

# DATA ON THE FRAGMENTATION OF PRODUCTION IN THE US

Thibault Fally\*  
University of Colorado-Boulder

March 2012

This document describes the files available on:

[spot.colorado.edu/~fally/data.html](http://spot.colorado.edu/~fally/data.html)

The files contain part of the data used in the paper “On the Fragmentation of Production in the US” available here:

[spot.colorado.edu/~fally/Fragmentation.pdf](http://spot.colorado.edu/~fally/Fragmentation.pdf)

and the paper “A Measure of Upstreamness of Production and Trade Flows” available here:

[spot.colorado.edu/~fally/ACFH\\_short.pdf](http://spot.colorado.edu/~fally/ACFH_short.pdf)

## List of files made available on this webpage

1. “List\_industries.xlsx”: List of industries in the 1967-1992 data, in Excel format.
2. “Concordance\_manufacturing\_SIC87.xlsx”: Concordance between 1967-1992 industry codes and 1987 SIC codes for manufacturing industries, in Excel format.
3. “stages\_panel.xlsx”: Index  $N$  (for the number of production stages) for 1967-1992 by industry, in Excel format.
4. “stages\_panel.dta”: Same as above, in STATA format.
5. “distance\_panel.xlsx”: Index  $D$  (for the number of stages to final demand or “upstreamness”) for 1967-1992 by industry, in Excel format.
6. “distance\_panel.dta”: Same as above, in STATA format.

---

\*Contact: Department of Economics, University of Colorado-Boulder, 256 UCB Boulder, Colorado 80309-0256, USA. [fally@colorado.edu](mailto:fally@colorado.edu), Ph: 303-492-5562.

7. “Stages\_by\_industry2002.xlsx”: Index  $N_i$  (for the number of production stages) for 2002 by industry, in Excel format.
8. “Stages\_by\_industry2002.dta”: Same as above, in STATA format.
9. “Upstreamness\_by\_industry2002.xlsx”: Index  $D_i$  (for the number of stages to final demand or “upstreamness”) for 2002 by industry, in Excel format.
10. “Upstreamness\_by\_industry2002.dta”: Same as above, in STATA format.
11. “Do-file-N-and-D-2002.do”: STATA do-file to construct both  $N$  and  $D$  indexes for 2002 using the file “iousedetail2002.dta”.
12. “Do-file-N-and-D-2002.txt”: Same as above in a .txt format.
13. “iousedetail2002.dta”: This is a STATA version of first 5 columns of the original file “iousedetail.txt” from the BEA available in the following zipped package:  
[http://www.bea.gov/industry/zip/2002detail\\_redef.zip](http://www.bea.gov/industry/zip/2002detail_redef.zip)
14. “Supplementary+Material+AERPnP”: Corresponds to the supplementary material associated with Antras, Chor, Fally and Hillberry (2012) available with the Papers and Proceedings version on the AEA website.
15. “concordance\_hs6\_IO.xlsx”: Concordance table from the HS classification to the 2002 input-output classification, in Excel format.
16. “concordance\_hs6\_IO.dta”: Same as above, in STATA format.
17. “Notes\_data\_fragmentation.pdf”: This is the present document.

## Classification of industries

Industry classifications for the 1967-1992 panel data and the 2002 data are both available at a very disaggregated level (6-digit level). Moreover, the 1967-1992 classifications can be fairly harmonized across years by aggregating industries when some industries have been split or merged over time. We end up with 378 industries from 1967 to 1992 and 426 industries for 2002.

## 1967 - 1992 data

In order to obtain a harmonized classification across years, some aggregations were necessary when some industries were either aggregated or split for some years but not other, or split in different ways depending on the year. Industries where such adjustment where made have a letter "A" or "X".

- industries ending with A are aggregating all sub-industries. For instance, industry "080A00" corresponds to Crude petroleum and natural gas. For some years, petroleum and natural gas were differentiated into 080100 and 080200, for some other years, it was classified into one unique category 080000. I thus aggregated 080100 and 080200 into one category renamed 080A00 (080000 was also renamed 080A00). Similarly, "76010A" is the aggregation of all sub-industries starting by "7601".
- An industry ending with "X" refers to an aggregation of sub-industries otherwise not listed. For instance, industry "24070X" ("Converted paper products, n.e.c.") is the aggregation of all industries in "2407" other than "240701", "240702" and "240703".

The classification of industries until 1992 is based on the SIC system. For manufacturing industries, I can provide a concordance table to the SIC classification (1987 version). In this concordance table, each SIC product is matched to a unique code in the input-output classification.

## 2002 data

For the 2002 data, I simply keep the same classification as in the BEA input-output table with no changes. The BEA classification is based the NAICS industry classification but some industries have been aggregated compared to the 6-digit NAICS classification.

I also provide a concordance table between the 2002 input-output classification and the harmonized system classification (HS) at the 6-digit level. This corresponds to an aggregation of the concordance table with the 10-digit HS classification provided by the BEA. While there is a unique input-output category for each 10-digit industry, there are sometimes several input-output categories for each 6-digit HS classification. I provide weights (the sum being one for each HS product) based on the number of 10-digit HS products associated with a particular IO code. This is how we matched trade data to our upstreamness index in Antras, Chor, Fally and Hillberry (2012).

