates are derived as the sum of the provincial estimates.

Where source data could not provide sufficient details or coherent weights, estimates were adjusted on a case by case basis. In certain cases, estimates of supply were used to determine demand or vice versa, depending on the relative quality of the source information.

Valuation and margins

As mentioned previously, the model is solved in basic prices. The margin components are subsequently allocated based on the basic price weights of each new commodity in the previously published commodity value from the balanced IO tables. The purchaser price estimates are derived as the sum of the basic price and margin components.

The symmetric allocation of margins according to the basic price proportions is reasonable under the assumption that sufficient homogeneity in margin rates exists within the group of commodities involved. While this may be a reasonable assumption for most of the commodities allocated, certain margin rates may nonetheless differ, even within a group of similar commodities, and a selective re-examination of the margin allocation weights was necessary, especially for some of the less homogeneous margins, such as taxes.

b) Adjustments for other classification changes

The new industry classification introduced for 2009 is based on the North American Industrial Classification System (NAICS) and thus does not represent any major departures from the previous classification system. The new industrial classification has, in general, less detail in the goods-producing industries and greater detail in services, mostly in the wholesaling and retailing industries.

[2] See section e) for further details on the optimization models used.

In the industry dimension, the revised IO tables were summed to the most detailed possible common aggregation between the old and new industry classifications. These have traditionally been referred to as the «Link» aggregations between different IO classifications. For example, the two separate industries for 'Wineries' and 'Distilleries' in the old classification are aggregated to match the 'Wineries and distilleries' industry in the new classification. Conversely, the 12 wholesaling and 12 retailing industries in the new classification are combined to match the two prior aggregate industries for total wholesale and retail trade. The new industry classification contains 235 industries whereas the common, "Link" industry structure contains 188.

Simple aggregation, however, is not used to resolve all differences between the two industry classification systems. The activities of Aboriginal government, included in the 'Other non-profit industries' in the old classification, are classified under the government sector in the new classification. Rather than creating a common aggregation between the non-profit and government sectors, it was deemed preferable to introduce instead statistical adjustments to the IO tables that would generate separate estimates for the non-profit and aboriginal government industries.

The fixed capital formation industries in final demand followed the same general principle set for the industries in the output and input tables. Aggregation was used where possible to create a common time series with the exception of the Aboriginal government and non-profit industries. The new category of intellectual property products incorporated the values of exploration that were previously in construction, software that were previously in machinery and equipment and the new estimates of research and development.

In final demand, the final consumption expenditures of households categories in the new classification are based on the international classification standard, the Classification of Individual Consumption According to Purpose (COICOP) and differ substantially from the previous classification of personal expenditures. Since estimates are available at the detailed level for these categories from the Provincial Economic Accounts, these categories were adopted as is. The commodity dimension of these categories was determined based on the previously published composition and a reexamination of source data for major definitional changes introduced in the comprehensive revision.

c) Adjustments for conceptual changes

There are three main conceptual revisions that have a substantial effect on the IO tables. They relate to the capitalization of expenditures on research and development (R&D), military weapons systems, and exploration services. The first two changes capitalize expenditures previously treated as intermediate consumption and represent new concepts introduced with the 2012 CMEA comprehensive revision. The third change removes the routing of the output of exploration services through the non-residential construction industries, a simplification resulting from the introduction of a new fixed capital formation category for intellectual property products in the comprehensive revision.

These three major adjustments are estimated based on source data. The Gross Domestic Expenditures on Research and Development (GERD) figures published by Statistics Canada were used to estimate own-account output and purchases of market R&D by industries. For weapons systems, some of the intermediate expenditures of the defence industry are reclassified to capital expenditures based on asset information from the Financial Management System of government statistics and the aggregate capital expenditures published in the Public Accounts of Canada.

Estimates of investments in exploration services from the Oil and Gas Extraction Survey (Statistics Canada) and the Survey of Mineral Exploration, Deposit Appraisal and Mine Complex Development Expenditures (Natural Resources Canada) are used to shift the appropriate values from the non-residential construction to the IPP categories in final demand. The outputs and related inputs of the oil and gas and other engineering construction industries are reduced by an equivalent amount.

d) Adjustments to align with PEA aggregates

The IO tables are further adjusted to match the PEA aggregates published in November 2013. These include for each province or territory, the income and expenditure GDP aggregates, the detailed series of final consumption expenditures of households and the current price GDP by NAICS industry[3], after accounting for differences due to sectoring in the IO tables.

