UNITED NATIONS CONFERENCE ON TRADE AND DEVELOPMENT

MONTH 2022

UNCTAD Research Paper No. 85
UNCTAD/SER.RP/2022/5

Onno Hoffmeister
UNCTAD/DSIB

Ronald Halim
Equitable Maritime Consulting

Nicholas-Joseph Lazarou
Joint Research Centre

Nour Barnat
UNCTAD/DSIB

David Cristallo
UNCTAD/DSIB

Dominik Englert
World Bank

Jan Hoffmann
UNCTAD/TLB

UNITED NATIONS

Developing a global transport costs dataset for international trade

Abstract

This paper describes the sources and methods used for the compilation of the new Global Transport Costs Dataset on International Trade (GTCDIT), a beta version of which is publicly available on UNCTADstat. GTCDIT records bilateral international merchandise trade in value and quantity, broken down by commodity group and mode of transport (air, sea, railway, road, other modes), alongside its associated transport costs, for 2016. The compilation of GTCDIT has been made possible by the availability of new variables in a recent upgrade of the UN Comtrade database and of new estimates on global transport distances derived with the help of geographic information systems. To obtain global coverage, the primary data on the new variables in UN Comtrade reported by some countries have been used to develop models that estimate the missing values of most other countries. As a result, GTCDIT covers around 87 per cent of global trade in terms of value.

Key words: International trade, freight transport, data editing, imputation

Contents

Acknowledgements

The authors gratefully acknowledge the scientific advice from Wendela Schim van der Loeff (University College of London), the work of Denis Gervalle (UNCTAD/DSIB), Yoann Chain (UNCTAD/DSIB) and Richard Chalveras (UNCTAD/DSIB) for the acquisition of the UN Comtrade data and the dissemination of the output data, the assistance in project coordination from Jennifer Brown (World Bank), Goran Dominioni (World Bank) and Ekaterina Chernova (UNCTAD/DSIB), the comments from Daniel Hopp (UNCTAD/DSIB), the managerial support from Stephen MacFeely (UNCTAD/DSIB) and Anu Peltola (UNCTAD/DSIB), and the financial support from the World Bank's Blue Economy Program (PROBLUE).

1. Introduction

The United Nations Conference for Trade and Development (UNCTAD) and the World Bank have joined forces to develop a Global Transport Costs Dataset for International Trade (GTCDIT). GTCDIT is aimed to represent a novel information source recording the transport costs and transport work for the shipment of commodities from exporting countries to importing countries, alongside the corresponding bilateral trade, in volumes and value, and an estimate of the distance over which the goods can be assumed to have travelled. That information is made available at the level of individual commodity groups and differentiated by mode of transport (MoT).

International transport is central to global trade and the economic development of countries. Transport costs, however, constitute an important impediment of trade. The costs of transporting commodities are perceived to be comparable in size to tariffs, exhibiting large variation across products, jointly altering the patterns of and gains from trade (Hummels, 2001, 2007; Hummels et al., 2009). As tariffs have been gradually declining across the last 50 years (Yi, 2003), the relative importance of transport costs as a trade barrier has been rising. Transport costs have been shown to impact economic development by a number of studies. These show that higher transport costs are most likely attributed to geographic disadvantage, in particular peripheral location (Krugman, 1980). Exporting countries frequently have to absorb transport costs so as to access foreign markets impeding export-led development, lowering returns to labour and welfare (Yeatz et al., 1996). Remoteness and lack of infrastructure inhibits market access for developing nations leading to losses from trade, weakened competitiveness of domestic products in international markets, and a high import bill (UNCTAD, 2022). Redding and Venables (2004) estimate that these losses account for around 68 per cent of gross domestic product (GDP) per capita on average. Often, small island developing States (SIDS), landlocked developing countries (LLDCs), and least developed countries (LDCs) spend more than the average country on the international transport and insurance of their merchandize imports (UNCTAD, 2017, 2021a).

Yet, due to a paucity of data, the various impacts of transport costs on international trade are not well known and difficult to quantify. This applies also to their determinants (Sanchez et al., 2003; Anderson and Van Wincoop, 2004; Korinek and Sourdin, 2009; Korinek, 2011). Data on transport costs collected at the transaction level often differ according to the type of contract the shipper has signed. For example, the costs of shipments can be expressed as dollars per ton or dollars per day travelled (Stopford 2008). Aggregating those figures to the country pair level with full coverage within a given time period becomes a formidable if not infeasible task. The extant international trade literature usually uses distance as the most suitable and widely available proxy for transport costs (Berthelon and Freund, 2007). While distance may explain a significant proportion of variations in the costs of transport, many other determinants exist.

One solution consists in measuring transport costs as the difference between the cost, insurance and freight (CIF) and the free on board (FOB) price of a good shipped from one location to another. This standardized way of reporting enables easy aggregation across goods, reporting countries and trading partners. Hummels (2001) and the United Nations Economics Commission for Latin America and the Caribbean (UN-ECLAC) (Hoffmann et al., 2002) carried out pioneering work in constructing transport costs statistics from that type of data. Hummels compiles CIF-FOB differentials based on detailed data from the national offices of the United States of America, New Zealand, Argentina, Brazil, Chile, Paraguay and Uruguay. The statistics from UN-ECLAC are derived from a database constructed from customs records on imports and exports from eleven Latin America countries.

In an attempt to obtain global coverage Gaullier and Zignago (2010) derive transport costs from the difference between the CIF value reported by the importing country and the FOB value reported by the exporting country for the same flow in UN Comtrade, the world's largest database of bilateral international trade broken down by product groups. UNCTAD (2017) applies the same approach to bilateral trade data from the IMF's Direction of Trade Statistics, to assess broad trends and patterns in transport costs in developing and developed countries, SIDS and LLDCs. The same method has been applied, though not publicly documented in detail, for the development of the World Input Output Database where a CIF-FOB adjustment is needed for the conversion from basic to purchasers prices (Streicher and Stehrer, 2013). However, this "implicit" (Miao and Fortanier, 2017) approach to the calculation of transport costs is strongly complicated by the fact that differences between reported CIF values and the mirrored FOB values are not caused by transport costs alone. They are to a large extent a reflection of cross-country discrepancies in the measurement and recording of international trade. Gaullier and Zignago use econometric modelling to adjust for this disturbance, based on the assumption that average trade asymmetries converge to the expected value of the CIF-FOB differential as the number of observations increases. This implies that the asymmetries caused by discrepancies in the measurement and recording of trade converges towards zero, an assumption which may not necessarily hold.

The most recent attempt to compile transport costs based on CIF-FOB differentials was the development of the International Transport and Insurance Costs of Merchandise Trade dataset by the Organization for Economic Cooperation and Development in 2016. Like the Hummels and the ECLAC datasets, ITIC is based on "explicit" CIF-FOB differentials observed on the importers side. The source data were obtained from the national statistical offices of eight developed and eight Latin American countries. To achieve global coverage, the data for the remaining countries were estimated using econometric modelling (Miao and Fortanier, 2017).

GTCDIT follows, in principle, the approach of calculating transport costs "explicitly" as the difference between the CIF and the FOB value reported on the importers side, benefiting however from the availability of new variables in a recent upgrade of the UN Comtrade database, known as UN Comtrade Plus (UNSD, 2021a), as a result of a change in official reporting guidelines for international merchandise trade statistics in 2010 (UNSD, 2011). These revised guidelines encourage countries to distinguish between MoTs, when reporting their bilateral trade figures, and to record both the CIF and FOB value for imports. This new information enables us to construct explicit CIF-FOB margins for the same flow, differentiated by MoT. Furthermore, advancements in geographic information systems (GIS) have enabled the identification of the shortest distance needed to ship goods between the main city centres of the countries of the world, depending on the MoT used. Thus, GTCDIT can make available data on transport costs for each flow of bilateral trade between countries worldwide, alongside the information on the value and volume of the corresponding trade, differentiating by around 5000 commodity groups and five main MoTs: air, sea, railway, road, and other (comprising non-standard modes such as pipelines, powerlines, post, etc., as well as inland waterways). The dataset also provides a summary measure of the mode-specific distance between the exporting and importing countries as well as derived indicators, such as the ad-valorem freight rate (the ratio of transport costs to the FOB value), unit transport costs, and unit transport costs per km.

Transforming the wealth of new information from UN Comtrade Plus and GIS sources into an integrated information product that accurately records bilateral trade and the associated transport service, allocated in correct proportions over MoTs and commodity groups, is a challenging endeavour which, to the best of our knowledge, has not been carried out before. The beta version of the database, publicly available on UNCTADstat since December 2020 (UNCTAD, 2021b), represents the result achieved after the first year of work. This first version is already, to our knowledge, the most complete mode-specific dataset of international transport and trade available to date. However, research is still ongoing and several measurement issues, discovered during the first project year, have not yet been entirely solved. These imperfections will be discussed in more detail at the end of section 2.1.

Other compilation challenges have been inherited from the source data. At the time the project was carried out, data coverage in the new variables used from UN Comtrade Plus was relatively low. The data gaps have been filled by econometric models, designed to learn from the reported data to make predictions about the non-reported values. Furthermore, international trade data are known to be prone to errors. To enhance accuracy, algorithms have been developed that comb the several million records of the dataset for apparent errors and carefully correct these. As a result of the low coverage with source data and the extensive imputations applied, most of the data included in the present version of GTCDIT, especially at individual MoT level, are statistical estimates. In its current form, GTCDIT should therefore be considered primarily as a synthetic dataset. Estimated values are distinguished from originally reported values by flags, so that users can consider that information about the data origin in their analysis.

Currently, the dataset utilises information from 136 importing countries recorded in UN Comtrade Plus. The output data report CIF values, FOB values, MoT breakdowns and variables derived thereof in 6.8 million records, covering about 87 per cent of global trade, in terms of CIF value, for 2016. As data coverage in UN Comtrade Plus increases in the coming years, and more research can be carried out, UNCTAD intends to successively enhance GTCDIT in accuracy and scope, mainly by developing solutions to persisting measurement challenges, refining the statistical methods used for data editing and filling of gaps, and by sourcing new and refined data from UN Comtrade Plus. The time coverage is intended to be extended further by adding data, as a first step, for 2017 and subsequent years. Once finalized, GTCDIT is envisaged to become a key resource at the disposal of researchers, policy makers, enterprises, non-profit organizations and the interested public.