## Treatment of “make” and “use” tables

“Make” and “use” industry-by-commodity tables are available from 1972 onward. For the years 1972 to 1992, I combine information from these two tables to construct a commodity-by-commodity table and estimate the amount of commodity  $j$  (input) used to produce commodity  $i$  (output).<sup>1</sup>

“Use” tables describe the value of purchases  $u_{kj}$  of input  $j$  by industry  $k$ , while “make” tables describe the value of production  $m_{ki}$  of output  $i$  for each industry  $k$ . I construct commodity-by-commodity input-output coefficients  $\mu_{ij}$  by taking the average share of input  $j$  in production of industry  $k$  weighted by the contribution of industry  $k$  to the production of output  $i$ :

$$\mu_{ij} = \sum_k \left[ \frac{m_{ki}}{\sum_{k'} m_{k'i}} \frac{u_{kj}}{\sum_{j'} m_{kj'}} \right]$$

where  $\sum_{k'} m_{k'i} = Y_i$  corresponds to total production of output  $i$  and  $\sum_{j'} m_{kj'}$  corresponds to total production of industry  $k$  – this method is based on the “industry-technology assumption” (see Guo et. al., 2002).

Note that this way of constructing input-output coefficients  $\mu_{ij}$  is consistent with the construction of coefficients  $\varphi_{ij}$  measuring the fraction of output  $i$  used for production of output  $j$  if they are defined as:

$$\varphi_{ij} = \sum_k \left[ \frac{u_{ki}}{(\sum_{k'} u_{k'i} + u_{Fi})} \frac{m_{kj}}{\sum_{j'} m_{kj'}} \right]$$

where  $\sum_{k'} u_{k'i} + u_{Fi}$  includes the use of product  $i$  by all industries plus final demand. In an open economy, this corresponds to total absorption  $Y_i + M_i - X_i$  i.e. domestic production plus net imports, as discussed in section 2.2. We can verify that:

$$\varphi_{ij} = \frac{Y_j \mu_{ji}}{Y_i + M_i - X_i}$$

Note also that this way to construct input-output coefficients is consistent with aggregation properties discussed in the text. In particular, we find that total value added  $\sum_i V_i$ , where value-added is defined by  $V_i = (1 - \sum_j \varphi_{ij})Y_i$  as in the text, equals total production  $\sum_{k,i} m_{ki}$  minus total use of inputs  $\sum_{k,j} u_{k,i}$ .

---

<sup>1</sup>The 1967 input-output table is treated as a commodity-by-commodity table. I obtain very similar results by extrapolating a “make” table from other years to adjust input-output coefficients. For 2002, both indexes  $D$  and  $N$  are based solely on the “use” table as we did in Antras *et al* (2012) for index  $D$ .

## Treatment of “non-comparable” and “transferred” imports

In the 1972 table and after, the sum of each column of the use table provides production for each industry (sum of value-added and intermediate purchases). Intermediate goods imports are reported as part of input usage  $u_{kj}$  as described above. Total imports and exports by product are also reported in two of the last columns.

A small share of imports, however, are reported as “non-comparable” and correspond to a distinct row in the list of inputs. These non-comparable imports correspond to product that are different from any product produced in the US such as coffee and cocoa beans. Since I need to have an estimate of the number of production stages necessary to produce all inputs (even if those goods are imported), I make changes in the data for two industries: I assume that all non-comparable imports by the coffee-roasting industry (industry 142800) and the chocolate industry (industry 142002) correspond to imports of coffee and cocoa beans respectively and are comparable to “tree nuts” (commodity 020401). These two changes reduce the amount of non-comparable imports of intermediate goods by more than half and the remaining non-comparable account for less than half a percent of total production value (and are thus dropped).<sup>2</sup>

The 1967 input-output table has a different treatment for imports and a few other corrections are needed. Imports are classified in two categories: “non-comparable” imports as described above and “transferred” imports. “Transferred” imports are recorded in two places and would be double-counted if not carefully taken into account. In particular, the column-sum of the 1967 I-O table gives the sum of domestic production plus “transferred” imports classified in the same product category. Hence we need to subtract “transferred” imports to obtain domestic output. Note however that “transferred” imports of intermediate goods also appear in input-output coefficient for each input category.<sup>3</sup> In terms of final consumption in 1967, some imports destined for final consumption are classified as “non-comparable” imports while they are actually quite comparable and may account for a large share of absorption in these industries: for instance, most imports of cars are classified as “non-comparable” in the 1967 consumption data. I use import data from the NBER trade database (Feenstra, 1996) to adjust the amount of imports for consumption in industries where more than half of imports are missing.

---

<sup>2</sup>Note that the 1992 table significantly reduced the “non-comparable imports” category by associating these imports with other classified commodities. In particular, the coffee-roasting and chocolate industries in 1992 exhibit large uses of inputs classified as “tree nuts” instead of non-comparable imports, which is consistent with the changes made on earlier tables.

<sup>3</sup>For instance, imports of crude petroleum to be used by the petroleum refinement industry appear twice: in the row for transferred imports in the column of crude petroleum, and also in the row for crude petroleum in the column for petroleum refinement.