iotables: an R Package for Reproducible Input-Output Economics Analysis, Economic and Environmental Impact Assessment with Empirical Data

Daniel Antal, Leo Lahti

14 February, 2022

Introduction

Several years have passed since the first release of the eurostat R package (Lahti et al. 2017), which has been designed to facilitate reproducible retrieval and analysis of Eurostat's more than many thousand statistical open data products. Eurostat produces European statistics in partnership with national statistical institutes and other national authorities in the EU Member States. This partnership is known as the European Statistical System (ESS). It also includes the statistical authorities of the European Economic Area (EEA) countries and Switzerland, and in cooperation with other developed nations, it often publishes comparable U.S., Japanese, or other data.

The rOpenGov community around the original package has developed several CRAN released extensions to manage the idiosyncratic problems of particular subsets of this large, real-life data source. The *regions* package (Antal 2021) retrospectively tracks the rather frequent boundary, name, and geographical code changes of sub-national areas, such as provinces, regions, departments and counties. The iotables package (Antal 2022) deals with a different but complementary problem, the analytical inter-dependency of many statistical data elements within the system of national accounts. What connects these packages is that they utilize standardized statistical metadata to improve the usability of the upstream *eurostat* package.

The supply and use tables (SUTs) and input-output tables (IOTs) provide a very detailed, partly empirically measured, partly statistically estimated picture of the economy. These tables present information on the supply and use of goods and services for industries' intermediate consumption and categories of final use (final consumption, capital formation and exports). They also provide details on the generation of income for each industry distinguishing the components of gross value added. The SUTs and IOTs provide empirical data for a wide range of economic analyses. They follow the seminal work of Wassily Wassilyevich Leontief (Leontief 1951), who won the Nobel Memorial Prize in Economic Sciences in 1973 mainly for developing this analytic toolkit.

The system of input-output tables are the most comprehensive, empirically measured data for many types of macroeconomic research or industry organization analysis, and they provide tools for various economic and environmental impact analysis.

There are several R packages that would allow the user to download the necessary input-output data from the Eurostat Rest API, for example, the datamart (Weinert 2014), or the rsdmx packages (Blondel 2021). However, we chose as a dependency the *eurostat* package, because it is highly customized to this particular data source, and it is also very widely used to access data from the European Statistical System.

The input-output system is a matrix algebraic system. The system of the input-output tables must be spread into at least four, interconnecting and compatible matrixes. Any further data for analysis (such as data on employment, or material flows like greenhouse gases) must be added to a system of matrix equations

in the form of conforming vectors or matrixes. Properly formed coefficient matrixes must be calculated from parts of the input-output table, and they must be translated into the Leontief matrix and its inverse. The *iotables* R package adds the functionality to *eurostat* to properly process the long-form data into many tables—in some cases, the bulk downloader returns more than 800 SUTs in one single long-form dataset. The `eurostat::get_eurostat` function retrieves the requested SUTs or IOTs in a tidy, long-form, but these tables cannot be meaningfully spread to a wide form without understanding the vocabulary of the System of National Accounts, and prefectly filtering and ordering the rows and columns into well-formatted matrice. This is the functionality that the *iotables* data importing functions add to the original dependency.

Since the first release of the *iotables* packages in 2018, we saw the appearance of two new packages with a partly overlapping functionality, but with a quite different focus. Most input-output economics uses can be described in a few matrix equations, partly, because in real life preparing the underlying matrixes is a greater challenge than their analysis. The *leontief* package overlaps with the analytical functionality of *iotables* in the way it selects appropriate vectors from the input-output tables, and uses them in matrix equations to create multipliers. The *ioanalysis* package calculates the fundamental IO matrices following Leontief's work and provides further support for various analytical applications (Wade and Sarmiento-Barbieri 2020) that are different from the current *iotables* analytics. The *iotables* packages does provide functionality for the most widely used economic and environmental impact analysis, but its focus is making the research workflow reproducible, and the help with the laborious data wrangling, and the painstaking matching of auxiliary or 'satellite' data to broaden the analytical capabilities. The *iotables packages* can be used as a data gathering and preparation application of the other two packages—particularly to overlap with *leontief* is so great that in the future we are planning to paralell develop both packages.