The aim of the present document is to provide users of GTCDIT with a solid understanding of the information contained in it, of the origins of that information, and of the way in which it has been compiled. The remainder of this paper is structured as follows. Section 2 provides a description of the GTCDIT variables and the structure of the database. It explains how the data can be accessed and viewed, and it presents main patterns in global transport costs revealed by them. Section 3 presents the data sources used for the compilation of GTCDIT. Sections 4 to 7 describe the methods applied to clean the data from suspect cases and to fill data gaps. Section 8 concludes the paper.

2. Database description and structure

GTCDIT records the value and volume of bilateral international merchandise trade, measured from the import side, alongside the transport costs it incurred during the year 2016, where trade and transport costs are broken down by commodity group and MoT. The dataset also provides derived variables, such as the ad-valorem freight rate (the ratio of transport costs to the FOB value), unit transport costs and unit transport costs per 10 000 km. The dataset covers the imports of 136 countries received from 235 exporting countries, broken down into 5 204 commodity groups, differentiating between five MoTs. It covers 92 per cent of global imports in 2016, in terms of (CIF) value. Transport costs and MoT breakdowns are available for data representing 87 per cent of global trade.

2.1. Definition of variables

The specific variables included in GTCDIT are presented in table 1, where we distinguish between *dimensions*, variables which identify observation units, and *facts*, variables which record observations on those units. The table also provides descriptions of what precisely these variables indicate or measure, and, for dimensions, the categories by which they differentiate or, for facts, the units of measure in which they are expressed.

To avoid misinterpretation, some particularities in the variable definitions require attention:

a) The MoT recorded under *TransportMode* may not be the only mode used during the transport from the country of origin to the destination country. Before reaching the final destination by the recorded MoT, goods may have been transported by other modes and transloaded on their way from the origin to the destination countries. In the present version of GTCDIT, the resulting inaccuracy in the breakdown of the variables of the dataset by MoT could only partially be adjusted for: in cases in which an MoT appears impossible to be the only means used – as for instance "sea" for the transport from or to a landlocked economy, or "railway" or "road" for the transport from or to an island state – the trade recorded for that MoT has been allocated to the other modes based on econometric predictions. For further details on how these adjustments have been made, see sections 6 and 7 below.

b) Since transport costs are derived as the difference between the CIF and the FOB value, they cover more than the pure costs of the service of moving the imported goods geographically. They also include the costs of insurance for losses and damage. Under free competition, these insurance costs can be taken as a measure of the average expected costs caused by those losses and damage. As insurance costs are part of the trade bill, it is reasonable to take them into account as a component of transport costs.

c) The transport costs recorded in the database are meant to refer to the transport up to the border of the destination country, whereas distance is measured up to city centres located within that country. Therefore, transport costs per unit per 10 000 km may be slightly underreported.

d) As the distance between the origin and the destination country is calculated as an average of the minimum distance needed to transport goods between main city centres of these countries, it should be considered as a first best estimate of the average distance which goods actually travelled. This average does not take into account the intra-country distribution of trade, which is not currently available. It should also be considered that, for logistic reasons, the direct route is not always the most cost-efficient one. Carriers may decide to let goods travel over a longer distance than technically needed.

e) As trade in GTCDIT is recorded from the importer side, a country's imports are principally exhaustively covered. However, the completeness of the recording of its exports depends on the completeness by which its trading partners are represented in the source dataset.

It should also be noted that the dataset contains values reported by countries as well as statistical estimates. The statistical estimates are distinguished from the reported values by the flag "(5)" (or data status "50" in the bulk-download file).

2.2. Viewing and retrieving the data

The beta version of GTCDIT is openly available in a dedicated domain in the UNCTADstat Data Centre, at the URL below:

https://unctadstat.unctad.org/wds/ReportFolders/reportFolders.aspx

Table 1. The variables of the dataset

a) Dimensions

CommodityProduct	
Description	The traded commodity
Categories	Commodity groups as defined by the sub-headings (6-digit level) of the Harmonized Commodity Description and Coding System, version 2012 (World Customs Organization, 2022)
Origin	
Description	The country of origin as reported by the importer. According to IMTS Concepts and Definitions (UNSD, 2011), this means the country in which the goods have been produced or manufactured, in accordance with the Revised Kyoto Convention, Specific Annex K / Chapter 1/ E1.
Categories	Economies, as defined by the UNCTAD classification of economies (UNCTAD, 2021c)
Destination	
Description	The country or territory that has reported an import in its international merchandise trade statistics submitted to the United Nations Statistics Division (UNSD).
Categories	Economies, as defined by the UNCTAD classification of economies (UNCTAD, 2021c)
TransportMode	
Description	The means by which imported goods enter the economic territory of the importing country, in accordance with IMTS Concepts and Definitions, article 7.1 (UNSD, 2011)
Categories	10 - Air 21 - Sea 31 - Railway 32 - Road 99 - Non-standard modes

b) Facts

CIF value	
Description	The CIF-type value of imports, as defined by IMTS Concepts and Definitions, article 4.6 (UNSD, 2011). It includes the transaction value of the goods, the value of services performed to deliver goods from their origin to the border of the exporting country and further to the border of the importing country. These services should include not only the service of bringing the goods to their destination, but also the procurement of insurance against the risk of loss or damage during the carriage, in accordance with the definitions in Incoterms (International Chamber of Commerce, 2019). The CIF value is usually derived by customs administrations from the invoice price and the terms of delivery indicated in the contract of sale, as well as from other supporting documents.
Unit of measure	United States dollars
FOB value	
Description	The FOB-type value of imports, as defined by IMTS Concepts and Definitions, article 4.6 (UNSD, 2011). It includes the transaction value of the goods and the value of services performed to deliver the goods from their origin to the border of the exporting country. It is usually derived by customs administrations from the invoice price and the terms of delivery indicated in the contract of sale, as well as from other supporting documents.
Unit of measure	United States dollars
Quantity	
Description	The quantity imported goods. Quantities expressed as weights usually refer to the net weight, thus excluding packaging.
Unit of measure	Various, depending on CommodityProduct (*e.g.* kilogram, carat, meters, square-meters, litres, 1 000 kilowatt hours, number of pieces). The unit of measure is provided at the end of the label of the *CommodityProduct* category.
Distance	
Description	The average distance over which goods need to be transported for their delivery from main city centres of the origin to main city centres of the destination country.
Unit of measure	Kilometers
Transport costs	
Description	The difference between *CIF value* and *FOB value,* as defined above.
Unit of measure	United States dollars
Transport costs to FOB value	
Description	The ratio of *transport costs* to *FOB value*
Unit of measure	Per cent
Transport costs per unit	
Unit of measure	Various, depending on *CommodityProduct,* in accordance with the unit of measure of *quantity* (see above), *e.g.* US$/kg, US$/carat, US$/meter, US$/m^2, US$/l, US$/10 000 kwh, US$ per piece.
Description	The ratio of *transport costs* to *quantity*
Transport costs per unit per 10 000 km	
Unit of measure	Various, depending on *CommodityProduct,* in accordance with the unit of measure of *quantity* (see above), *e.g.* US$ per 10 ton-km, US$ per 10 000 carat-km, US$ per 100 hectolitre, US$ per 10 0000 item-km.
Description	The ratio of transport costs to the product of quantity and distance. The product of quantity and distance is a measure of the volume of used transport service.

For technical reasons, the data have been split into twelve dataset tables on UNCTADstat, two for each MoT (air, maritime, railway, road, other) as identified by the measure *TransportMode,* as well as for the total of all modes. Six of the dataset tables contain 'input variables', namely the CIF value, the FOB value, and quantity; six other datasets contain 'output variables', namely transport costs, transport costs per unit, and transport costs per unit per km.

Each of these tables provide the usual facilities for browsing UNCTADstat data, including sorting, pivoting, filtering, searching and downloading extracts in a pre-specified format. A typical extract from the dataset looks like that of figure 1.

Figure 1. Typical extract from the GTCDIT, transport costs values

Sea Mode - Transport cost, transport cost per unit, and unit transport cost per km

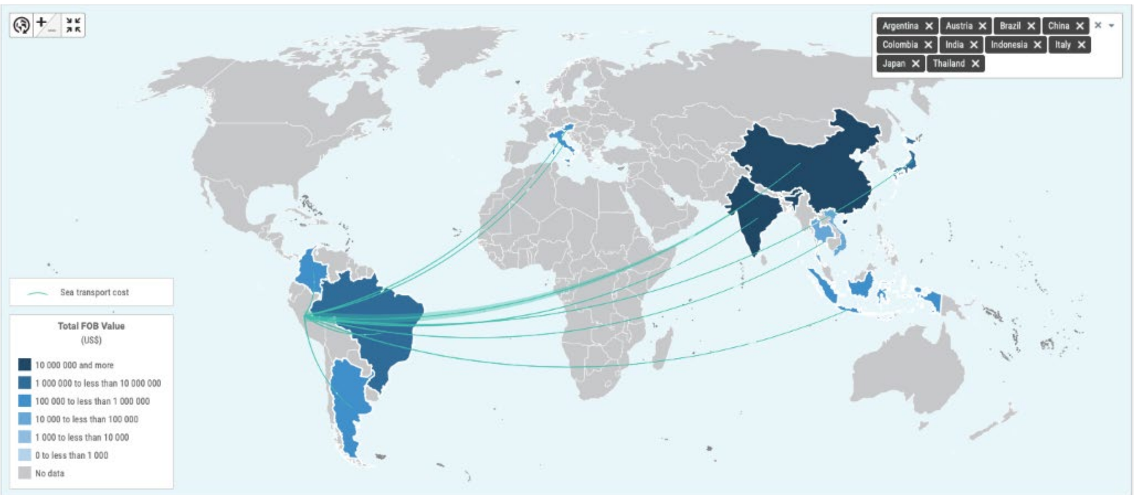
Other:

