



Scale-Invariant Export Contents from I/O-Tables

Luzius Meisser, luzius@meissereconomics.com

The Leontief inverse is the standard tool to calculate the import content of exports from given input-output tables. Unfortunately, it only yields correct results when inputs are proportionally used for the production of outputs, which rarely is the case. As a consequence, the value calculated by the Leontief inverse depends on the resolution of the underlying data. With the WIOD dataset, global import contents of exports climb from 17% to 25% as the resolution is increased from one to thirty-five sectors per country. I address this known problem by constructing a new, scale-invariant computation method, resulting in more credible 32%.

1 Introduction

Import reuse (also known as import content of exports, foreign contents of exports, foreign value added share of gross exports, foreign value added and vertical specialization Hummels et al. (2001))¹ 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 countries. (Auer et al., 2016)

The known 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

¹ These terms are not always used consistently and do not always refer to the same metric. However, all of them suffer from the same problem when calculated with the Leontief inverse.

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 effect was already observed in 1991 by Mark Planting who compared using 6800 versus 536 sectors, which is why the OECD considers the Leontief result a "conservative indicator of the actual input activity". (OECD, 2000)

Koopman et al. (2012) as well as Kee and Tang (2013), both focusing on China, address this problem by enriching input-output tables with additional data to isolate processing trade. Koopman et al. summarize their approach as follows: "The basic idea is to use information from the standard I/O table to determine sector-level total imports/exports, and information from trade statistics to determine the relative proportion of processing and normal exports within each sector." Similarly to the findings presented in here, they find that actual import reuse is much higher than what the traditional described by Hummels et al. (2001) suggests. In contrast to these previous approaches, I present a method to better estimate import reuse that does not rely on having additional data, but instead calculates a more plausible result from input-output tables alone under the assumption of self-similarity.

Generally, I treat economic trade flows as a graph and not as a matrix, a choice motivated in section 2, which also describes the used data more closely. Section 3 specifies the applied methods, in particular how the introduction a hidden variable *domestic consumption preference* allows the calculated import reuse to be tuned such that it becomes robust against changes in resolution (scale-invariant). This is followed by results section 4, that presents selected findings, and a concluding section 5. All source code, input data, and outputs can be found on github.com/meisserecon/importreuse.

2 Data

2.1 WIOD

The world input-output database (WIOD) presented by Timmer et al. (2012) serves as data source. It covers 40 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. Furthermore, it does not depend as strongly on the proportionality assumption

as other data sources.

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 node² as they are negligible and not having negative edge weights is a prerequisite for the Edmonds-Karp algorithm (Edmonds and Karp, 1972) and other graph operations. All other flows in WIOD are already 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 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 , and then also removing row i and adding it to row j in the matrix view, 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. Generally, the graph view is more compact than the matrix view when the matrix is sparse and enables representations of the data that strike me as more intuitive than a matrix.

² This can happen when capital is reduced or inventories decrease.

3 Method

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.³ This is no 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 25% as the resolution is varied between one and thirty-five sectors, a result presented 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 that can flow into domestic consumption (for the minimum) or into exports (for the maximum), given the edges including their weights. This is known as the *maximum flow problem* and can for example be solved with the Edmonds-Karp algorithm. (Edmonds and Karp, 1972) The true import reuse must lay within these 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 This is *not* what is usually called the proportionality assumption. The proportionality assumption in the context of input-output analysis refers to proportionally assigning an import like steel across the domestic sectors that use that import, depending on their size. In contrast, proportional use of inputs refers to proportionally assigning the inputs of a single sector to its outputs.