Package functionality

Data retrieval and processing

The creation of an input-output table from raw data is the most complex task in the production of governmental economic statistics, and it is beyond the scope of *iotables*. However, it must be that due to the complexity of this task, even developed countries usually produce a new input-output table every five years, and they may choose various data sources that result in slightly different (but compatible) matrixes.

The current version of *iotables* works with a metadata table that contains four metadata dictionaries to spread the long-form data retrieved from the Eurostat Rest API, or other sources to the correct interconnecting matrixes. These matrixes use the `NACE` or `CPA` statistical coding and labelling of various macroeconomic and industry classification information (Eurostat 2008b, 2008a; EUR-Lex 2006, 2008). As we elaborate on the end of the article, this limitation means that we can provie the full workflow including *automated data retrieval and data wrangling* with only about 33 developed countries, and we hope that we will be able to build similar future functionality to other data sources. However, the rest of the functionality works with any SUTs or IOTs, provided that the user can import it from a spreadsheet in a correct format.

The Eurostat Supply, use (SUTs) and input-output tables (IOTs) distinguish 64 industries and 64 products. This means that the they aggregate data in a bit idiosyncratic way, because they often aggregate `NACE` or `CPA` categories. In NACE/CPA the division of the economy is hierarchival system:

- The first alphabetic code refers to the *section* of the economy;
- Adding the first numeric code defines a narrower *division* of the section, for example, multiplier_create(input_vector = emission_coeffs[2,], Im = I_de, multiplier_name = "CO2_multiplier," digits = 4)

, like `(CPA_)A` and `(CPA_)F` agriculture or construction, or , for example, `A01` or `CPA_A01` Products of agriculture, hunting and related services), and sometimes, they combine several divisions of the same sector together. They never go down to the level of groups (`A012`— perennial crops) or classes (`A0121` — grapes) or beyond (`A012111` — Table grapes.) There are far more detailed economic statistics available outside the scope of IOTs. What makes SUTs and IOTs especially useful is that they describe the relationship among economy with a system of 64x64=4096 comprehensive indicators to describe only the economic inter-relationships. The data processing functions of iotables make sure that this data is wrangled into a strict column/row order so

that it can be used in a system of matrix equations. This processing is probably the most valuable part of the package functionality: ordering many thousand indicators into strict column and row order is an error-prone process. Errors are sometimes easy to capture, when they breach some algebraic condition and later steps result in a syntax error. But in many cases, an ordering error results in a logical error that may yield a credible-looking, but false result. Our quality control with many unit tests try to exclude such logical errors.

The `iotables_download()` and the `iotable_get()` functions download and retrieve a single, properly processed input-output table from the Eurostat data warehouse. In this example, which is a simplification of the Introduction to iotables vignette, we use the built-in simplified input-output table for Germany taken from the *Eurostat Manual*.

```
library(iotables)
germany_io <- iotable_get( source="germany_1995", labelling = "iotables" )
input_flow <- input_flow_get (
                data_table = germany_io,
                households = FALSE)

de_output <- primary_input_get ( germany_io, "output" )
print (de_output[c(1:4)])
```

```
##     iotables_row agriculture_group industry_group construction
## 15        output             43910        1079446       245606
```

```
#>     iotables_row agriculture_group industry_group construction
#> 15        output             43910        1079446       245606
```

The various data wrangling functions of `household_column_get`, `output_get`, `primary_input_get` help to subset often used sub-matrixes or vectors from the input-output table, which is often hidden with labelling not immediately familiar to the analyst. The `total_tax_add` and `supplementary_add` help merging tax row and adding auxiliary data to input-output table with maintaining the strict ordering and labelling of the matrixes. we also adopted various tidyverse functions, for example, `vector_transpose_longer` and `vector_transpose_wider` which keep the key column, and the strict row/column order of various vectors needed in the input-output system.