MEASURE	Transport/Insurance costs (US dollars)								
ORIGIN	Belgium	Canada	China	France	Germany	India	Japan	Mexico	United Kingdom
DESTINATION	United States of America	United States of America	United States of America	United States of America	United States of America	United States of America	United States of America	United States of America	United States of America
COMMODITYCODE									
Spirits obtained by distilling grape wine or grape marc - liters	(5) 34 760	(5) 455	(5) 9 875	(5) 75 189 865	(5) 27 699	..	..	(5) 58 368	(5) 163 487
Wine of fresh grapes, incl. fortified wines, and grape must whose fermentation has been arrested by the addition of alcohol, in containers of <= 2 l (excl. sparkling wine) - liters	(5) 186 170	(5) 329 147	(5) 57 082	(5) 55 215 353	(5) 5 934 365	(5) 11 751	(5) 6 999	(5) 71 973	(5) 486 114
Light oils and preparations, of petroleum or bituminous minerals which >= 90% by volume """"incl. losses"""" distil at 210°C """"ASTM D 86 method"""" (excl. containing biodiesel) - kilograms	(5) 53 022 407	..	(5) 4 557 456	(5) 51 103 359	(5) 4 334 744	(5) 142 501 767	(5) 4 843 982	..	(5) 232 245 760
Parts of turbojets or turbopropellers, n.e.s. - kilograms	(5) 4 054 218	(5) 4 364 231	(5) 3 956 121	(5) 39 455 734	(5) 10 835 927	(5) 591 722	(5) 20 487 494	(5) 776 498	(5) 10 034 747
Vodka - liters	(5) 48 514	(5) 469 990	(5) 20 236	(5) 37 703 284	(5) 68 401	(5) 9 507	(5) 563 739	(5) 17 262	(5) 2 651 403
Sparkling wine of fresh grapes - liters	(5) 15 594	(5) 1 193	..	(5) 36 349 175	(5) 155 581	(5) 2 710	..	(5) 93	(5) 116 645
Paintings, e.g. oil paintings, watercolours and pastels, and drawings executed entirely by hand (excl. technical drawings and the like of heading 4906, and hand-painted or hand-decorated manufactured articles) - number of items	(5) 3 903 677	(5) 28 892	(5) 2 495 436	(5) 30 200 165	(5) 10 931 255	(5) 1 047 971	(5) 2 271 857	(5) 1 014 866	(5) 15 209 385
Medium oils and preparations, of petroleum or bituminous minerals, not containing biodiesel, n.e.s. - kilograms	(5) 70 136 725	..	(5) 36 520 315	(5) 29 384 832	(5) 9 805 132	(5) 81 942 658	(5) 34 850 545	..	(5) 15 711 028
Perfumes and toilet waters (excl. aftershave lotions, personal deodorants and hair lotions) - kilograms	(5) 15 297	(5) 19	(5) 838 436	(5) 15 935 903	(5) 834 783	(5) 386 946	(5) 24 096	(5) 3	(5) 2 806 439
Liqueurs and cordials - liters	(5) 6 503	(5) 6 885 596	(5) 57 625	(5) 12 254 473	(5) 7 274 782	..	(5) 74 562	(5) 1 380 980	(5) 1 362 258
Uranium enriched in U 235 and its compounds: plutonium and its compounds; alloys, dispersions, incl. cermets, ceramic products and mixtures containing uranium enriched in U 235, plutonium or compounds of these products [Euratom] - kilograms	(5) 270	(5) 152	(5) 8 411 881	(5) 12 052 947	(5) 32 038 104	..	(5) 1 977	..	(5) 28 479 500
Casks, barrels, vats, tubs and other coopers' products parts thereof, of wood, incl. staves - kilograms	(5) 7 091	(5) 12	(5) 58 637	(5) 11 686 303	(5) 7 141	(5) 261	(5) 5 432	..	(5) 21 809
Turbojets of a thrust > 25 kN - number of items	(5) 6 747	(5) 1 990 100	(5) 289 706	(5) 11 531 507	(5) 11 056 177	(5) 350 163	(5) 268 391	(5) 26 666	(5) 7 624 261
Aluminous cement - kilograms	..	(5) 245 532	(5) 465 002	(5) 10 330 520	(5) 8 853	..	(5) 22 360	..	(5) 306 980

Source: UNCTADstat (UNCTAD, 2021b)

An interactive map is also provided, connected with the data, that enables users to visualize the flow of trade in individual groups of commodities and the related transport costs between country pairs (see figure 2). Users have at their disposal the option to choose a particular importer and product and observe the origin and value of the shipment in FOB terms.

Figure 2. Extract from the Global Transport Costs Database for International Trade displaying motorcycle transport costs to Peru



Source: UNCTADstat (UNCTAD, 2021b)

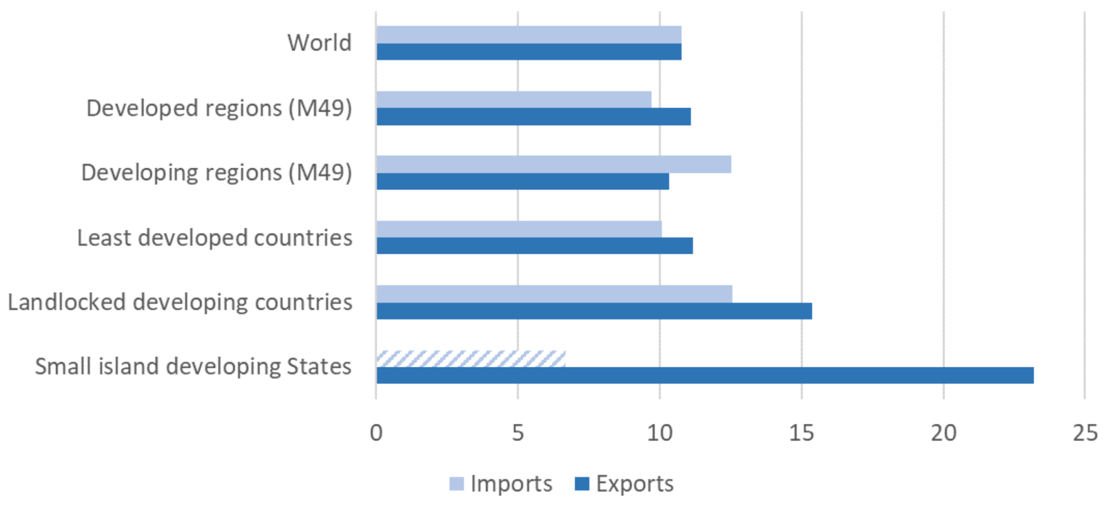
Like all dataset table released in the UNCTADstat Data Center, GTCDIT is also available as a compressed flat file in 7z-format (a free open-source format with a high compression ratio) in the bulk-download section:

https://unctadstat.unctad.org/EN/BulkDownload.html

2.3. Global patterns revealed by the data

Figure 3 shows the average ad-valorem freight rate, *i.e.* the ratio of transport costs to FOB value, for the world total and for selected country groups as revealed by the GTCDIT data. The global rate averages to 10.8 per cent. It thus lays in the upper range of the ad-valorem freight rate compiled at the country level by Hummels (2001) and Miao and Fortanier (2017). Comparing the figures from GTCDIT across regions, we note that developing and developed countries pay on average a similar price for the transport of their exports and imports, ranging from slightly below 10 to just above 11 per cent. The rate for LDCs also falls within that range. However, LLDCs are faced with transport costs per FOB value of their imports of 12.6 per cent, almost 3 percentage points higher than developed countries. They pay a price of 15.4 per cent for their exports, which is 4.3 percentage points more than developed countries do. SIDS, finally, appear to be faced with exceptionally high ad-valorem freight rates on the export side, amounting to 23.2 per cent on average. On the import side, however, based on information from four SIDS, [1] they are estimated to pay only 6.7 per cent, much less than the world average. This discrepancy between exports and imports may be due to different product compositions, different distances to and connectedness with main trading partners and/or due to the proportions in which transport is allocated over the modes air and sea, between exports and imports. GTCDIT allows exploring the factors behind these findings in more detail.

Figure 3. Average transport costs per FOB value in selected groups of countries
(Percentage)



Note: The figure for imports of SIDS is based on information from only four out of 56 SIDS: the Dominican Republic, Jamaica, Mauritius and Singapore. It therefore has limited significance.

[1] On the import side, freight rates can be calculated only for countries that reported their imports to UN Comtrade. As these include only four SIDS, these figures may not be representative for the SIDS of the whole. On the export side, by contrast, the calculated freight rates are based on mirror flows reported by 136 countries vis-à-vis principally all SIDS. If we restrict the calculation on the export side to the four SIDS for which freight rates on imports are available, we obtain a very similar rate as for the SIDS as a whole: 24.4 per cent (as compared to 23.2 per cent). This confirms the finding of a high difference in freight rates between exports and imports at least for these four SIDS.

Table 2 illustrates how transport costs of international trade, per FOB value and per quantity, vary across commodity groups. Raw materials are usually heavier relative to monetary value than manufactured goods, which generally makes them more costly to transport. This is well indicated by the difference in the average ad-valorem freight rate between cereals, iron and steel, on one hand, and pharmaceutical products, electric motors, passenger cars, and calculating machines on the other. While transport costs for the former two product groups amount to 16.3 and 19.1 per cent of the FOB value, respectively, the costs for the latter groups range between 4.6 and 7.0 per cent. International trade in wool and cotton, representing relatively light raw materials, is faced with a lower ad-valorem freight rate than trade in cereals and in iron and steel, but still with a higher rate than the aforementioned types of manufactured goods. Such differences between raw materials and manufactured goods may explain part of the discrepancies in ad-valorem transport costs paid by LLDCs and SIDS for their imports and exports, presented above, considering that the export of these groups of countries are dominated by raw materials, while their imports are largely comprised of manufactured goods. The SIDS' ocean location likely allows them to save costs by realizing their imports over sea, in sharp contrast to LLDCs, which need to rely more on costly overland transport. For the delivery of SIDS' exports, largely comprised of perishable goods, such as fish and other food products, transport by air may often be required, which is a relatively expensive MoT.

Table 2. Global transport costs in trade with selected commodity groups

Commodity group	HS code	Transport costs to FOB value *(Percentage)*	Unit transport costs *(US$ per unit)*
Cereals *(tons)*	10	16.3	15.38
Wool and Cotton *(tons)*	51, 52	8.8	76.64
Iron and steel *(tons)*	72	19.1	4.86
Pharmaceutical products *(tons)*	30	7.0	36.75
Electric motors *(number of items)*	8501	4.7	0.36
Motor cars for passengers *(number of items)*	8703	5.2	567.32
Calculating machines *(number of items)*	8470	4.7	0.71

Transport costs per quantity unit also show a high heterogeneity across commodity groups, reflecting differences in product characteristics. International transport of iron and steel costs on average US$4.86 per ton. The transport of cereals is two times more costly, per ton, that of pharmaceuticals seven times, and that of wool and cotton 15 times more costly than the transport of iron and steel. For the shipment of an electric motor, traders pay on average 36 cents of transport costs; for the shipment of a calculating machine on average 71 cents; and for a passenger car an average of US$567, according to the new GTCDIT data.