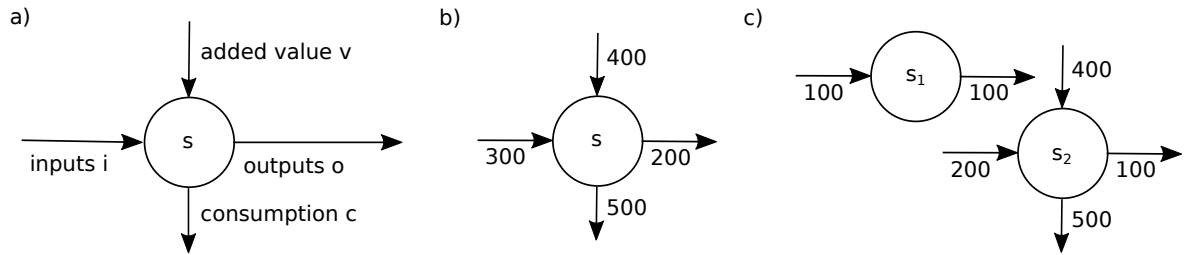


Figure 1: In order to separate processing trade, the flows of a sector a are divided into four groups.

There is the sum of all inputs i , flows towards domestic consumption c , and the sum of all other outputs o , thereby implicitly providing the added value $v = c + o - i$. Given a processing trade share s_{proc} , the existing sector a is split into two new nodes, with $a1$ only having horizontal processing flows of magnitude $i_{a1} = o_{a1} = \min(o, s_{proc}i)$. In the shown example split, s_{proc} is $\frac{1}{3}$.

3.3 Processing Trade

To seek for plausible values of import reuse within the given bounds, I introduce a parameter *processing trade share* $s_{proc} \in [0, 1]$. It is defined as the share of inputs of a node that are directly assigned to its non-consumption outputs, before proportionally distributing all the remaining flows. This is equivalent to splitting the node in two, namely a processing node and a normal node as illustrated in figure 1. I also tested other variants of reassigning the flows of a node, all yielding similar results. The separation of processing trade flows was chosen because it is simple and the most compatible with existing literature. Other tested approaches involved recursive separation of inputs into imports and domestic sources, as it is done by the maximum flow algorithm. Such approaches allow to explore a wider range of values of import reuse, but the range provided by the presented purely local approach already provides sufficient freedom.

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 use of inputs, and $p_{dom} = 1$ being somewhere between the proportional result and the maximum flow result. Even though $p_{dom} = 1$ does not quite reach the theoretic maximum, it is good enough for our purpose.⁴ In future research, it might be desirable to redefine domestic consumption preference such that it can cover the whole range given by the

⁴ In some cases, maximum flow can only be reached by redirecting locally created value over multiple nodes towards consumption. The Edmonds-Karp algorithm does this, but my method does not.

bounds of maximum flow, including negative values that would correspond to a preference for foreign goods. In the data at hand, the only country that would have a slightly negative domestic consumption preference in some of the given years is Russia, so restricting $p_{dom} \geq 0$ is unproblematic.

3.4 Scale-Invariance

Equipped with a parameter to tune, one can try to find values of p_{dom} that make the calculated 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 simply by updating import contents in the worldwide input-output graph until the values stabilize.

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 use of inputs. Under the assumption that it is also robust against further 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 according to Song et al. (2005) and that power laws are common in economics (Gabaix, 2016), this might be the best possible estimate in the absence of more fine-grained data.

4 Result

4.1 Varying the Resolution

Figure 2 shows how the global import reuse calculated with the traditional method of applying the Leontief inverse depends on the resolution of the underlying data.⁵ As the number of sectors is in-

⁵ The raw data [Figure1Resolution.out](#) for this figure has been generated with script [Figure1Resolution](#), which outputs the average result of 100 randomized runs.