Analysis in the input-output system

Apart from the automated retrieval, data wrangling and unit testing of SUTs and IOTs available in the Eurostat data warehouse, the rest of the package functionality works with any tables—provided that the user imported them in a correct format. Most of our long-form documentation (tutorials) and the package examples do not even follow the European SNA 2010 table format, because their 64x64 indusry size make any results impossible to display on a single screen.

The analytical functions of *iotables* make sure that otherwise relatively simple algebraic equations in input-output analysis are performed on meticoulusly matched, conforming matrixes. These functions are accompanied by many unit tests, and meaningful error messages. Whenever the user tries to work with non-conforming matrixes, a simple base R error message would be triggered. We tried to build in more meaningful error messages to explain where and how these objects are likely to be incompatible. We also give in some cases meaningful error messages when the matrixes contain logical errors, but they do not breach the basic algebraic conditions, and the underlying matrix equation delivers an errorneous result. Because the standard Eurostat matrix is has at least 4096 elements, tracing such logical mistakes is extremely time-consuming. During three years of practical use, we tried to cover as many exceptions with meaningful error messages as possible to make the debugging of logical errors more efficient.

The *matrix processing functions* of `coefficient_matrix_create`, `input_coefficient_matrix_create`, `output_coefficient_matrix_create`, `output_coefficients_create` create various coefficient matrixes with dividing the appropriate elements in the proper ordering and with retaining the proper labelling. The input coefficient matrixes are likely to be used in the demand-driven, original Leontief input-output model

(Leontief 1951), and the output coefficients in the supply-driven dual model of Ghosh (Ghosh 1958).

Whilst input-output economics has a fairly standard analytical method, it has a long history or rather different applications in macroeconomic analysis, antitrust, tourism economics, cultural economics, or environmental impact assessment, to give a few examples. Different disciplines have incorporated the use of the input-output system with a slightly different terminology. The names of our data wrangling and anaytical functions follow the conventions of *Eurostat Manual*, because the original aim of our package was to give a programmatic, reproducible access to the tables harmonized by the Eurostat statistical agency. However, in the package documention we have described the other commonly used names for these matrixes. For example, the *input flow matrix* of the Eurostat Manual is often called the inter-industry matrix in other literature.

```
de_input_coeff <- input_coefficient_matrix_create(
    data_table = germany_io,
    digits = 4)

print(de_input_coeff[1:3, 1:3]) # use knitr::kable instead of plain print, for a nicer output?
```

```
##         iotables_row agriculture_group industry_group
## 1 agriculture_group            0.0258         0.0236
## 2    industry_group            0.1806         0.2822
## 3      construction            0.0097         0.0068
```

The `leontief_matrix_create` and `leontief_inverse_create` create the most important object of input-output analysis, first described, and named in honour of Leontief. The `ghosh_inverse_create()` the inverse from the 'supply-driven' input-output model.

```
L_de <- leontief_matrix_create(technology_coefficients_matrix = de_input_coeff)
I_de <- leontief_inverse_create(de_input_coeff)

print(I_de[,1:3])
```

```
##                 iotables_row agriculture_group industry_group
## 1          agriculture_group        1.03390950     0.03501839
## 2             industry_group        0.28968590     1.42923465
## 3               construction        0.02070433     0.01910180
## 4                 trade_group        0.12698600     0.12146217
## 5     business_services_group        0.18418167     0.20706399
## 6        other_services_group        0.04945504     0.02953987
```

Industrial linkages

Backward linkages show the buying linkages towards suppliers, and often understood as the strength to pull the supplier base when a given industry is growing. Industries with a strong pull tend to create many purchasing orders when they are growing within the same economy. Forward linkages show the supply side effects when the industry in question is growing. The abundance of supply, with normal goods associated with falling prices, creates more opportunities within the same economy for users of this intermediate product. Industries with a strong push tend accumulate many purchasing orders from others.

