

# A scale-invariant calculation of import contents of exports

Luzius Meisser, [luzius@meissereconomics.com](mailto:luzius@meissereconomics.com)

*The most popular tool to calculate import content of exports (also known as ...) from given reuse of exports from given input-output tables. Unfortunately, it yields wrong results when the inputs of sectors are not proportionally used for the production of their outputs, which usually is the case. As a consequence, input reuse as calculated by the Leontief inverse depends on the resolution of the underlying data, monotonically increasing from 17% to 24% globally as the resolution is increased from one aggregate to 35 sectors per country. To address this problem, I postulate a hidden variable 'gravity of local consumption' and introduce a scale-invariant calculation method to estimate the true import reuse under the assumption of self-similarity, resulting in about 32% globally.*

## 1 Introduction

Import reuse, sometimes also called the import content of exports, is a basic metric that is typically derived from input-output tables and often serves as an intermediate step for more elaborate calculations, for example as a contributing factor when estimating exchange-rate pass-through. It measures the extent to which exports consist of previous imports. Its usual method of calculation is to derive the composition of exports by applying the Leontief inverse to the input-output table of the exporting country. ?? More accurate results are achieved when basing the calculation on a worldwide input-output table such as WIOD, as this allows to account for circular flows of goods between country. ??

The subtle problem with this traditional approach is that the Leontief inverse implicitly assumes proportional use of inputs. In practice, however, firms that import much also tend to export much. ?? When aggregating such firms together with firms that have low import intensity, information about the actual use of specific imports gets lost and the import reuse as calculated by the Leontief inverse decreases. Thus, when the import intensity and the export intensity of sectors beyond the resolution of the data at hand are positively correlated, the Leontief inverse systematically underestimates import reuse. This is the case for all the 42 countries present in the WIOD database, except for Russia, which exhibits a negative correlation.

After illustrating how strongly the reported import reuse depends on the resolution of the data, I construct a simple heuristic to artificially increase the resolution, thereby allowing to better estimate the true import reuse under

the assumption of self-similarity between sectors at different resolutions. The source code of the used computer program is provided in a public repository<sup>1</sup>, allowing anyone skilled in the art to reproduce these results within minutes and to calculate import reuse more accurately on their own datasets.

Subsequent section 2 describes the used data. It is followed by a detailed description of ...

Koopman, Robert, Zhi Wang, and Shang-Jin Wei. "Estimating domestic content in exports when processing trade is pervasive." *Journal of development economics* 99.1 (2012): 178-189. Recognize short-comings of Leontief Inverse, manually introduce extra nodes to separate imports that are sent abroad again, which they call "processing trade". Estimate China to have 50

Seminal paper with 2600 citations: Hummels, David, Jun Ishii, and Kei-Mu Yi. "The nature and growth of vertical specialization in world trade." *Journal of international Economics* 54.1 (2001): 75-96. Take into account back-and-forth flows between countries. The vertical specialisation (VS) share of world trade is about 25%. Matrix view, apply leontief inverse. Problematic: varying numbers of sectors across time (?), page 27

The literature on vertical trade aims at measuring sequential trade in vertical production chains by looking at the import content of exports. "Mexico's manufacturing exports have a foreign content share of approximately 66%". Literature World Bank Collection.pdf

## 2 Data

### 2.1 WIOD

The world input-output database (WIOD) presented by [3] serves as data source. It covers 42 countries and 17 years starting from 1995 with a resolution of 40 sectors per country. Unlike national input-output tables, the WIOD tables contain disaggregate flows between the individual sectors of different country. For simplicity, I merge each country's destination sectors such as consumption and inventory changes into one node "consumption and capital formation" that is treated specially and generally not counted as a sector. I.e., when later reducing the number of sectors per country to one, there will actually be two nodes left per country, one active sector and the consumption node. Furthermore, I ignore negative flows into the consumption node<sup>2</sup> as they are negligible and not having negative edge weights is a prerequisite for the later applied Edmonds-Karp algorithm [1]. All other flows in WIOD are zero or positive.