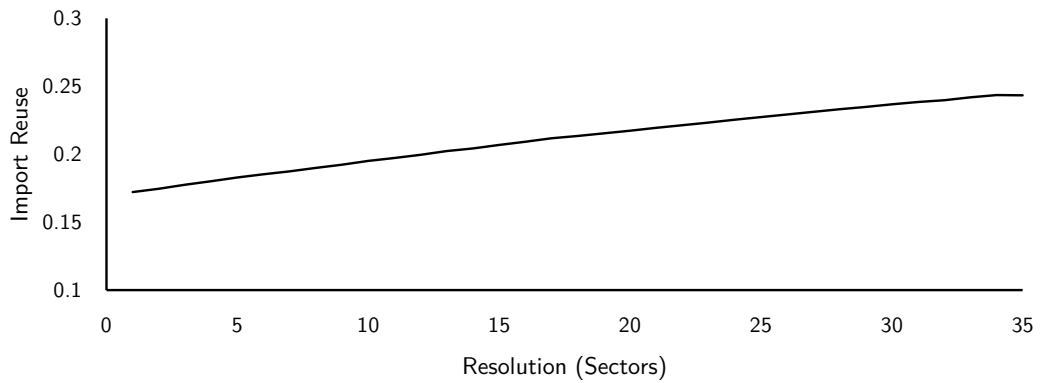


Figure 2: The average global import reuse as calculated by the Leontief inverse grows from 17% to 25% as the resolution is increased from one to thirty-five sectors, based on 2011 WIOD data.

creased from one to the thirty-five, import reuse continuously climbs from 17% to 25%. 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 2 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. 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 raise as the resolution increases. For Russia, the curve slightly declines, at least for some of the years in the dataset, indicating that the domestic consumption preference of Russia is slightly negative.

4.2 Scale-Invariant Domestic Consumption Preferences

As a next step, domestic consumption preference p_{dom} is varied as described in section 3.4. Figure 3 shows how the slope of the previously shown curve changes as p_{dom} is increased from 0.0 (equivalent to the previous figure) to 1.0 in steps of 0.05. 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 4 shows how domestic consumption preference, scale-invariant import reuse, and Leontief import reuse

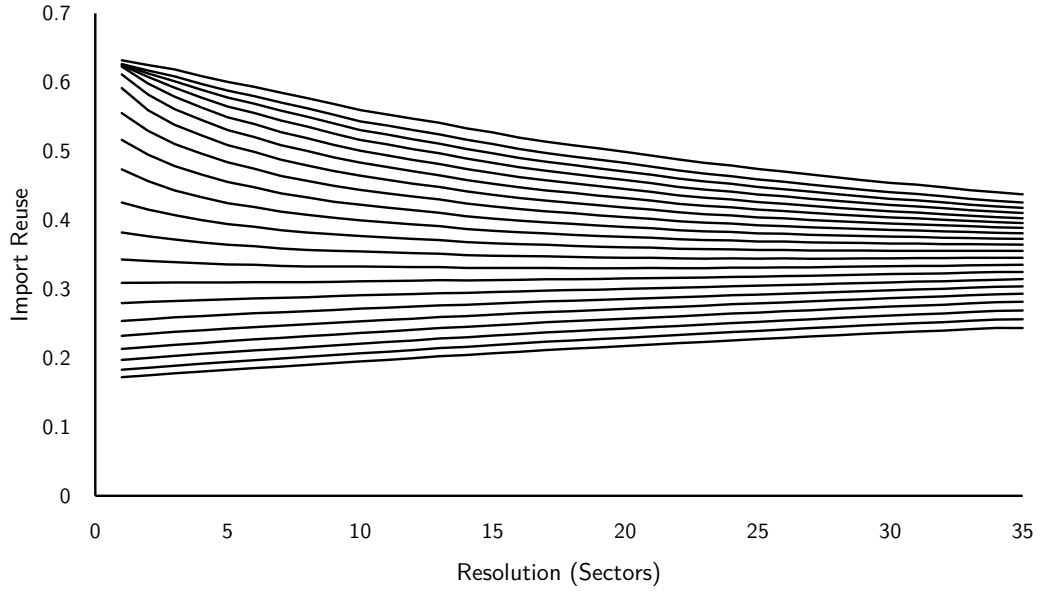


Figure 3: Increasing domestic consumption preference from 0.0 to 1.0 in steps of 0.05 results in 21 curves with varying slope, all converging towards the true import reuse. If the input-output graph is self-similar, a flat curve will stay flat. If not, i.e. if the structure changes at higher resolutions, the curves can take turns and converge to any value within the bounds given by the Edmonds-Karp algorithm (dotted lines).

change over time for the United States. All variables have comparable variances. Table 1 ranks countries by domestic consumption preference of 2011.