The analysis of backward linkages is often an important starting point in development economics: foreign direct investment that finances new activities with high backward linkages is likely to increase the production, employment, wages, and tax receipts of a developing nation. Backward and forward linkages can play an important role in the understanding of vertical problems in competition economics, or analyzing the competitiveness of an economy (Botrić 2013).

Economic impacts

The advantage of working with symmetric input-output tables is that they give a detailed portrait of an economy, including the inter-linkages of various sectors. Eurostat's input-output tables detail by default 63x63 economic activities (or products of those activities). This means that for each activity, such as power generation, we can analyze the impacts on an entire supplier (upstream) and purchases (downstream) supply chain of 62 other industries. For example, power generators increasing production, and buying more extracted natural gas, and selling the power via energy merchants to car manufacturers, banks, or health providers.

The Working With Eurostat Data vignette explains economic impact analysis with *iotables* in greater details. It compares the input, output multipliers, the employment direct and indirect effects, and the inter-industry linkages in the Slovak and Czech national economies. It contains a similar calculation that was used in the **Slovak Music Industry Report* [Správa o slovenskom hudobnom priemysle] to compare the various employment, gross value added and production related tax potentials of the development of music industry compared to other sectors of the Slovak national economy (Antal 2019).

Environmental impacts

When a particular form of environmental impact, for example, the emission of greenhouse gases, if a function of the technologies applied by an industry, the input-output system is a powerful tool to understand the adverse impacts of various economic growth scenarios. The Environmental Impacts vignette explains environmental impact analysis with *iotables* in a greater detail.

Quality Control

During the development of iotables, we have tried to replicate analytical findings from reliable, publicly available sources. The inter-industry matrix of an IOT can be aggregated to as small as a dimension of 2x2 or 6x6. Obviously, researchers prefer to use much larger matrixes, but they are more difficult to use for replication and unit-testing in a transparent manner. Statistical manuals therefore usually contain relatively small IOTs which can be easily read in a printed book, and can be replicated, too. We have given a priority to such published results.

We try to avoid as many hard-to-detect errors for the user as possible, by replicating reliable input-output analysis built into our unit-testing infrastructure (and the vignette documentation.) For example, the Introduction to iotables vignette article replicates the calculations of the Chapter 15 "Applications" of the *Eurostat Manual*. We have also cross-checked results with the of the Chapter 20 of the *Handbook on Supply and Use Tables and Input-Output Tables with Extensions and Applications* published by the United Nations (Beutel et al. 2018).

The type-II indicators and multipliers consider the induced effect of changes in household demand. Neither of the above publications published type-II examples, so we chose to replicate the *Input-Output Multipliers – Specification sheet and supporting material* from the *Spicosa Project Report*, because it contained a very detailed and useful documentation with a small IOT for the Netherlands in 2006 (D'Hernoncourt, Cordier, and Hadley 2011).

Another, much larger cross-validation is the comparison of direct effect indicators and multipliers with the statistical publication of the *United Kingdom Input-Output Analytical Tables 2010*. The inter-industry matrix published by the Office for National Statistics is unusually large, as it has 127 rows and columns. We have written functions for the reproducible download of data and published analytical results from the website of the UK statistical authority, and compared our results with their published results with the help of the editor of original publications. This important cross-validation is published as a separatte 'vignette' article on the website of iotables titled United Kingdom Input-Output Analytical Tables.