Industry estimates were also examined in light of the revisions to the source data incorporated in the comprehensive revision. The commodity structure of the outputs and inputs of industries was adjusted to align with those revisions as well as new patterns introduced in the 2009 and subsequent benchmark IO tables.

e) Balancing the revised tables

The combination of steps described above inevitably introduces a large number of incoherencies into the data. A two stage process is used to remove these incoherencies. In the first stage, large imbalances in the supply-use framework or outliers from a time series perspective are examined and adjusted on a case by case basis. In a second stage, the remaining smaller incoherencies are eliminated through the use of a balancing algorithm, which is presented in a semi-generalized form.

The balancing algorithm generates the minimum variations necessary to remove all imbalances in the tables through the following optimization:

$$\min_{x^P,x^N} \sum_{ijklm} w_{ijklm}\left(x^P_{ijklm} + x^N_{ijklm}\right) \tag{1}$$

[3] See CANSIM 384-0037, 384-0038, 384-0041, and 379-0030.

subject to:

$$\sum \alpha_{ijklm}(x^{P}_{ijklm} - x^{N}_{ijklm} + a_{ijklm}) = 0 \quad (2)$$

$$x^{P}_{ijklm} \geq b^{P}_{ijklm} \quad (3)$$

$$x^{N}_{ijklm} \geq b^{N}_{ijklm} \quad (4)$$

where x^{P}_{ijklm} and x^{N}_{ijklm} are the decision variables representing the values of positive and negative variations respectively associated with each non-zero value in each province *i*, IO table *j*, commodity *k*, industry *l*, and valuation *m* and for the remaining exogenously set constants, w_{ijklm} specifies the weight of each variable, a_{ijklm} is equal to the value of each element in the initial unbalanced tables, α_{ijklm} is a term that is equal to +1 or -1 to define the balancing equations, b^{P}_{ijklm} is the upper bound for the positive variations, and b^{N}_{ijklm} is the upper bound for negative variations.

For example, for a given commodity *k*, setting α_{ijklm} to +1 for the elements of supply and to -1 for the elements of demand will generate the values of x^{P}_{ijklm} and x^{N}_{ijklm} necessary to remove imbalances between supply and demand. The values of w_{ijklm} allow the incorporation of information on the relative reliability of different elements; thus, final expenditures, where detailed estimates are deemed to be of high reliability are given a high value whereas intermediate consumption detail, deemed to be less reliable, is given a relatively lower value.

A slightly different form of equation (2) is also used to match exogenously set values. Equation (5) shows how any subset of elements can be set equal to a known value c_n, where *n* is the number of related, appropriately ordered known values:

$$\sum (\mathrm{x}^{\mathrm{P}}_{\mathrm{ijklm}} - \mathrm{x}^{\mathrm{N}}_{\mathrm{ijklm}} + \mathrm{a}_{\mathrm{ijklm}}) = \mathrm{c}_{\mathrm{n}} \quad (5)$$

Finally, inequalities and proportionality relationships between variables can also be specified. For example, setting *m* =1 as the purchaser valuation and *m* = 2 as one of the margin valuations, margin values can be subject to an upper bound proportion of the purchaser value as in equation (6) or a lower bound proportion of the purchaser value as in equation (7):

$$(u_{ijkl} - 1)(\mathrm{x}^{\mathrm{P}}_{\mathrm{ijkl1}} - \mathrm{x}^{\mathrm{N}}_{\mathrm{ijkl1}} + \mathrm{a}_{\mathrm{ijkl1}}) \geq 0 \quad (6)$$

$$(\mathrm{l}_{\mathrm{ijkl}} + 1)(\mathrm{x}^{\mathrm{P}}_{\mathrm{ijkl1}} - \mathrm{x}^{\mathrm{N}}_{\mathrm{ijkl1}} + \mathrm{a}_{\mathrm{ijkl1}}) \leq 0 \quad (7)$$

where u_{ijkl} is the upper bound on margin rates and l_{ijkl} is the lower bound on margin rates.

The linearity of the balancing algorithm presents a great advantage in solving large-scale problems. While such algorithms provide expedient solutions, the resulting mechanically-rebalanced estimates are not of equal quality to those estimated directly by experienced national accounts compilers.