How much do transport costs vary across transport modes? As table 3 shows, the highest ad-valorem freight rates are paid in trade realized by air (7.7%) and road (7.8%), and the lowest in trade by railway (2.3%). However, these rates do not take into account the different distances over which goods are transported. Railway and road are on average chosen for shorter distances than air and sea. This effect is adjusted for in the ad-valorem freight rates per 10 000 km, shown in the column on the right of the table. Based on that measure, sea turns out to be the most economic and road the most expensive transport medium. Certainly, these figures are affected by the different types of products transported by the different modes. For example, light products, with reference to their value, are preferably transported by air whereas heavy products are preferably transported by sea, to save costs. Thus, unit transport costs expressed for the same type of commodity, can be expected to show a much higher difference between sea and air than the ad-valorem freight rates for the product totals. Such types of analysis can be easily carried out, at a detailed product level and differentiated by trading partner, with GTCDIT.

Table 3. Global transport costs by mode

Mode of transport	Transport costs to FOB value *(Percentage)*	Transport costs to FOB value per km *(Percentage per 10 000 km)*
Air	7.7	7.1
Sea	5.4	5.4
Railway	2.3	6.3
Road	7.8	9.5
Others	6.2	/

3. Data sources

GTCDIT was built from two sources: the UN Comtrade Upgrade 2019, also referred to as "Comtrade Plus", and a distance matrix constructed from GIS.

3.1. UN Comtrade Plus

UN Comtrade is a well-known and widely recognized comprehensive database on international trade, constructed from international trade data reported by national statistical agencies, mainly based on the customs records presented at country borders, and made available in harmonized format by UNSD (UNSD, 2021b). Among other data, the database records the values and volumes of merchandise exports and imports, broken down by reporting country, trading partner and product group. The data obtained from countries are released in UN Comtrade almost in their raw form, after only basic editing and validation checks have been applied.

With the "Upgrade 2019", new variables have been added to UN Comtrade, opening up new opportunities for the measurement of transport costs detailed by mode. Firstly, while in the conventional UN Comtrade database each flow is recorded on a CIF or FOB basis only – imports usually on a CIF basis and exports on an FOB basis – for some reporting countries both values have now become available for imports (UNSD, 2021c). This combined reporting of CIF and FOB values has enabled us to calculate transport costs as the difference between the two (see section 2.1 above). Secondly, for some countries, trade flows are now broken down by the MoT by which the goods entered the destination country. That information has been used to allocate imports over the five MoT categories of the variable *TransportMode*: air, sea, railway, road, and others (see section 2.1 above).

From the wealth of information stored in UN Comtrade Plus, we have restricted ourselves to the data of annual merchandise trade in 2016 that are classified by commodity groups at a 6-digit level based on HS 2012, the latest revision of HS in which 2016 data are recorded. The following variables of the UN Comtrade Plus dataset have been used for the construction of GTCDIT:

- the flow type (flowCode) to distinguish between imports and exports;
- the code of the reporting country (reporterCode),
- the code of the partner country (partnerCode);
- the six-digit commodity code (cmdCode);
- the code that identifies the MoT (motCode);
- the code that identifies the quantity unit used for the measurement of volumes (qtyUnitCode);
- the CIF value of trade in United States dollars (CIFValue);

- the FOB value of trade in United States dollars (FOBValue);
- and the volume of trade (qty).

For the reference year 2016, combined CIF and FOB values could be found only for imports, not for exports, throughout UN Comtrade Plus. For that reason, and considering that goods are usually recorded more accurately when they arrive in a country than when they leave a country, GTCDIT uses the reported data on imports as the primary source of information, while data on exports are employed for filling data gaps and performing accuracy checks by way of mirroring.

The coding and definition of economies in UNCTADstat is similar to the coding and definition used in UN Comtrade Plus, as both are based on the M49 standard (UNSD, 2020). However, a few differences exist. Therefore, some UN Comtrade Plus codes for reporting country and partner country have been aggregated to obtain the corresponding UNCTADstat economy codes. Certain flows, for which the individual trading partner was not specified, had to be dropped. The list of the UNCTADstat codes of economies, and their correspondence with M49, can be found on the UNCTADstat website, in the tab "Documentation" (UNCTAD, 2021c).

While UN Comtrade Plus allows in principle distinguishing between more than ten MoTs, in GTCDIT the distinction is made by only five main modes, in order to limit complexity and size: air, sea, railway, road, and all others (labeled "non-standard mode"). The conversion of the MoT categories of UN Comtrade into the MoT categories for GTCDIT has been conducted in accordance with the correspondences presented in table 4.

Table 4. Correspondence between MoT categories in UN Comtrade Plus and GTCDIT

UN Comtrade Plus		GTCDIT	
Code	Label	Code	Label
0	All	0	All
10	Air	10	Air
21	Sea	21	Sea
22	Inland waterway	99	Non-standard mode
29	Water n.e.s.	99	Non-standard mode
31	Railway	31	Railway
32	Road	32	Road
39	Land n.e.s.	99	Non-standard mode
41	Pipelines and cables	99	Non-standard mode
90	Others, non-specified	99	Non-standard mode

For the compilation of GTCDIT, out of the more than 50 UN Comtrade Plus variables, the data of the nine variables above were retrieved from UN Comtrade Plus and integrated into the UNCTAD data warehouse via a server-to-server exchange, where the UN Comtrade coding of economies was mapped and aggregated to the coding system of UNCTADstat, as described above, and MoTs were aggregated to the five categories needed in the final GTCDIT dataset in accordance with table 2. As UN Comtrade Plus provides in some cases several records for the same observation, reflecting different methods of measurement, filters have been developed that ensure that only the one record per observation that is consistent with the specified formats for GTCDIT is retained.

The thusly extracted and converted UN Comtrade Plus data contained CIF values of imports from 136 countries, split up into around 23 000 bilateral trade connections (combinations of origin and reporting country) and, differentiating further by commodity groups, 7.5 million records. All countries that report CIF values also report almost exhaustively the corresponding quantity. However, as can be seen from table 5, the coverage of FOB values is much lower than that of CIF values. Only six countries have reported import values on both a CIF and

FOB basis (coverage types B and D). Breakdowns of imports by MoT are available for 23 countries (coverage types C and D), but only three of them (Angola, Peru and Madagascar) also reported FOB values (coverage type D). The individual reporting countries – represented as import destinations in the final dataset – that are associated with the four different coverage types are listed in table 16 in the Annex.

Table 5. Coverage of bilateral data on annual merchandise imports in UN Comtrade Plus, 2016

		FOB value	
		No	Yes
MoT break-down	No	*Coverage type A:* 110 reporting countries 18 297 connections 5 902 254 records	*Coverage type B:* 3 reporting countries 531 connections 162 920 records
	Yes	*Coverage type C:* 20 reporting countries 3 729 connections 1 329 238 records	*Coverage type D:* 3 reporting countries 523 connections 133 993 records

Note: All data coded in HS 2012 format with specified country of origin are considered.

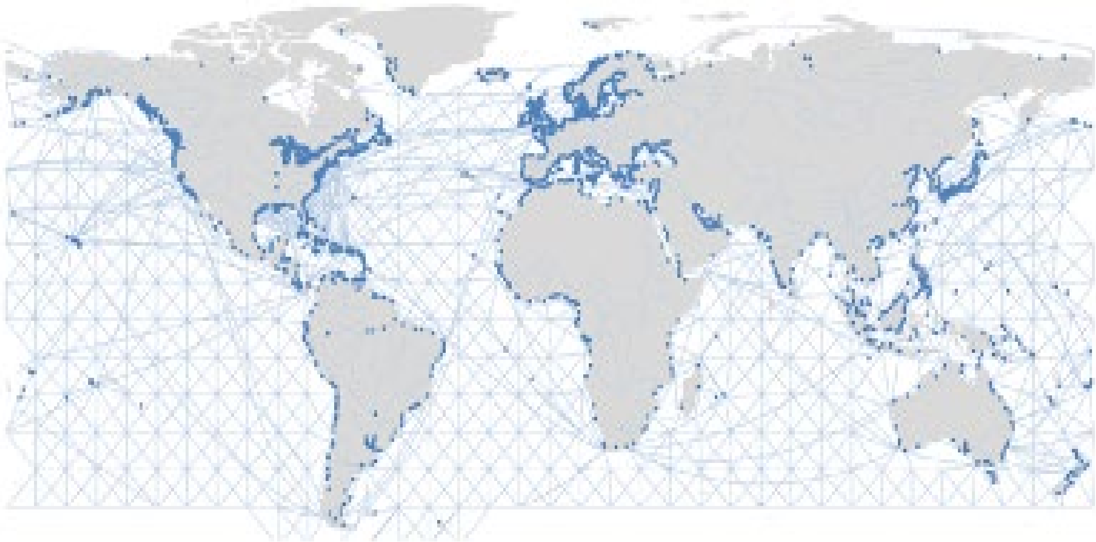
The fact that the new variables and breakdowns available in UN Comtrade Plus data were only to a limited extent populated with data from official sources has represented a major challenge for the compilation of GTCDIT. To cope with that challenge, most of the data gaps have been filled with estimates, as described in Section 7 below.

3.2. Distance matrix

The variable *distance* was compiled from a distance matrix constructed in the scope of this project. This matrix records estimates of the average distance goods need to travel at minimum to transport them from main city centres of one country to main city centres of another using a specific MoT. For the development of the distance matrix, in the first step, distances computed between 333 origins and destinations at major city level (centroids) were computed using a shortest-path model that identifies the shortest distance between locations, based on a GIS, after endogenizing travel times by type of infrastructure, intermodal transport, dwelling times, and mode-specific travel times (see figure 4). For more information on the construction of the algorithm and the underlying data, see Halim et al. (2018). In a second step, the distances between the centroids of the origin and destination countries were aggregated to distances between countries by calculating the median.

For modes that cannot technically serve as single means to transport goods between two specific countries – for example "railway" and "road" for goods delivered to or from an island state, or "sea" for goods delivered to or from a landlocked country – the corresponding cell of the matrix was flagged as 'unreachable'. This information was later used in the estimation of the breakdown of trade by MoT and the cleaning of the data from implausible cases (see section 7 below).