### 2.2 Graph View

Usually, economists treat input-output tables as matrices. An equally valid and in this case more insightful view is to treat the input-output table as a weighted

<sup>1</sup> [github.com/kronrod/importreuse](https://github.com/kronrod/importreuse), with more detailed instructions in the repository itself.

<sup>2</sup> This can happen when capital is reduced or inventories decrease.

directed graph. Every square matrix can be represented as a weighted directed graph and a weighted directed graph can be converted into a square matrix as long as there is at most one edge in each direction between each pair of nodes. In the graph representation of input-output tables, each node represents a sector in a country and is named accordingly. Each weighted edge  $e = (a, b, w)$  represents a flow from node  $a$  to node  $b$  of volume or weight  $w$ . Flows from a node to itself are allowed and correspond to diagonal elements in the matrix view.

When merging two nodes  $a$  and  $b$  of a graph into a new node  $c$ , the nodes  $a$  and  $b$  are replaced by  $c$  in all edges and then all edges that connect the same nodes in the same direction are aggregated into a single new edge whose weight is the sum of the old weights. This is equivalent to first removing column  $i$  and adding it to column  $j$  in the matrix view, and then also removing row  $i$  and adding it to row  $j$ , with  $i$  and  $j$  being the indices of the two sectors represented by nodes  $a$  and  $b$ . When doing so, the indices of other sectors may change, making it less convenient to track a specific sector in the matrix view than in the graph view, where the names of unaffected nodes stay the same when others are merged. Generally, the graph view is more compact than the matrix view when the matrix is sparse.

### 3 Empirical Part

#### 3.1 Varying the Sector Resolution

The usual method to calculate input reuse is to apply the Leontief inverse to the input-output table. A known flaw of this approach is that it implicitly assumes proportional use of inputs.<sup>3</sup> This should not be a problem for disaggregate data with atomic resolution, but when applied to aggregate data, the Leontief inverse is systematically biased with countries that prefer consuming domestic products over foreign products. The import reuse as calculated by the Leontief inverse usually grows as the resolution of the underlying data is increased, which is a strong hint that it cannot be trusted. In case of the WIOD data, world input reuse grows from 17% to 24% as the resolution is varied between one and thirty-five sectors, a result discussed in section 4.1. Thus, the Leontief inverse is not scale-invariant. A scale-invariant computation method that is consistent across different resolutions would yield more credible results, which is exactly what will be constructed in the subsequent sections.

#### 3.2 Maximum Flow

Given the input-output graph of a country, one can derive the minimum and maximum possible import reuse by computing the maximum amount of imports

<sup>3</sup> This is *not* what is usually called the proportionality assumption. The proportionality assumption in the context of input-output analysis usually means proportionally assigning an import like steel across the domestic sectors that use that import. Here, proportional use of inputs refers to proportionally assigning the inputs of a single sector to its outputs. To avoid confusion, I refer to the latter as *proportional assignment*.

that can be directed towards domestic consumption (for the minimum) or towards foreign nodes as exports (for the maximum). This problem is known as the *maximum flow problem* and can for example be solved with the Edmonds-Karp algorithm. [1] The true import reuse must lay within the calculated bounds, with the proportional variant being somewhere in between. Unfortunately, these bounds turn out to be so wide that they cannot provide much guidance as to where the true import reuse might be.

### 3.3 Domestic Consumption Preference

To seek for plausible values of import reuse within the given bounds, I introduce a parameter *domestic consumption preference*  $p_{dom} \in [0, 1]$ . It is defined as the fraction of domestic inputs of a sector that directly flows into domestic consumption, before proportionally distributing all the remaining flows.<sup>4</sup> For example, if there are two firms in the beer sector of a country, one producing beer from domestic hops and one producing beer from imported hops, a positive domestic consumption preference indicates that the locals prefer to drink the former and export the latter, whereas the proportional assignment implies indifference.

