

Estimating granular product-level trade flows in global value chains using publicly available data

Topic: Input-Output Modelling: Trade and Global Value Chains policies (3)

Author: Oscar Lemmers

Policymakers need detailed insights into global value chains (GVCs), from critical product dependencies to tariff impacts, but current data fall short. Trade data do not link products to industries. Furthermore, they do not capture indirect relationships, such as when an industry depends on another industry that uses titanium. In contrast, multi-country input-output tables provide these links, but lack detailed product information. For example, a dependence on “products of the mining and quarrying industry” does not tell whether one is dependent on gas, crude oil, metal ores or something else.

This raises the central research question of this paper: how can one estimate product-level trade flows within global value chains in a way that is consistent with multi-country input-output structures while presenting detailed product information? This methodological paper develops and implements a method to estimate such granular product-level GVC trade flows using exclusively publicly available data. The dataset covers about 5,000 products at the Harmonised System 6-digit level.

The data underlying the paper stems from public sources only. First, the FIGARO multi-country input-output tables and the corresponding multi-country supply and use tables (Eurostat/Joint Research Centre), covering about 50 countries and 64 industries over 2010-2023. Second, detailed supply and use tables from the U.S. Bureau of Economic Analysis, containing about 400 industries and products, for 2007, 2012 and 2017. Third, the Balanced International Merchandise Statistics of the OECD, with all countries in the world and products at HS6 level, for 1995-2024. The current implementation in the paper focuses on 2017.

The methodology proceeds in three main steps. First, input-output analysis using FIGARO yields GVCs with limited product detail, consistent with the multi-country input-output table. Second, the U.S. supply and use tables are used to infer a detailed structural prior for other countries. Because production structures differ across countries, this prior is adjusted using the observed trade data. For example, in the prior, Norwegian agriculture would export tropical fruits. Using the trade data, this is filtered out. Third, the product detail of the previous step is further enlarged using the OECD trade data. This step yields granular product flows, for example that French metal manufacturers supply frames to German carmakers and screws to German furniture manufacturers.

Given the relevance of product detail in GVCs, several approaches have been proposed to fill this particular data gap. They differ in their methodology and data requirements. Some rely on qualitative mapping through interviews and desk research (Gereffi et al.), others approximate product detail without exploiting detailed production structures (Berthou et al.), while other work uses machine learning techniques to infer product networks (AIPNET, Fetzner et al.) or relies on commercial databases with firm information (Mau et al.).

The contribution of this paper is threefold. First, it provides a transparent method to estimate product-level trade flows that remain fully consistent with multi-country input-output tables. Second, it demonstrates that substantial product granularity (around 5,000 HS6 products) can be achieved without relying on proprietary data, making it accessible to researchers with limited resources. Third, the modular structure of the method allows researchers to substitute components of the methodology or data sources, making the framework adaptable to alternative IO tables, more granular trade data, or country-specific information.

The resulting dataset supports various applications, including the identification of risk-prone products in GVCs of specific industries, tracing critical raw materials, assessing tariff exposure along GVCs, developing early warning indicators based on shifts in trade flows, and analysing inflationary pressures transmitted through GVCs. Earlier versions of the data and methodology were applied to the first three applications. The dataset based on the current paper and the underlying R-scripts will be made publicly available.