Figure 4. Schematic representation of global maritime transport networks on a geographic information system



4. Framework for data editing and imputation

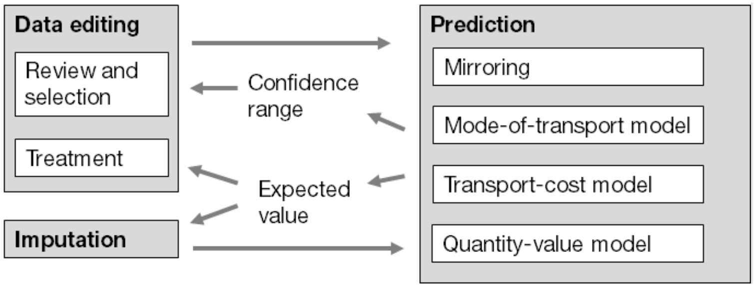
4.1. Underlying rationale

As shown in the previous section, UN Comtrade Plus offers a wealth of new information on transport costs and international merchandise trade at a considerable level of detail. However, it lacks by far all the data required for a full recording of transport costs worldwide, at detailed country, commodity and MoT levels. Thus, many data gaps needed to be filled. Furthermore, as the data are disseminated by UNSD in a relatively raw form, after only basic cleaning procedure applied, they should not be expected to be free from errors. Cleaning the data from obvious errors was therefore undertaken to ensure reliability. Due to the sheer size of the dataset, performing these tasks manually was not feasible. Therefore, automatic procedures were put in place. These are broadly outlined in this section and documented in more detail in sections 5 to 7.

Gap filling is guided by the aim of imputing estimates that are as close as possible to the corresponding true values missing in the source data. Cleaning data from apparent errors, also known as data editing, can be differentiated into three steps: (i) examining the data and identifying potential problems (review step); (ii) selecting cases for specified further treatment (selection step); and (iii) changing the selected data in a way considered appropriate for improving the data quality (treatment step) (UNECE, 2019).

The procedures put in place for data editing and imputation for GTCDIT rely on predictions of expected values in principal for all data cells of the dataset, as well as on estimations of the distribution of their error. Once these statistics have been compiled, confidence ranges can be constructed and used for accuracy checks (in the review and selection step). The in-sample predictions can be used to replace previously identified suspect values (in the treatment step), and the out-of-sample predictions can be used for the filling of gaps. This logical framework is portrayed in figure 5.

Figure 5. Logical framework for data editing and imputation



Applying the considerations above, data editing and imputation for GTCDIT proceeds according to the following steps:

1) Prediction, where we compile estimators representing expected values of the specific variables based on previously developed algorithms and models
2) Review and selection of suspect cases, where we carry out accuracy checks, based on a comparison of actual values with confidence intervals constructed around their corresponding predicted values, taking into account the estimated distribution of the prediction error
3) Treatment of suspect cases and filling of gaps, where we replace missing values and values identified as suspect in step 2 with estimators derived from the predictions made in step 1, a task known as imputation.

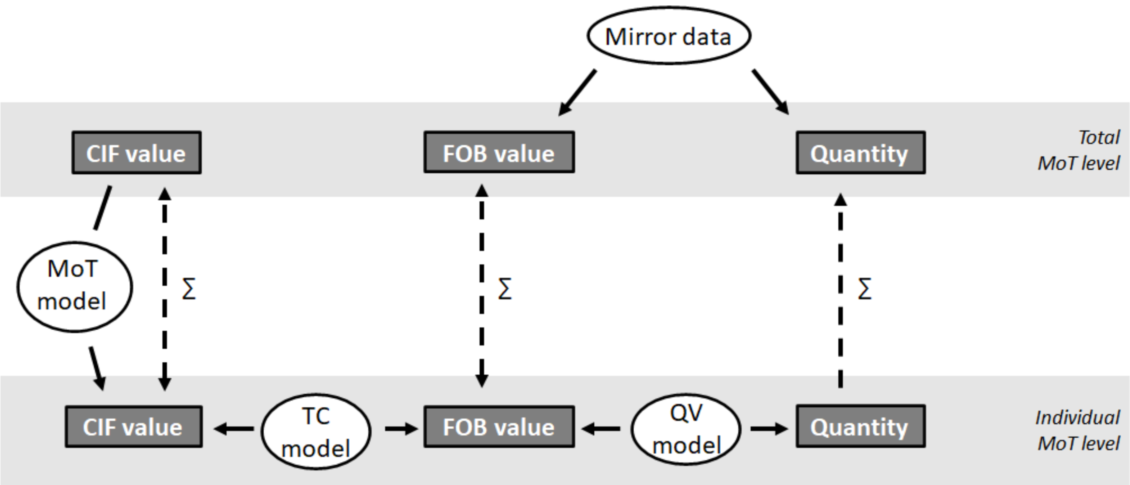
Once, cases identified as suspect have been replaced, the predictions made on their basis can be rerun and will yield predictions of expected values with higher precision than before. And once empty cells in some variables have been filled with imputation, extended opportunities exist for out-of-sample predictions of other variables, so that more data gaps can be filled. Running several iterations of prediction and subsequent imputation thus allows successively increasing data coverage and accuracy. Due to time constraints, for the present version of GTCDIT only three iterations were run. The full potential of this framework has thus not yet been exploited.

4.2. Prediction

Predictions were made based on mirroring and based on econometric models suited to estimate expected values of the different variables in the dataset. These models comprise: an MoT model which estimates the distribution of CIF values of imports over MoTs; a transport cost (TC) model that estimates elasticities of transport costs with respect to CIF values and distance; a quantity value (QV) model that estimates the elasticities of FOB prices with respect to quantity. These estimation and prediction techniques are described in detail in section 5 below.

Figure 6 gives an overview of which variables can be predicted, in principle, by the different methods and on which input variables these predictions can rely. The MoT model, by estimating mode shares, can predict CIF values at individual MoT level based on CIF values observed for MoT totals. The TC model, by estimating transport costs, can predict the FOB value based on a given CIF value or, alternatively, the CIF value based on a given FOB value, at the individual MoT level. The QV model can predict quantity based on the FOB value, and vice versa, at the individual MoT level. Mirror data, consisting of exports reported by trading partners, can be used for predictions of the FOB value and quantity at the total MoT level. Mirror data cannot be used, however, for predictions of CIF values and predictions FOB values or quantities at the individual MoT level, as exports are not reported in that form by any country in the UN Comtrade Plus data used. Finally, the CIF value, FOB value and quantity at total MoT level can also be obtained by summing up their corresponding entries in individual MoTs.

Figure 6. Prediction techniques and their applicability to input and output variables



4.3. Review and selection

In the review and selection step, we browse through the dataset to identify cases in which the recorded value deviates from its predicted expected value by an unusually high margin as compared to the usually observed error in a given segment of the dataset. With the help of the TC model, in principal, an implausible CIF value can be identified based on its corresponding reported FOB value, or an implausible FOB value can be identified based on its corresponding reported CIF value. Only one of these alternatives is possible, as the prediction of one variable – CIF or FOB value – by the model is always conditional on the value observed in the other. Likewise, with the help of the QV model, either an implausible entry for quantity can be identified based on its corresponding entry for the FOB value, or vice versa. Considering that the CIF value of imports is most often represented by the invoice value directly observed in the customs declaration and that the recording of monetary values at customs is usually more precise than the recording of quantity, the accuracy checks were carried out as follows. At the individual MoT level,

- the plausibility of the CIF value was evaluated using the MoT model, based on predictions obtained from the corresponding CIF value observed for the MoT total, whenever available;
- the plausibility of the FOB value was evaluated using the TC model, based on predictions obtained from their corresponding CIF values, whenever available;
- the plausibility of quantity was evaluated using the QV model based on predictions obtained from the corresponding FOB value, whenever available.

At the total MoT level, the plausibility of the FOB value was evaluated using mirror data. This process is described in more detail in section 6 below.

4.4. Treatment and filling of gaps

In the treatment and gap-filling step, the predicted expected value is imputed for any value previously identified as suspect or missing. Contrary to the review and selection step, the CIF and the FOB value are never simultaneously available for imputations done with the TC model, and the FOB value and quantity are never simultaneously available for predictions done with the QV model. Therefore, imputations can proceed in two directions along the arrows shown in figure 6. Accordingly, at the individual MoT level,

- a missing or suspect CIF value was replaced by the expected value derived from the CIF value observed for the MoT total, applying the MoT model, whenever available;

- a missing or suspect FOB value was replaced by either
  - the expected value derived from the CIF value, applying the TC model,
  - the expected value derived from quantity, applying the QV model,
  - or an estimator that combines both, depending of the availability of these predictions;
- a missing or suspect entry for quantity was replaced by the expected value derived from the FOB value, applying the QV model, whenever available.

At the total MoT level, a missing FOB value or quantity was replaced by either

- the corresponding mirror value,
- or the sum of values observed or estimated at individual MoT level,
- or an estimator that combines both, depending on the availability of mirror data and MoT breakdowns.

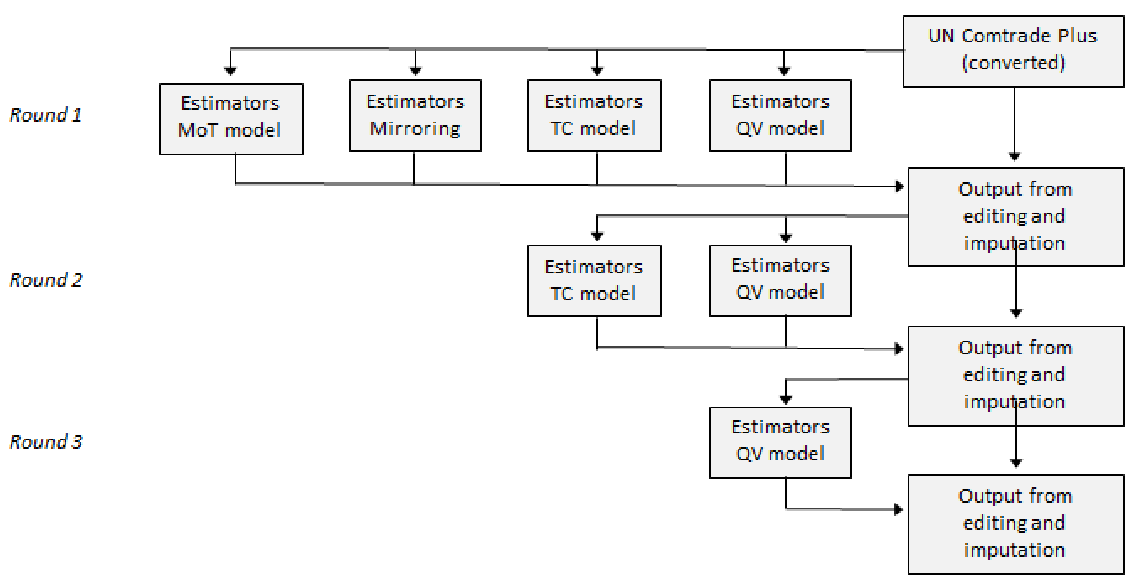
This process is described in more detail in section 7 below.

