Computational aspects of a large environmentally extended input-output system

Reinout Heijungs & Arjan de Koning
Institute of Environmental Sciences, Faculty of Science, Leiden University, Netherlands

Abstract

Environmentally extended input-output analysis is a tool for supporting policies on sustainable production and consumption. Present-day statistical offices collect and make available a huge amount of data for this. The computational challenge is to organize these data in the right way, and to develop the well-known matrix-based computational steps into efficiently implemented algorithms. This paper sketches some of these developments in the EU-financed EXIOPOL project.

Keywords

environmentally extended input-output analysis (EEIOA), relational database, matrix algebra

Introduction

Input-output analysis (IOA) is a technique for organizing economic data and using this for all sorts of analytical purposes (Miller & Blair, 1985). Wassily Leontief (1936, 1941) operationalized in the 1930s and 1940s Quesnay's *Tableau Économique*. Leontief's consistent labour on this has been well awarded. Besides being awarded the 1972 Nobel Prize in economics, Leontief's ideas have been welcomed by statistical agencies throughout the world. The main idea of collecting information on economic transactions between firms and final demand by households and other consuming agents has become a prominent part of the work of the statistical divisions of the UN, the EU, the OECD, and so on.

An interesting development was also initiated by Leontief (1970): the addition of separate accounts for pollution, waste, and other "environmental repercussions" of production and consumption. From this grew the NAMEAs (de Haan, 1994), representing the "national accounting matrix with environmental accounts". Such NAMEAs are available for a number of developed countries and are being developed for the full EU27, but are not available for the entire world at sufficient level of detail to be very useful.

IOA is used nowadays intensively in the context of environmental economics, environmental life cycle assessment, scenario and policy studies, embodied pollution of trade, and similar subjects that relate to sustainable production and consumption. Obviously, in an increasingly globally connected world, trade links between countries form an ever more important ingredient. Input-output tables (IOTs), on the other hand, are to a large extent collected for individual countries, with some overall aggregated information on imports and exports. As the basic layout of the IOTs differs sharply between countries, it is difficult to connect several domestic IOTs into a multi-national table of high quality.

Within the EU-FP6 research scheme, the EXIOPOL project has been set-up to address the compilation of an EU-wide system of environmentally extended IOTs (EEIOTs); see Tukker et al. This obviously involves a large amount of data collection, estimation and transformation. But it also involves a number of steps with computational relevance. This paper addresses some of these issues. It describes some of the features of EXIOBASE, the database management system that is in development and that will become available in 2011.

Data storage

In EXIOBASE, 43 countries and 1 "Rest of the World" region are distinguished covering the global economy. The countries are the 27 member states of the EU and their main trade partners (Norway, Switzerland, US, Canada, Japan, China, Brazil, etc.). For each of these countries, the so-called supply and use tables (SUTs) provide the working material to create the IOTs. These are defined in terms of 129 industries (such as cultivation of wheat, forestry, casting of metals, and air transport) and 129 products (such as poultry, fish products, gas oils, and hotel and restaurant services). Thus, for 44 countries, two tables of size 129*129 are maintained. In addition, there are tables for final demand, imports and exports, labour, pollutants, etc. It is clear that a simple spreadsheet no longer suffices, but that a well-defined relational data model is needed to be able to store the data in a flexible yet robust manner.A fragment of the tables and their relations is shown in Figure 1.

Figure 1. A small part of the database structure of EXIOBASE.

Figure 1 shows the implementation in MS Access, but given the cross-platform project requirements, the database also runs under MySQL and Derby. To give an indication of the size, the present (not yet complete) version of the Access file is more than 250 MB.

Data transformation

In fact, the screenshot shown represents just the very input side of EXIOBASE. Figure 2 shows the overall architecture of EXIOBASE.

