Resources Futures
A Chatham House Report

Methodology for the Chatham House Resource Trade Database

Read more
www.chathamhouse.org/resourcesfutures
Chatham House Resource Trade Database

Structure and content

To improve understanding of global flows of resources, Chatham House has developed a database tracking bilateral trade in natural resources and resource products between more than 200 countries and territories between 1998 and 2010. The database covers the weight and value of trade in over 1200 different types of natural resources and resource products—including agricultural, fishery and forestry products, fossil fuels, metals and other minerals. It allows for a detailed examination of new and growing resource-related dependencies among countries and regions and flows of resources through global value chains.

The principal sources for the Chatham House database are the International Merchandise Trade Statistics (IMTS), which are collected by national customs authorities and available through the United Nations Commodity Trade Statistics Database (UN COMTRADE). Specifically, we rely on the International Trade Database at the Product Level (BACI), a particular version of COMTRADE statistics that has been developed by the French Research Centre on International Economics (CEPII). BACI reconciles the raw data on trade flows reported separately by importers and exporters in COMTRADE. This method relies principally on adjustments for transport costs and on assigning a score to countries that is estimating the reliability of the data that they report. This score is then used as weight when averaging COMTRADE data obtained from the importing and exporting reporter into a single data point in BACI.

The Chatham House database, like COMTRADE and BACI, is organized according to the Convention on the Harmonized Commodity Description and Coding System (or HS nomenclature), which assigns 6-digit codes to different types of traded products in a hierarchical structure. The first two digits (HS-2) identify the chapter the goods are classified in, e.g. 26 = Ores, slag and ash. The next two digits (HS-4) identify groupings within that chapter, e.g. 26.01 = Iron ores and concentrates, roasted iron pyrites. The next two digits (HS-6) are even more specific, e.g. 26.01.11 Iron ore, concentrate, not iron pyrites, unagglomerated. HS nomenclature has evolved historically as a pragmatic taxonomy of the broad range of products that are traded globally. Products that have a long history of being traded extensively are captured in much greater detail than products that are traded less frequently. For example there is a single HS code associated to rare earths, while there are several hundred codes associated to steel and steel products.

Like in the case of other international merchandise trade statistics, an individual entry into the Chatham House database consists of the aggregate value and weight flowing from one country to another in an individual 6-digit code over the

---

1 This is compiled by the United Nations Statistics Division (UNSD). See http://comtrade.un.org/.
2 See www.cepii.fr/anglaisgraph/bdd/baci.htm.
4 The HS nomenclature is revised in regular intervals to account for the changing nature of global trade. The database relies on the 1996 version of HS codes, which is the revision used by the BACI database.
period of a year. Data are reported in US dollars and metric tons. The aggregate value and weight of crude petroleum oils (HS96 code 270900) exported from Saudi Arabia to China in 2003 (14.59 million tonnes worth 2.87 billion US dollars) would for example constitute one such data point. Any crude petroleum that would be exported back from China to Saudi Arabia in the same year would be captured in a separate data point. In total, the Chatham House database contains over 12 million data points for the years 1998 to 2010.

IMT statistics and HS nomenclature include not only natural resources but all types of traded goods, including manufactures. In order to track resource trade flows, only HS codes identifiable as raw materials or relatively undifferentiated intermediate products were selected into the database. For example both wheat and wheat flours were included, while bakery wares were excluded. Traded waste products, such as metal scrap or wood pulp are also included. It should be noted that natural resources typically go through incremental stages of processing and value addition and HS nomenclature does not always capture these different stages with much granularity. As result, not all HS codes can be classified unambiguously as ‘natural resources’ or be allocated to specific types of resources. 1253 HS codes were selected in this way out of the over 5000 HS codes available in BACI.

These 1253 HS codes were then re-organized to allow for the tracking of specific resources along the value chain. This is necessary because the common HS hierarchy is mainly organized around different types of products, rather than the materials that they consist of. For example, in COMTRADE ores and concentrates of various metals are classified in the chapter with the two-digit HS code 26, while refined metals and metal products are part of a different set of chapter (two-digit HS codes 72-83). In contrast to this, the Chatham House database groups HS codes by different types of resources. Copper ores and concentrates, various intermediate copper products (such as mattes, bars or wires) and copper scrap are for example all classified into a single ‘copper’ category. This allows tracking global copper trade (or other individual types of resources) at different stages in the value chain. These individual resources are then further organized into groups such as non-ferrous metals, cereals, or oilseeds.