By exogenously varying  $p_{dom}$ , an interval of potential true values of import reuse can be explored, with  $p_{dom} = 0$  corresponding to the proportional assignment, and  $p_{dom} = 1$  being somewhere between the proportional result and the maximum flow result. It does not quite reach the maximum as merely redirecting the domestic flows of every node towards local consumption is not sufficient. In some cases, it is necessary to redirect locally created value through multiple nodes towards consumption, which is what maximum flow implies, but not  $p_{dom} = 1$  as defined above. In future research, it might be desirable to redefine domestic consumption preference such that it can cover the whole range given by the bounds of maximum flow, including negative values that would correspond to a preference for foreign goods. In the data at hand, this actually seems to be the case for Russia in some years. But for the time being, the definition at hand is broad enough.

### 3.4 Scale-Invariance

Equipped with a parameter to tune, one can try to find values of  $p_{dom}$  that make the calculate import reuse scale-invariant. One can do this globally with one global  $p_{dom}$ , or nationally with a separate domestic consumption preference  $p_{dom,c}$  for each country  $c$ . In both cases, the parameter is adjusted such that the covariance  $cov(r, f)$  of resolution  $r$  (the number of sectors per country) and resulting import reuse  $f$  is minimized. I calculate these values iteratively

<sup>4</sup> To be more precise, the direct flow from sector  $s$  to the consumption node before proportionally distributing the remaining flows is  $\max(w_{s,c}, p_{dom}(v_s + i_{dom,s}))$ , with  $w_{s,c}$  being the weight of the edge connecting  $s$  to the consumption node  $c$ ,  $v_s$  being value creation of  $s$ , and  $i_{dom,s}$  being the total value of the domestic inputs of  $s$ . As a prerequisite for calculating  $i_{dom,s}$ , the import contents of all inputs of  $s$  must be known.

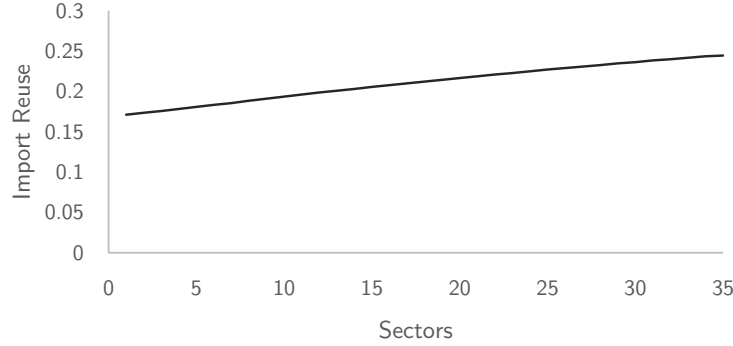


Fig. 1: The calculated import reuse depends on the resolution of the underlying data. The plotted curve represents the average of 100 randomized runs and uses the data of 2005. The standard deviation of each plotted point is less than  $10^{-5}$ , so the curve can be considered free from visible random variation.

simply by updating import contents in the worldwide input-output graph until the values stabilize.<sup>5</sup>

Being robust against the tested changes in resolution, the import reuse resulting from assuming the scale-invariant domestic consumption preference is more plausible than the one calculated with the proportional assignment of inputs. Under the assumption that it is also robust against changes in resolution beyond the limits of the data at hand, the calculated scale-invariant import reuse matches the true import reuse. This assumption is valid if the input-output graph is self-similar, i.e. if the network similarly structured at all scales. Given that a wide range of real-world networks exhibit self-similarity [2], this might be the best possible estimate in the absence of more fine-grained data.