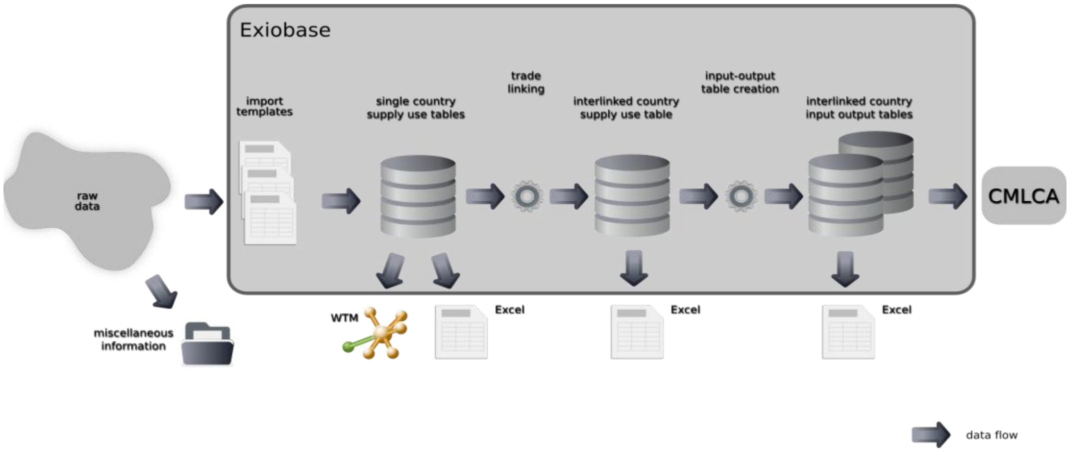

Figure 2. Overall structure of EXIOBASE. The part shown in Figure 1 is a zoom in of the “single country supply use tables”.

The SUTs apply to the individual 44 countries, but they are only linked in an aggregate way. The trade linking algorithm converts these loosely connected SUTs to one large SUT with tightly interlinked regions. The computational steps needed for this are partly described in literature (Oosterhaven, 1984), partly new (Bouwmeester & Oosterhaven, 2009). They involve an iterative readjustment process on a set of matrices of approximately 6000 rows and columns. The results of this is referred to as the “interlinked country supply use table”, and it is stored in the relational database.

A second transformation step is the conversion from the SUT to the IOT. This is a mathematical operation for which a number of alternative models are available, see Ten Raa & Rueda-Cantuche (2007). Typical steps involve the inversion of a diagonalized matrix and the multiplication and transposition of matrices. Remembering that the matrices are 6000*6000, it is clear that the implementation of these routines should be done critically, using sparse matrix storage and operations whenever profitable, and with a keen eye on memory requirements and CPU time. One example may illustrate this. the mathematical treatment of Ten Raa & Rueda-Cantuche (2007) gives formulas like

$$\mathbf{VA} = \mathbf{W}\left(diag\left(\mathbf{S}^T\mathbf{e}\right)\right)^{-1}\mathbf{S}^T$$

A naive implementation leads to a straightforward algorithm that is computationally very inefficient, involving for instance the inverse of a diagonal matrix. Working out the math provides a – from a computational point of view – much simpler form:

$$i = 1...I,\ j = 1...J : va_{ij} = \sum_{k=1}^{K}\left[w_{ik}\left(1\Big/\sum_{l=1}^{L}s_{lk}\right)s_{jk}\right]$$

No inverse needed, and not 4 products of vectors or matrices, but just two.

Analytical use of data

EXIOBASE is a structure that contains data and two important computational steps. It also comprises a limited functionality in terms of presenting data on screen, using SQL-based forms. But the real analytical power, connecting it to scenarios for final demand, calculating impacts on economy and environment, converting these into impact indicators (such as global warming), etc. takes place outside EXIOBASE, in the CMLCA program, a program that was originally developed for LCA (Heijungs & Suh, 2002), but that has been adapted to cover IOA, EEIOA and hybrid LCA/EEIOA as well (Heijungs et al., 2006). CMLCA is a freely available program for Windows, written in Delphi. The exchange of data from EXIOBASE to CMLCA takes place through XML technology. Although XML files can be extremely large, rapid and streaming-based parsers are available for Java and Delphi, so that the actual exchange of data can take place within a few minutes.