Data quality and identification of outliers

Given well-known issues with data quality in IMTS statistics and the size of the dataset, the identification and treatment of outliers is a central challenge for the reliable tracking global resource flows. The reconciliation of exporter and importer reporting in BACI provides a significant improvement, but it does not provide an assessment of the reliability of entire database or of individual data points.

To address this issue, Chatham House has developed a measure of data quality for IMT statistics that works through on comparing value-weight ratios (or prices per ton) of individual data points. This method is based on the assumption that in a given year, resources belonging to the same six-digit product group can be expected to trade at roughly similar prices per ton, irrespective of the country pair that trades them.

In a first step a ‘distance indicator’ is calculated for each data point in the Chatham House database. This measure compares the price per ton of the individual data point with the median price per ton at which the product with the same HS code is traded among all countries in the same year (the median is used instead of the mean because it is less sensitive to large outliers). The measure is also normalized so that same deviations above or below the median price result in the equal value for the distance indicator (i.e. two data points that respectively trade at a third of the median world market price and at triple the median world market price will have the same distance value). Mathematically this can be expressed as:

5 The full classification of HS codes developed for the database is available from the authors on request.

6 See UN Comtrade (no date), ‘Read Me First (Disclaimer): Every User of UN Comtrade should know the coverage and limitations of the data’. Available: http://comtrade.un.org/db/help/uReadMeFirst.aspx.
The resulting measure can be interpreted as the differential between the price-per-ton at which a country pair is trading a given product in a given year, and the median world market price-per-ton for the same product in the same year. A value of 5 for an individual data point, for example, indicates that the price at which the country pair trades the product is five times larger (or five times smaller) than the median price at which the same product was traded by all country pairs in the same year.

Significant price differentials in COMTRADE and BACI data are to be expected. For example, resources may experience large price swings over the period of one year and global trade for individual commodities is often fragmented into regional markets, where the same commodity can trade at a substantial price premium or discount. Other factors, such as variations in transport costs, differences in product quality that are not captured in HS codes, different seasonal trade patterns, quantity discounts for larger orders, differences in bargaining power between trading partners, or long-term supply contracts may also contribute to such differentials. The natural gas and LNG market is a good example, where a million British thermal units (Btu) of gas in Asia in 2010 cost approximately 10.91 US dollars, while selling at 8.01 dollars in Europe and just 4.39 in the US.\(^7\)

However, opportunities for arbitrage create pressures limiting the size of such differentials. The larger the trade flow and the greater the distance is from the median world market price the stronger these pressures will become. It conceivable that two countries could trade a good in a given year at a premium of several hundred per cent compared to the median world market price (in the gas case the mark-up for Asian LNG over US gas prices was nearly 150 per cent). A premium of several thousand per cent is much less plausible. Where calculated distances for individual data points in IMT statistics are very large, data quality issues can therefore be suspected.

Figure 1 shows the distribution of data points in the Chatham House resource trade database by the size of their distance indicator. As expected, the vast majority of data points cluster tightly around the world market median price: 7.7 million data points trade at more than half and less than double the median world market price. Another 2.8 million points trade at more than a fifth and less than five times the median price. However roughly one in ten observations is off from median world market prices by much larger margins, with prices per ton hundred or even thousand times higher than median world market prices. These large differentials are likely to result from inaccurate statistics rather than reflect actual trading patterns.

To improve the reliability of the database, data points where prices per ton are less than a tenth or more than ten times the world market price are classified as outliers and excluded from the dataset. This concerns 1.33 million data points. Another 137.8 thousand data points are excluded because weight data are missing. Together, these constitute roughly 12% of the total observations. While excluding such a large number of data points could appear problematic, Figure 2 shows that they consist overwhelmingly of relatively insignificant observations: together, the excluded data points account for less than 3% of the total value of global resource trade flows that are recorded in the database. This is not surprising as the likelihood that trade flows are recorded wrongly and not corrected subsequently in COMTRADE is much higher for very small and insignificant flows than it is for large and valuable trade flows.

Figure 3 demonstrates that over 95% (1.40 of the 1.47 million) of the observations that were excluded based on the distance measure concern trade flows worth less than one million per year. In contrast, there are only 119 excluded trade flows with a value that is larger than one billion. Together, these 119 data points contain nearly 40% of the total value that was excluded as outliers.
In order to improve the accuracy of the data and to avoid excluding resource trade streams of significant size, these 119 data points were corrected individually and re-included in the database. The corrections were made based on alternative sources such as the original COMTRADE data (as opposed to the BACI version of the data), FAO Trade Stat, Eurostat, and the yearbooks of national customs and statistics authorities. Ultimately, only little more than 1 per cent of the total value of global trade flows is excluded from the database.