Limitations and Directions for Development

The package grew out of the *eurostat* package, and its reproducible data importing functionality relies on the Eurostat data warehouse, which contains IOTs from the European Economic Area and a few select developed nations like the United States and Japan. The dictionaries that process the data from a long-form dataset to correct matrixes (the current `metadata` dataset) follows the European classifications and dictionaries of economic activities (NACE) and products (CPA). These classifications are slightly modified versions of ISIC and CPC classifications of the United Nations (Eurostat 2008b; EUR-Lex 2006).
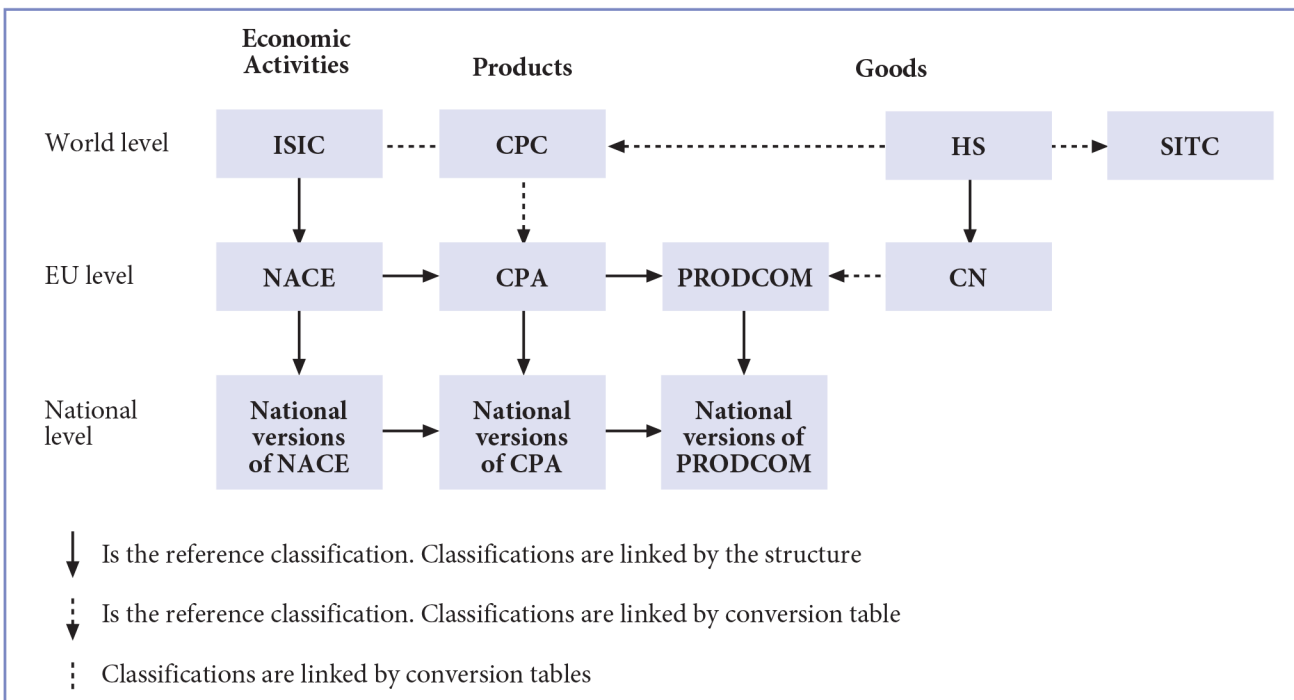
Source: NACE Rev. 2. Statistical classification of economic activities in the European Community (Eurostat)

Within Europe, there may be slight national modifications of NACE, or even the IOT definition. For example, Switzerland uses IOTs that are in most applications work perfectly well with our functions, as the country-specific modification is extremely unlikely to enter any real life application, but still, it is not fully harmonized with the Eurostat format and it must be imported manually from Excel spreadsheets. Outside of Europe, a case-by-case adjustment is usually very straightforward. Differences from ISIC and CPC are marginal and easy to reconcile. In the future, we will add a tutorial on such a manual adjustment, but with an understanding of IOTs, this should not be a problem for a knowledgable user.

A case-by-case adjustment with an R code can keep the research flow reproducible, but with coding input. In future releases we will look for other large data sources for extending the functionalities of the importing functions `iotables_download` and `iotable_get` for the data structuring, formatting and labelling idiosyncraticities from non-European data sources that contain more than one country's tables.