Discussion

IOA is full of developments that touch on the limits of computational power. Leontief did his first studies on an economic system that was subdivided into no more than 9 sectors. Even then, the “total number of multiplications involved in the practical solution of our problem exceeds 450,000. This task alone would mean a two-year job, at 120 multiplications per hour. Fortunately, the recent invention of the Simultaneous Calculator by Professor Wilbur of the Massachusetts Institute of Technology has made it possible to perform all the necessary computations in a small fraction of the time they otherwise would have required. This apparatus solves nearly automatically a system of nine simultaneous linear equations [...].”(Leontief (1941, p. 74).)

In 1950, Waugh introduced the technique of a power series expansion to approximate the inverse of a matrix. The idea rested on the theorem that

$$\left(\mathbf{I}-\mathbf{A}\right)^{-1} = \mathbf{I}+\mathbf{A}+\mathbf{A}^2+\mathbf{A}^3+\cdots$$

provided that A satisfies certain conditions that are usually met in IOA. This trick obviates the computationally demanding step of inverting a matrix, and replaces it by the easier operations of multiplication and addition.

The limit of nine sectors and the sequential approximation of the inverse matrix have for long been outdated by the digital revolution. But the increase of computer power should be employed in a well-considered way. The system that is available in EXIOBASE poses new challenges:

- data structures become too large for MS Excel or programs
- the number of records becomes so large that automated checking procedures and integrity warranting mechanisms become indispensible

- the choice of data representation (e.g., single or double precision, sparse matrix operations) becomes essential
- the design of the computations (e.g. the order of loops) and the choice of algorithm (e.g., for inverting a matrix) can become decisive
- special precautions may be needed to ensure the stability of the numerical steps.

An example of this latter is provided by the need to rescale a matrix before subjecting it to inversion, and to scale back the obtained inverse (cf. Heijungs & Suh, 2002, p.148-150).

Altogether, there appears to be room for interdisciplinary groups, combining the content (economics, environment, etc.) with the various disciplines that relate to data storage, data retrieval, data processing, algorithms, and numerical mathematics. The era of data abundance and computer power offers lots of opportunities, but it poses new demands and skills.

Acknowledgement

This paper has been written in the context of the EU-FP6 EXIOPOL project.

References

Bouwmeester, M. & Oosterhaven, J. Methodology for the construction of an international supply-use table. 17th International Conference on Input-Output Techniques, Sao Paulo, 2009

Haan, de M.; Keuning, S.J.; Bosch, P.R. Integrating indicators in a National Accounting Matrix including Environmental Accounts (NAMEA), National Accounts Occasional Papers NA-60, Statistics Netherlands, Voorburg/Heerlen, 1994.

Heijungs, R.; Suh, S. The computational structure of life cycle assessment. Kluwer Academic Publishers, Dordrecht, 2002

Heijungs, R.; de Koning, A.; Suh, S. & Huppes, G. Toward an information tool for integrated product policy. Requirements for data and computation. Journal of Industrial Ecology 10:3 (2006), 147-158

Leontief, W.W. Quantitative input and output relations in the economic system of the United States. The Review of Economic Statistics XVIII:3 (1936), 105-125

Leontief, W W. The structure of American Economy, 1919-1939. An empirical application of equilibrium analysis. Oxford University Press, New York, 1941/1953

Leontief, W. Environmental repercussions and the economic structure: an input-output approach. Review of Economics and Statistics LII (1970), 262-271

Miller, R.E.; Blair, P.D. Input-output analysis: foundations and extensions. Prentice Hall, Englewood Cliffs, 1985

Oosterhaven, J. A family of square and rectangular interregional input-output tables and models. Regional Science and Urban Economics 14 (1984), 565-82

Raa, T. ten; Rueda-Cantuce, J. A Generalized Expression for the Commodity and the Industry Technology Models in Input-Output Analysis. Economic Systems Research 19:1 (2007), 99-104

Tukker, A.; Poliakov, E.; Heijungs, R.; Hawkins, T.; Neuwahl, F.; Rueda-Cantuche, J.M.; Giljum, S.; Moll, S.; Oosterhaven, J.; Bouwmeester, M. Towards a global multi-regional environmentally extended input–output database. Ecological Economics 68:7 (2009), 1928-1937

Waugh, F.V. Inversion of the Leontief matrix by power series. Econometrica April (1950), 142-154