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 $s = \text{imports} / (\text{consumption} + \text{exports})$. Import share $s_{c,y}$ of a country c in year y can already explain 72% of import reuse $f_{c,y}$. Together with domestic consumption preference $p_{dom,c,y}$, the R^2 is improved to 92% when regressing:

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

Both variables are highly significant (t-stat above 40). Given that $p_{dom,c,y}$ is the variable used for minimizing $\text{corr}(f_c, r)$ in each year, this is not surprising. However, $p_{dom,c,y}$ even has significant explana-

Country	Dom. consumption pref.	Scale-invariant reuse	Leontief reuse
Russia	0	0.06093049	0.06193057
Indonesia	0.13330078	0.16617329	0.14628616
China	0.23779297	0.29366705	0.21562566
Luxemburg	0.2569336	0.62844914	0.61236697
Ireland	0.25927734	0.49142683	0.44567347
Australia	0.26220703	0.175439	0.13742872
Latvia	0.31137696	0.32521242	0.24283995
Estonia	0.33120117	0.38466322	0.32733306
Korea	0.34760743	0.49048623	0.40077987
Romania	0.35039061	0.33335215	0.23907939
Malta	0.35117188	0.46638846	0.39485261
Bulgaria	0.36523438	0.47888938	0.346733
Hungary	0.37539062	0.54578656	0.45960322
Taiwan	0.37773436	0.51886541	0.47280994
Canada	0.39414063	0.29781705	0.2022918
Czech Republic	0.39414063	0.55580229	0.46454751
Slovenia	0.39726561	0.42307466	0.36454636
Netherlands	0.41088867	0.46661356	0.3924354
Slovakia	0.4144043	0.51753253	0.41957343
Finland	0.41879883	0.44981	0.34448382
Denmark	0.42319337	0.42534709	0.37145129
United Kingdom	0.42319337	0.30213764	0.21590012
Germany	0.4275879	0.36432379	0.27240592
Sweden	0.43637696	0.41681033	0.31875575
Belgium	0.43813476	0.54269993	0.45950785
Austria	0.44340819	0.45071855	0.34209871
Lithuania	0.44604492	0.44599289	0.33458024
Cyprus	0.46533203	0.36767864	0.26017001
Poland	0.47509766	0.49195138	0.34321958
Brazil	0.48388672	0.2490387	0.11892759
Japan	0.49267578	0.28432676	0.16841628
Italy	0.49658203	0.42533097	0.27062058
Turkey	0.50676268	0.37992546	0.22193924
Spain	0.52072752	0.43563309	0.29691294
Greece	0.52824706	0.32894555	0.24268779
Portugal	0.53791505	0.44851738	0.27894515
India	0.54328614	0.38483444	0.21687698
United Stats	0.56279296	0.29106486	0.14915283
Mexico	0.56455076	0.40132162	0.30031794
France	0.59091794	0.44226861	0.28516418

Table 1: Ranking countries by domestic consumption preference in 2011. Unlike import reuse, p_{dom} does not depend much on a country's size. But does it describe a meaningful quality?

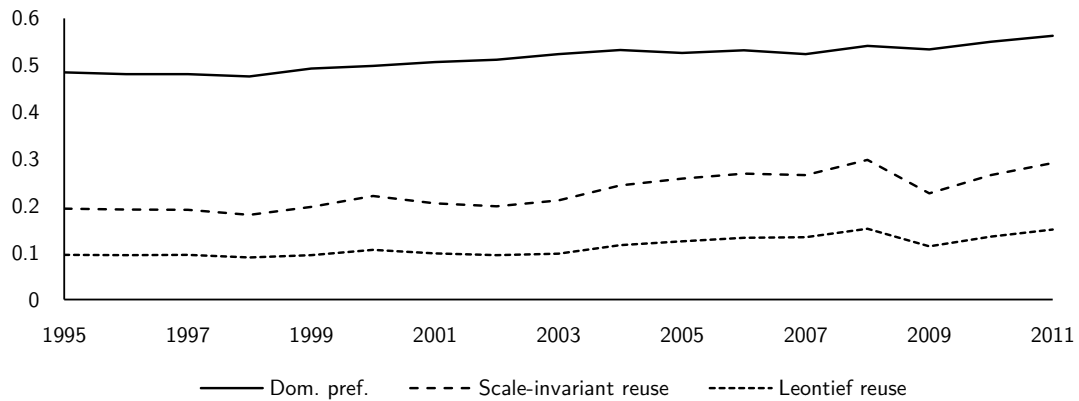


Figure 4: Historic changes of domestic consumption preference, scale-invariant import reuse, and Leontief import reuse of the United States.