4.5. Sequence of iterations

As mentioned above, the sequence of prediction, review and selection, and treatment and gap-filling can be repeated for several iterations, so that the accuracy and coverage of the dataset is successively enhanced. For the compilation of the beta version of GTCDIT, three rounds of those iterations were run.

Figure 7 outlines the data generated in each round as well as their sources and flows. In the *first round*, all three models were run to make predictions. The predictions from the QV and TC models were used to identify outliers in FOB values and quantities at the detailed MoT level in records of coverage type D (see table 5 above), and to create new observations on FOB values, first, at the detailed MoT level, in records of coverage type C and, second, via aggregation, for MoT totals, in records of coverage type A. For records of coverage type A, new FOB values at the total MoT level were also generated by mirroring. When predictions of FOB values were obtained from various sources, these were reconciled with each other as described in section 7 below. Furthermore, the predictions from the MoT model were used in round 1 to identify outliers in CIF values and trade in infeasible MoTs, among records of coverage types C and D, and also to create new breakdowns of CIF values by MoT, thereby transforming type-A into type-C records.

Figure 7. Data flows in the three rounds of editing and imputation



In the *second round*, only the TC and QV models were run, now on data cleaned from outliers in the first round, to improve the predictions of FOB values and quantities, and to add FOB values to the MoT breakdowns of CIF values newly created by the MoT model in the first round. Thereby, most of the new records of coverage type C were further upgraded to coverage type D.

In the *third*, and final, *round*, only the QV model was run to carry out further refined outlier checks in quantities on cases of coverage type D, and to predict the quantities that fit the FOB values newly predicted in the second round.

5. Prediction methods

Let us now turn to a description of the specific prediction methods applied.

5.1. Mirroring

A first prediction method we used was mirroring. As mentioned above, GTCDIT is primarily constructed from the data on imports. The idea of the mirroring is to use data on exports for the filling in of gaps and for cross-checking and correcting the data reported from the import side.

In the extracted UN Comtrade Plus data with reference year 2016, data on export values are reported on an FOB basis only, never on a CIF basis, and they are never broken down by MoT. This is in line with IMTS recommendations (UNSD, 2011), which encourage countries to report the FOB value of imports, complementarily to the CIF value of imports, not making any recommendation about the CIF value for exports. Therefore, a comparison with mirror data is possible for FOB values only, not for CIF values, and only for data recorded at total MoT level.[2]

Table 6 shows the number of records for which a mirror entry can be found, out of the total number of records, differentiating by the coverage type as defined in table 5. For all four coverage types, the mirror data coverage is slightly below 50 per cent. For around 3.4 million out of 7.2 million records with missing FOB values in imports (coverage types A and C), a FOB value from mirror data is principally available for imputation. Furthermore, in the 145 thousand out of around 297 thousand cases with data coverage type B or D, reported FOB values could in principle be cross-checked against FOB values found in the mirror data.

Table 6. Mirror data coverage in UN Comtrade Plus

Coverage type	MoT breakdown	FOB value	Total number of records	Mirror data correspondence	
				Number of records	Per cent
A	no	no	5 902 254	2 780 326	47.1
B	no	yes	162 920	80 825	49.6
C	yes	no	1 329 238	636 048	47.9
D	yes	yes	133 993	64 502	48.1
All			*7 528 405*	*3 561 701*	*47.3*

However, the potential for using the extracted mirror data for filling of gaps and for verification of reported values is considerably constrained by cross-country discrepancies in the measurement and recording of international trade. Research has shown that the quantities and FOB values of the same flow reported by the importing and exporting country are often considerably different, due, for instance, to differences in

[2] However, investigation of 2017 data confirmed that several countries started reporting both CIF and FOB values from both import and export side. This is promising for future updates of the transport cost database.

classification, time of recording, exchange rates movements, different recording of triangular trade, underreporting, measurement errors, and probably processing errors. For that reason, in many cases, the difference between the CIF value of imports and the FOB value of the corresponding exports reported by partner countries for the same group of commodities cannot be explained by transport costs alone (Carrère and Christopher, 2014; Fortanier, 2016).

Based on the considerations above, mirroring was applied only in cases in which discrepancies in measurement and recording can be assumed to be small. To test for the absence of those discrepancies, we follow Gaulier et al. (Gaulier et al., 2008). Consider two countries, *o* and *d*, trading a given quantity *q* of a good. In the absence of combined reporting of the CIF and the FOB value by one reporting country, the CIF-FOB ratio (*z*) is sometimes estimated "implicitly", using the CIF value (*x*) observed on the importer's side and the corresponding FOB value (*y*) observed on the exporter's side:

(1) $$\hat{z}_{o,d}^{Imp} = \frac{x_{o,d}^{Imp}}{y_{o,d}^{Exp}}$$

Quantities, unlike the CIF and the FOB value, are observable on both the exporter's and importer's side in Comtrade Plus. Gaulier et al. (2008) suggest making use of that information and calculating a price-based estimator of the implicit CIF-FOB ratio, which can be seen as the value-based estimator of the implicit CIF-FOB ratio from formula 1, normalized by the quantity (*q*) reported on the importer and exporter side:

(2) $$\hat{z}_{o,d}^{*Imp} = \frac{\frac{x_{o,d}^{Imp}}{q_{o,d}^{Imp}}}{\frac{y_{o,d}^{Exp}}{q_{o,d}^{Exp}}}$$

The terms in the numerator and the denominator represent the CIF price declared by the importer and the FOB price declared by the partner country, respectively. The rationale behind this estimator arises from the fact that errors in quantity and value can be expected to be correlated, for example due to probable edit issues and cross-country differences in the methods of recording trade at the country borders.

Hummels and Lugovskyy (2006) define a reasonable range of the CIF-FOB ratio as [1, 2], implying ad valorem transportation costs between 0 and 100 per cent. We should expect the normalized implicit CIF-FOB ratio, z^*, to lie within the same range, as the minimum rate per unit should be higher than one and the maximum rate per unit should not exceed one plus the price of the good transported.

Table 7 presents descriptive statistics of the distributions of the value-based and price-based implicit CIF-FOB ratios defined above throughout the data extracted from UN Comtrade Plus. The second column describes the distribution of value-based ratio ($\hat{z}$), when all observations are considered. Although the median is slightly higher than 1, many observations fall far below 1 and far above 2. As one can deduct from the figures, 20 per cent of the cases lie beyond the range from 0.09 to 12.56. The third column provides the same statistics for the price-based implicit CIF-FOB ratio ($\hat{z}^*$). This distribution appears narrower than the former, but it again has long tails at both ends. Still, 20 per cent of the cases show rates smaller than 0.23 or greater than 4.05. Columns 4 and 5 show the results obtained when only values between 1 and 2 are retained. These remaining cases make up slightly more than one tenth of all observations. Without any correction, these data should still not be considered usable as such, as they are likely distorted by asymmetries.

Gaulier et al. (2008) suggest a prior quality check based on the coherence of the quantities declared by the trading partners. They propose dropping the cases that do not satisfy the criterion:

(3) $$\frac{Min\left(q_{o,d}^{Imp}, q_{o,d}^{Exp}\right)}{Max\left(q_{o,d}^{Imp}, q_{o,d}^{Exp}\right)} > 0.9$$

thus, the cases in which the absolute relative difference in the measurement of quantities on importer and exporter side exceeds 10 per cent. The columns 6 and 7 of table 7 report the same distribution statistics as above when only observations within the range from 1 to 2 are considered imposing the additional constraint that the quantity measured from the exporter side matches that measured from the importer side based on the criterion of Gaulier et al. (2008). The distribution of implicit CIF-FOB ratios appears now more in line with expectations. Although still relatively low, the median is now 1.10. Also, the distribution around this value is much narrower, with 80 per cent of the cases showing CIF-FOB ratios between 1.02 and 1.46. The range is far more satisfactory than in the previous cases, in which the size of the discrepancy in reported quantities was not yet controlled for. Moreover, the 90th percentile is now 1.46, which is a plausible figure for commodities with the highest transport costs. Unfortunately, the number of cases which pass that additional constraint represent only 6 per cent of the total observations. These 123 thousand cases were used for imputation and accuracy checking based on mirror data.

Table 7. Distribution of the value- and price-based implicit CIF-FOB, under different constraints

	All observations		Observations within range [1;2]		… of which discrepancies in quantity < 10%	
	Value-based	Price-based	Value-based	Price-based	Value-based	Price-based
Median	1.03	1.04	1.18	1.15	1.10	1.10
Percentile 10	0.09	0.23	1.02	1.02	1.02	1.02
Percentile 90	12.53	4.05	1.70	1.61	1.46	1.46
Number of observations		2 088 618		228 702		122 612
Percentage		100		11		6

Note: Observations with CIF and FOB values greater than 0 are considered.

5.2. The mode of transport model

The MoT model predicts the proportions in which the importing of traded goods from an origin (*o*) to a destination (*d*) country is realized by different MoTs (*m*). MoT models can be formulated as discrete choice models built on "utility maximization theory" (Cascetta, 2009). They describe and predict the choices of shippers given a set of alternative MoTs and their determinants. Within this framework, modes which give higher utility for shippers have a higher probability to be chosen. There are different types of discrete choice models that can be used to estimate modal share of transport modes. The most widely used are logit and probit models (Bierlaire, 1998).

In the context of GTCDIT, we invoked a multinomial logit model. These types of models feature a choice probability function that is easy to interpret. They are parsimonious yet able to capture choice behaviour of shippers across multiple transport modes. Our model estimates the probability (P) of choosing a particular mode *m* relative to all available alternatives in a sample, based on observed values of regressors:

(4) $$\Psi_{o,d,g',m} = \hat{\beta}_{0,m} + \hat{\beta}_{1,g'} a_{o,d,m} + \hat{\beta}_{2,g'} b_{o,d,m} + \hat{\beta}_{3,m} D_{o,d}^{contig} + \hat{\beta}_{4,o} + \hat{\beta}_{5,d}$$

$$\mathrm{P}_{o,d,g',m} = \frac{e^{\Psi_{o,d,g',m}}}{\sum_{i=1}^{M} e^{\Psi_{o,d,g',i}}}$$

where g' is an identifier of the HS 2-digits commodity group, Ψ is the choice utility, *a* is transport distance, *b* is transport time and D^{contig} is a dummy variable for the contiguity between the origin and the destination country.