## 4 Results

### 4.1 Varying the Resolution

Figure 1 shows how the global import reuse calculated with the traditional method of applying the Leontief inverse depends on the resolution of the underlying data.<sup>6</sup> As the number of sectors is increased from one to the thirty-five, import reuse continuously climbs from 17% to 24%. This is a clear signal that the true global import reuse is higher than what is typically assumed.

The program to calculate the data behind figure 1 first parses the given world input-output table into a graph, and then repeatedly calculates import reuse and reduces the number of sectors in each country by one. This is done by randomly selecting two sectors in each country and merging them as defined in section 2.2.

<sup>5</sup> This is done in script `BendingByCountry`.

<sup>6</sup> The raw data for this figure has been generated with script `Figure1Resolution`.

The reported import reuse is a global volume-weighted average of the import content of all exports, which in turn are calculated by repeatedly updating the input composition of each node until all values have converged towards a stable equilibrium. This latter step equivalent to applying the Leontief inverse to the matrix view.

When plotting the same curve for the 40 available countries individually, results vary, but all except Russia's import reuse raises as the resolution increases. For Russia, the curve slightly declines, at least for some of the years in the dataset, indicating that the domestic input preference of Russia is slightly negative.

## 4.2 Scale-Invariance

As a next step, domestic consumption preference  $p_{dom}$  is varied as described in section 3.4. Figure ?? shows how the slope of Germany's reuse-resolution curve changes as  $p_{dom}$  is varied. For  $p_{dom} = XXX$  it is flat, with  $cov(r, f) = 0.0001XXX$ . Depending on the country, the scale-invariant value of  $p_{dom}$  varies. Also, it varies over time. Figure ?? shows how Greece's import reuse and domestic consumption preferences changed over time. Both variables have comparable variances over time. TODO: figure with traditional and new import reuse, plus preference for Greece.

Table ?? ranks countries by domestic consumption preference of 2011. The raw data for all other years and other countries for the aforementioned figures can be found in file hrefasdXXX. It is unclear whether much meaning should be attributed to this hidden variable, and if so, what it it should be exactly. Is there a pattern behind the countries with low domestic consumption preference? This might be an interesting question for future research.

## 4.3 Drivers of Import Reuse

Import reuse varies across countries. The most important driver of import reuse is the import volume relative to the country's economy, here measured as import share  $s = imports / (consumption + exports)$ . Import share  $s_{c,y}$  of a country  $c$  in year  $y$  can already explain 72% of the import reuse  $f_{c,y}$ . Together with domestic consumption preference  $p_{dom,c,y}$ , the  $R^2$  is improved to 92% when regressing:

$$f_{c,y} = b_0 + b_1 s_{c,y} + b_2 p_{dom,c,y}$$

Interestingly, domestic consumption preference and size show no significant correlation ( $corr(s_{c,y}, p_{dom,c,y}) = 0.01$ ), confirming that  $p_{dom}$  measures a phenomenon that is independent of a country's size.

## 5 Discussion

- a. Robustness vis-à-vis sub-samples / sub-period
- b. Correlation or causality?
- c. Limitations and further questions

Thought: Considering capital accumulation as consumption lowers the measured import reuse, as it can mask the foreign origin of domestic capital.

## 6 Conclusion

### References

- [1] Jack Edmonds and Richard M Karp. Theoretical improvements in algorithmic efficiency for network flow problems. *Journal of the ACM (JACM)*, 19(2):248–264, 1972.
- [2] Chaoming Song, Shlomo Havlin, and Hernan A Makse. Self-similarity of complex networks. *Nature*, 433(7024):392–395, 2005.
- [3] Marcel Timmer, Abdul A Erumban, Reitze Gouma, Bart Los, Umed Temurshoev, Gaaitzen J de Vries, I-aki Arto, Valeria Andreoni AurŽlien Genty, Frederik Neuwahl, Joseph Francois, et al. The world input-output database (wiod): contents, sources and methods. Technical report, Institutue for International and Development Economics, 2012.