Acknowledgments

References

Antal, Daniel. 2019. "Slovak Music Industry Report [Správa o slovenskom hudobnom priemysle]." https://doi.org/10.17605/OSF.IO/V3BE9.

———. 2021. *Regions: Processing Regional Statistics.* https://regions.dataobservatory.eu/.

———. 2022. *Iotables: Reproducible Input-Output Economics Analysis, Economic and Environmental Impact Assessment with Empirical Data.* https://iotables.dataobservatory.eu/.

Beutel, Joerg, Simon Guerrero, Satoshi Inomata, Soren Larsen, Brian Moyer, Isabelle Remond-Tiedrez, José M. Rueda-Cantuch, et al. 2018. *Handbook on Supply and Use Tables and Input-Output Tables with Extensions and Applications.* Edited by Sanjiv Mahajan. Vol. 1. Studies in Methods, F 74. New York: United Nations.

Blondel, Emmanuel. 2021. *Rsdmx: Tools for Reading SDMX Data and Metadata.* https://CRAN.R-project.org/package=rsdmx.

Botrić, Valeria. 2013. "Identifying Key Sectors in Croatian Economy Based on Input-Output Tables." Ekonomski institut - The Institute of Economics. http://www.eizg.hr/Download.ashx?FileID=010e9ee3-e20a-4d41-a59d-d65fd3d400cf.

D'Hernoncourt, J, M Cordier, and D Hadley. 2011. "Input-Output Multipliers – Specification Sheet and Supporting Material, Spicosa Project Report." Brussels: Université Libre de Bruxelles – CEESE. http://www.coastal-saf.eu/output-step/pdf/Specification%20sheet%20I_O_final.pdf.

EUR-Lex. 2006. "Regulation (EC) No 1893/2006 of the European Parliament and of the Council of 20 December 2006 Establishing the Statistical Classification of Economic Activities NACE Revision 2 and Amending Council Regulation (EEC) No 3037/90 as Well as Certain EC Regulations on Specific Statistical Domains (Text with EEA Relevance)." *Official Journal* L 393 (1). http://data.europa.eu/eli/reg/2006/1893/2019-07-26.

———. 2008. "Regulation (EC) No 451/2008 of the European Parliament and of the Council of 23 April 2008 Establishing a New Statistical Classification of Products by Activity (CPA) and Repealing Council Regulation (EEC) No 3696/93 (Text with EEA Relevance)." *Official Journal* L 145 (65). ttp://data.europa.eu/eli/reg/2008/451/oj.

Eurostat. 2008a. "CPA 2008 Introductory Guidelines." Eurostat. https://ec.europa.eu/eurostat/documents/1995700/1995914/CPA2008introductoryguidelinesEN.pdf/df1e8d19-1156-4a1c-b384-4f95a12515e5.

———. 2008b. *NACE Rev. 2. Statistical Classification of Economic Activities in the European Community.* Methodologies and Working Papers. Luxembourg: Office for Official Publications of the European Communities. https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF.

Ghosh, A. 1958. "Input-Output Approach in an Allocation System." *Economica* 25 (97): 58–64. http://www.jstor.org/stable/2550694.

Lahti, Leo, Janne Huovari, Markus Kainu, and Przemyslaw Biecek. 2017. "Eurostat r Package." *R Journal* 9 (1): 385–92. https://journal.r-project.org/archive/2017/RJ-2017-019/index.html.

Leontief, Wassily W. 1951. "Input-Output Economics." *Scientific American* 185 (October): 15–21. https://doi.org/10.1038/scientificamerican1051-15.

Wade, John, and Ignacio Sarmiento-Barbieri. 2020. *Ioanalysis: Input Output Analysis.* http://www.real.illinois.edu.

Weinert, Karsten. 2014. *Datamart: Unified Access to Your Data Sources.* https://CRAN.R-project.org/package=datamart.