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

Daniel Antal, CFA

1/21/2022

Introduction

Several years have passed since the first release of the eurostat R package (Lahti et al. 2017), which helps the reproducible retrieval and analyzis 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 idiosynchartic 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 deals (Antal 2022) with a different problem, the analytical inter-dependency of many statistical data elements within the system of national accounts. What connects these packages is that they utilize standarized statistical metadata to improve the usability of the upstream *eurostat* package.

The supply and use tables (SUTs) and input-output tables (in short: IOTs) provide a very detailed, and empirically measured or estimated picture of the economy. The 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.

These tables show the inter-linkages of a national or regional economy both from a supply and demand perspective. Where various sectors of the economy typically purchase goods and services from each other as they work themselves towards the final consumers or towards exporting outside of the economy described with the data. How the economy distributes corporate, personal and government income, how they finance consumption and investments. They system of input-output tables are the most comprehensive, empirically measured data for many types of macroeconomic research, industry organization analysis, or they provide tools for various economic and environmental impact analysis.

There are several R pacakges 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 sources, and it is also very widely used to access data from the European Statistical System.

The `eurostat::get_eurostat` function retrieves the relevant supply, use, symmetric-input output or margin

tables in a tidy, long-form. These tidy, long datasets cannot be meaningfully spread into a tidy, wide format without understanding and properly ordering the rows and columns of those datasets. 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.

Since the first release of the *iotables* packages, 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 to process hard-to-access empirical data for input-output analysis without error, and to allow the similarly 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 too packages. Particularly to overlap with *leontief* is so great that in the future we are planning to paralell develop both packages.

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. The analytical functions of *iotables* make sure that otherwise relatively simply algebraic equations in input-output analysis are performed on meticoulusly matched, conforming matrixes. These functions contain many unit tests. Whenever the user would try to work with non-conforming matrixes, a simple base R error message would show the problem. However, matrixes with wrong ordering or labelling may give analytically possible but logically wrong results that are next to impossible to debug from the results alone. 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*. The United Kingdom Input-Output Analytical Tables vignette replicates the calculations of the *United Kingdom Input-Output Analytical Tables 2010* document of the Office for National Statistics for testing the analytical functions and type-I multipliers (Wild 2014). The type-II multipliers were testing with the replication of the *Input-Output Multipliers – Specification sheet and supporting material, Spicosa Project Report* for testing type-II multipliers (D'Hernoncourt, Cordier, and Hadley 2011). And at last, we 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 `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*.

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.

```
library(iotables)
germany_io <- iotable_get( 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 *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. Whilst input-output economics has a fairly standard analytical method, it has a long history, and very different applications in macroeconomic analysis, antitrust, tourism economics, cultural economics, or environmental impact assessment, to name a few. Different disciplines have incorporated the use of the input-output tables with a slightly different vocabulary. 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, reproducbile access to the tables harmonized by the Eurostat statistical agency. However, in the package documention we tried to describe what are the other often used names for these matrixes. For example, the input flow matrix in 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])
```

```
##        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 `leontieff_matrix_create` and `leontieff_inverse_create` create the most important object of input-output analysis, first described, and named in honour of Leontief (Leontief 1951). The `ghosh_inverse_create()` the inverse from the 'supply-driven' input-output model developed by (Ghosh 1958).

```
L_de <- leontieff_matrix_create (
  technology_coefficients_matrix = de_input_coeff
  )
I_de   <- leontieff_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 greater details.

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: Importing and Manipulating Symmetric Input-Output Tables.* 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.

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.

Wild, Richard, ed. 2014. "United Kingdom Input-Output Analytical Tables 2010." Office for National Statistics. https://webarchive.nationalarchives.gov.uk/ukgwa/20160114044923/http://www.ons.gov.uk/ons/rel/input-output/input-output-analytical-tables/2010/index.html.