tory power for the traditional Leontief import reuse (t-stat 17.58). All the underlying data is available as [reusedrivers.dta](#) and [reusedrivers.csv](#), and the stata output as [reusedrivers.pdf](#).

5 Conclusion

I demonstrated that the standard approach of calculating input reuse from input-output tables strongly depends on the resolution of the underlying data and thus cannot be trusted when applied to the highly aggregate data usually provided by input-output tables. To address this flaw, I postulate a hidden variable *domestic consumption preference* for each country and year, and tune it such that the calculated import reuse becomes scale-invariant, thereby fixing the aforementioned flaw. Assuming that input-output tables stay self-similar when increasing their resolution beyond the resolution of the available data, this improved approach yields the true import reuse.

When applied to atomic data,⁶ all discussed approaches must yield the same result, as they must stay within the bounds given by maximum flow, which converge towards the true import reuse as the resolution is increased. In theory, any value within these bounds could be the true value and the slope of the curve shown in figure ?? could change multiple times in all conceivable directions when the resolution is further increased. However, in absence of evidence to the contrary, the simplest assumption should

⁶ A hypothetical input-output table with atomic resolution would allow to track the exact trade path of every single apple as well as every other good at its atomic quantity.

be considered the most probable, namely that the structure of the data remains roughly the same at smaller scales. In that case, the scale-invariant estimate is the best guess for the true import reuse.

The presented approach raises the important question whether the introduced hidden variable *domestic consumption preference* is purely technical or whether it has broader implications. Deriving and testing such implications could constitute worthwhile future research. But first, one should verify whether this proposed approach yields consistent results when applied to other, and in particular more fine-grained data. Also, one could dig deeper into the specific ways in which the input-output graph must be self-similar in order for the results to be robust. A related question is whether the assumption taken for the construction of the input-output tables (for example the proportionality assumption) causes systemic distortions. Lastly, it would also be nice to improve algorithmic elegance and computational efficiency.

References

- R. A. Auer, A. A. Levchenko, and P. Sauré. International inflation spillovers through input linkages. 2016.
- J. Edmonds and R. M. Karp. Theoretical improvements in algorithmic efficiency for network flow problems. *Journal of the ACM (JACM)*, 19(2):248–264, 1972.
- X. Gabaix. Power laws in economics: An introduction. *Journal of Economic Perspectives*, 30(1):185–206, February 2016. doi: 10.1257/jep.30.1.185. URL <http://www.aeaweb.org/articles?id=10.1257/jep.30.1.185>.
- D. Hummels, J. Ishii, and K.-M. Yi. The nature and growth of vertical specialization in world trade. *Journal of international Economics*, 54(1):75–96, 2001.
- H. L. Kee and H. Tang. Domestic value added in chinese exports: firm-level evidence. *Johns Hopkins University mimeo*, 2013.
- R. Koopman, Z. Wang, and S.-J. Wei. Tracing value-added and double counting in gross exports. Technical report, National Bureau of Economic Research, 2012.
- OECD. The oecd input-output database, sources and methods. Technical report, OECD, 2000. URL <http://www.oecd.org/dataoecd/48/43/2673344.pdf>.

- C. Song, S. Havlin, and H. A. Makse. Self-similarity of complex networks. *Nature*, 433(7024):392–395, 2005.
- M. Timmer, A. A. Erumban, R. Gouma, B. Los, U. Temurshoev, G. J. de Vries, I.-a. Arto, V. A. A. Genty, F. Neuwahl, J. Francois, et al. The world input-output database (wiod): contents, sources and methods. Technical report, Institute for International and Development Economics, 2012.