β_0 is an MoT-specific constant, β_1 is a transport distance coefficient, β_2 is a travel-time coefficient, β_3 a contiguity coefficient. β_4 and β_5 are origin and destination fixed effects, respectively.

(5) $$\Psi_{o,d,g',m} = \hat{\beta}_{0,m} + \hat{\beta}_{1,g'} a_{o,d,m} + \hat{\beta}_{2,g'} b_{o,d,m} + \hat{\beta}_{3,m} D_{o,d}^{contig} + \hat{\beta}_{4,o} + \hat{\beta}_{5,d}$$

The estimation of the parameters of the multinomial logit model is typically done using maximum likelihood (Bierlaire M, 1998). However, due to the aggregate characteristics of the dataset, a dedicated estimation method had to be developed. This method uses a non-linear optimization method based on the evolutionary algorithm class "non-dominated sorting genetics algorithms II (NSGAII)" (Deb et al., 2000). The goal of this method is finding a set of parameter values that minimize the root mean square error (RMSE) between observed and modelled MoT shares across all modes and to maximize the coefficient of determination. We used observations of trade values aggregated at HS-2 level to estimate the coefficients of the model.

We estimated 97 MoT models, one for each of the 97 HS-2 commodity groups. The observation data used cover trade between 127 origin (exporting) and 19 destination (importing) countries in all MoTs except the "non-standard mode" (code 99). At HS 6-digit level, 4 128 316 observations were available, spread over 5 193 commodity groups. For the estimation, these were aggregated to 287 764 observations at HS-2 level. Despite the resulting reduction in degrees of freedom, the observation data basis is still sufficiently large. It has a broad country coverage, encompassing all development status categories.

Table 17 in the Annex presents summary statistics of the estimation results. As can be seen from the R-squared values, the estimated model yields varying degrees of accuracy in the different commodity groups. That is, it shows a very good fit and predictive power in some HS commodity groups, while others have relatively low R-squared values. Relatively low predictive power could be the result of unexplainable heterogeneity in the data used to estimate the model, possibly caused by error in the reporting of observation data. This indicates a need for additional variables that can better explain this heterogeneity. To better capture the unexplained heterogeneity across HS groups at 6-digit level, the model might need to be augmented with more detailed commodity-specific coefficients or additional independent variables which can explain the variation in the observation data. This further development of the model may be a subject of future research.

Predictions were made by applying the estimated choice probabilities to the CIF value observed at total MoT level:

(6) $$\hat{x}_{o,d,g,m} = x_{o,d,g} \hat{\mathrm{P}}_{o,d,g',m}$$

Note that common choice probabilities estimated for entire 2-digit groups of HS, identified by g', are used to predict the mode-specific CIF values at HS 6-digit level, as identified by *g*. This aggregated model specification has been done to retain simplicity of the model and reduce the computational complexity of model estimation. The model predicted within-sample modal shares of the modes air, sea, rail and road with a root mean squared error (RMSE) equivalent of 0.765 standard deviations of the observed modal shares. Translating the modal shares into disaggregated CIF values, the RMSE amounts to 0.787 times the standard deviation.

5.3. The transport cost model

A linear model was invoked to estimate the elasticities of transport costs with respect to the value of imports, expressed CIF, and to the distance (*a*) between the origin (*o*) and destination (*d*) country, where transport costs are measured as the difference between the CIF (*x*) and the FOB (*y*) value, in accordance with the definition in table 1 above. The model is specified as:

(7) $$\log\left(x_{o,d,g,m} - y_{o,d,g,m}\right) = \beta_{0,g,m} + \beta_{1g'm}\log\left(x_{o,d,g,m}\right) + \beta_{2g'm}\log\left(a_{o,d,m}\right) + \omega_{o,d,g,m}$$

where *m* designates the MoT used, g' is an identifier of product groups at the 2-digit level, and *g* an identifier of the product group at the 6-digit level of HS. β_0 represents a product-group specific fixed effect, and β_1 and

β_2 are the elasticities with respect to the CIF value of trade and distance, respectively. ω is an independent and identically distributed error term orthogonal to the regressors.

The model was estimated with least squares separately for each combination of commodity group (at 2-digit level of HS) and MoT. Insignificant coefficients, evaluated at the 5 per cent significance level, were removed from the model using backward elimination. The regressions were run on a total of 201 639 observations. They yielded a joint coefficient of determination (R squared) of 0.883. Table 18 in the Annex presents summary statistics of the estimation results.

Using the estimated elasticities and fixed effects, FOB values were predicted from observed CIF values, as follows:

(8) $$\hat{y}_{o,d,g,m} = x_{o,d,g,m} - e^{\hat{\beta}_{0,g,m}} x_{o,d,g,m}{}^{\hat{\beta}_{1g'm}} a_{o,d,m}{}^{\hat{\beta}_{2g'm}}$$

The model predicted within-sample logarithmic transport costs with a RMSE of 0.341 times the standard deviation, and within-sample logarithmic FOB values with a RMSE of 0.097 times the standard deviation.

5.4. The quantity value model

The QV model estimates the elasticity of the FOB value (*y*) to quantity (*q*), controlling for the effect of transport distance (*a*). The econometric specification is given by:

(9) $$\log(y_{o,d,g,m}) = \beta_{0,g,m} + \beta_{1g'm}\log(q_{o,d,g,m}) + \beta_{2g'm}\log(a_{o,d,m}) + \omega_{o,d,g,m}$$

where, as above, β_0 represents a product-group specific fixed effect. This fixed effect accounts for group-specific differences in prices and for the fact that quantities of different product groups are measured in different units. β_1 and β_2 represent the elasticities with respect to quantity and distance. ω is an independent and identically distributed error term orthogonal to the regressors.

The model was estimated with least squares separately for each combination of commodity group, at 2-digit level of HS, and MoT. As with the TC model, insignificant coefficients (at 5 per cent significance level) were removed from the model using backward elimination. The regressions were run on a total of 192 438 observations and yielded a joint coefficient of determination (R squared) of 0.761. Table 19 in the Annex presents summary statistics of the estimation results.

Once the coefficients of the model were estimated, FOB values could be predicted from observed quantities as follows:

(10) $$\hat{y}_{o,d,g,m} = e^{\hat{\beta}_{0,g,m}} q_{o,d,g,m}{}^{\hat{\beta}_{1g'm}} a_{o,d,m}{}^{\hat{\beta}_{2g'm}}$$

And quantities could be predicted from observed FOB values using the formula

(11) $$\hat{q}_{o,d,g,m} = \frac{y_{o,d,g,m}{}^{\frac{1}{\hat{\beta}_{1g'm}}}}{e^{\frac{\hat{\beta}_{0,g,m}}{\hat{\beta}_{1g'm}}} a_{o,d,m}{}^{\frac{\hat{\beta}_{2g'm}}{\hat{\beta}_{1g'm}}}}$$

The model predicted within-sample logarithmic quantities with a RMSE of 0.489 times the standard deviation.

6. Accuracy checks

Using the models above we obtain estimators of the CIF value, the FOB value and quantity of bilateral trade, detailed by commodity group, for each individual MoT. These can be interpreted as the values we would expect based on the statistical relationships observed in the UN Comtrade Plus data. Once we have an idea of the probability distribution of the error of these predictions, we can use that information for evaluating the plausibility of observed differences between actual and estimated values.

6.1. Method applied

When using the estimated error distribution of predictions for the identification of suspect values, we should consider that the distributions of the CIF value, the FOB value and quantity in source dataset are strongly skewed to the right, as our primary inspections of the data have shown. This is in line with the fact that the domains of the variables are bounded from below, at zero, and unbounded from above. The upper tails are often made up of extremely high values, usually representing imports or exports of large trading economies, such as China, Germany or the United States of America. In logarithms, however, the shape of the distributions is almost symmetric and bell-shaped, as the graphs in the next section will demonstrate (see figure 7). The transformation into logarithms effectively cleans the distributions from the effect of the differences in size among observation units, so that the central limit theorem applies. Therefore, in the following, when using the error distribution of a variable in data editing and imputation, we express that error and its standard deviation with reference to the logarithmic form of the variable.

The standard deviation of the prediction error $\hat{\sigma}_{log(\widehat{\Theta})}$ of the logarithm of the point estimator $\widehat{\Theta}$ was estimated by calculating the difference between the predicted and the observed value for all observations included in a specific stratum of the source dataset, and calculating its standard deviation. The strata were usually defined with reference to commodity group at two-digits level of HS (*g'*) and MoT (*m*). However, for the mirror data predictions at total MoT level, where less observations are available, only the first digit of HS was used for the construction of strata.

We applied the standard formula for the calculation of the standard deviation of the prediction error:

(12) $$\hat{\sigma}_{log(\widehat{\Theta}),g',m} = \sqrt{\frac{1}{n_{g',m}} \sum_{i=1}^{n_{g',m}} \left(\widehat{\omega}_{g',m,i} - \overline{\widehat{\omega}}_{g',m,i}\right)^2} \quad \text{with } \widehat{\omega}_i = log\left(\widehat{\Theta}_i\right) - log(\Theta_i)$$

where *i* is an index running over all combinations of origin country, destination country and six-digit commodity group within a stratum; and *n* is the stratum size. Θ is the actual and $\widehat{\Theta}$ the predicted value. Values smaller than one, in Θ and $\widehat{\Theta}$, were set to one, so that their logarithms were truncated at 0.

We defined confidence intervals as the range from five error standard deviations below to five error standard deviations above the logarithmic value of the point estimator:

(13) $$log(\Theta) \notin \left[log\left(\widehat{\Theta}\right) - 5\hat{\sigma}_{log(\widehat{\Theta})}; log\left(\widehat{\Theta}\right) + 5\hat{\sigma}_{log(\widehat{\Theta})}\right]$$

Any analyzed figure from the UN Comtrade Plus extraction which, transformed into its logarithm, did not fall into that range was classified as suspect and accordingly flagged for being replaced by an estimate. Note that this is equivalent to evaluating the accuracy of the non-transformed UN Comtrade figures against the confidence interval below, defined in relative terms:

(14) $$\Theta \notin \left[\frac{\widehat{\Theta}}{e^{5\hat{\sigma}_{log(\widehat{\Theta})}}}; e^{5\hat{\sigma}_{log(\widehat{\Theta})}}\widehat{\Theta}\right]$$

This means, for example, a discrepancy between an observed value of US$90 million and a predicted value of US$100 million is treated the same way as a discrepancy between an observed value of US$90 000 and a predicted value of US$100 000.

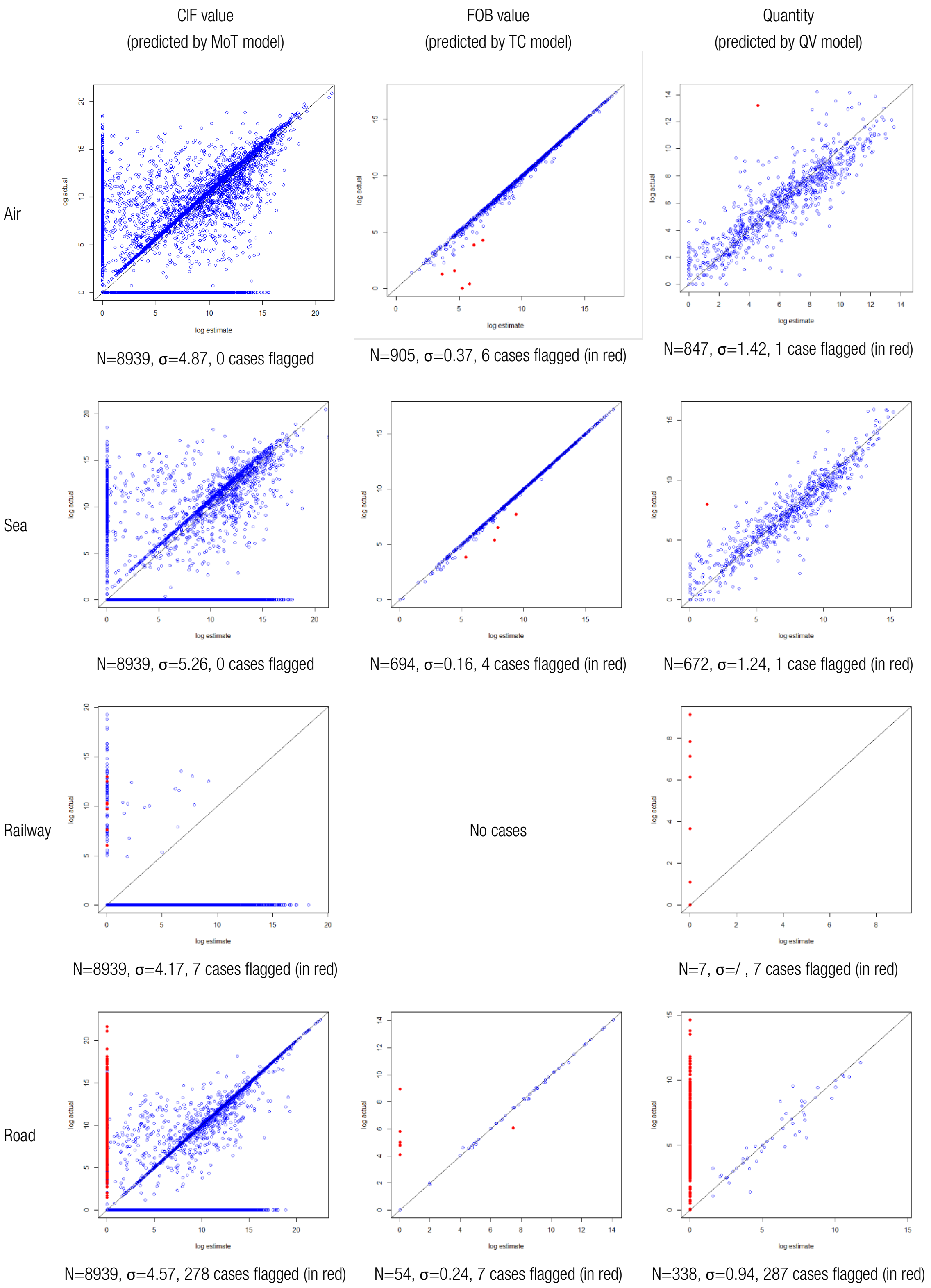
The accuracy checks above were applied to all three variables in the dataset, setting Θ for the CIF value (*x*), the FOB value (*y*) or quantity (*q*). respectively. As mentioned above, at individual MoT level, the accuracy of CIF values was evaluated using the predictions made by the MoT model based on the CIF values observed at total MoT level; the accuracy of the FOB values was evaluated using the predictions made by the TC model, based on the observed (disaggregated) CIF values; and the accuracy of quantities was evaluated using the QV model, based on the observed FOB values. However, as an overriding criterion, whenever the distance matrix indicated that transport by a given mode is not feasible as single MoT, for a specific pair of origin and destination country, any CIF value, FOB value or quantity greater than zero was also flagged for being set to zero. For MoT totals, the accuracy of FOB values was evaluated based on mirror data.

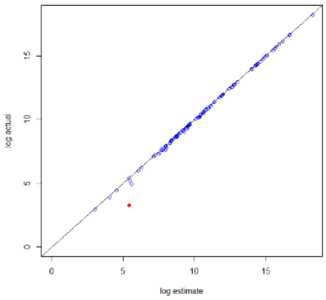
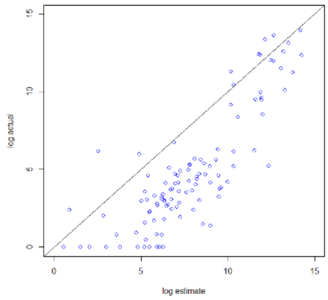
6.2. Illustrations based on the data

Figure 8 shows scatterplots of observed values (on the y-axis) against predicted values (on the x-axis), both in logarithms, for a randomly selected 2-digits commodity group: pharmaceutical products (HS code 30). The diagonal lines represent perfect match between observed values and their predictions. The farther away a point is from that line the greater the discrepancy, and the more reason we have to doubt the accuracy of the reported value. Cases selected as suspect, due to an extraordinarily large discrepancy between the estimated and the observed value, in accordance with the method above, are marked in red. In general, the prediction of the FOB value appears to have higher precision than the prediction of the CIF value and quantity. For the CIF value, more data points can be analyzed than for the FOB values and quantity, as the MoT model, in contrast to the TC and QV models, does not require the disaggregated FOB value as input. It can be run on all records with coverage type C or D (see table 5), whereas the TC and the QV model are limited to records with coverage type D.

It is worth noting that the plots for the CIF value show several points directly located on the y-axis. These represent cases in which the observed value is practically zero (less than 1), but not its prediction. Other points are located on the x-axis, representing cases in which the prediction suggests no trade via the given mode although such trade is actually observed. In those cases, the MoT seems misclassified based on the prediction from the MoT model. However, the error standard deviation of that prediction is not small enough to flag those cases as suspect with sufficient certainty. By contrast, several cases located on the y-axis, in the modes railway and road, were flagged as suspect. These are cases in which, based on the distance matrix, it is not realistic to assume that the given MoT was used as the single transport means. Examples include specific medicaments in doses or packed for retail sale (HS code 300490) of a value of US$2.6 billion (CIF) and specific antisera and blood for medical use (HS code 300210) of a value of US$1.5 billion (CIF), both declared by Germany as imports by road from the United States of America. Other examples consist of specific medicaments in doses or packed for retail sale (HS code 300490) of a value of US$400 thousands (CIF) declared by Hungary as imports by railway from Japan, as well as goods of the same category of a value of US$300 thousands (CIF) declared by Slovakia as imports by railway from the United States of America. For each of these cases, judged as misclassified, apart from the CIF value a corresponding entry of quantity exists. These are also flagged as misclassified and show up as red points in the scatterplots for railway and road in the right-hand column of the figure.

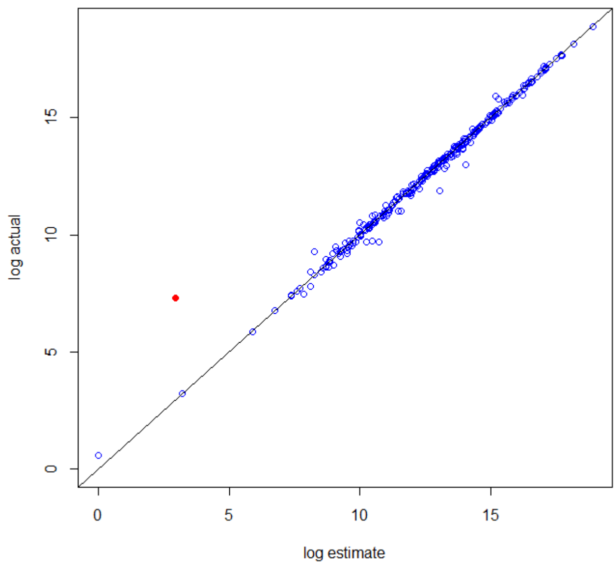
Figure 8. Identification of suspect values in trade with pharmaceutical products (HS code 30) at detailed MoT level, using model predictions



	CIF value (predicted by MoT model)	FOB value (predicted by TC model)	Quantity (predicted by QV model)
Other	No cases	 N=110, σ=0.21, 1 case flagged (in red)	 N=110, σ=2.06, 0 cases flagged (in red)

The TC and the QV model which require MoT specific observations as inputs, and the MoT model which breaks down CIF values by MoT are used for direct predictions at the individual MoT level only. For MoT totals, the mirror-flows, consisting of exports reported by trading partners on an FOB basis, can be used for accuracy checks once these mirror data have been cleaned from asymmetries. For pharmaceutical products (HS code "30"), only two such mirror flows can be found. Thus, the degrees of freedom would not be sufficient to estimate the standard deviation of the errors within this stratum. Due to the overall smaller number of observations among mirror flows than among model predictions, broader strata than in the previous cases have been used for the identification of suspect values based on mirror data. These strata are identified by the first digit of HS only. As figure 9 shows, the comparison of imports in HS product group "3" with their corresponding mirror data can rely on 280 cases of which only one case is identified to be out of range. This case is represented by specific vinyl chloride polymers (HS code 391620) imported by Argentina from Costa Rica, for which the reported FOB value is almost 10 times higher than its corresponding mirror value suggests.

Figure 9. Identification of suspect values in trade with products of HS group "3" at total MoT level, using mirror data



N=280, σ=0.33, 1 case flagged (in red)

6.3. Results

Table 8 summarizes the outcomes of the accuracy checks described above. The fourth and the fifth column present summary measures of the precision of the underlying prediction method. The RMSE expressed in units of the standard deviation is a (unit-free) measure of the accuracy with which the different methods predict the observed values in the dependent variables on average. The bias, also expressed in units of the standard deviation, measures the relative amount by which that prediction should be expected to be above (positive value) or below (negative value) the actual value, on average. For example, a RMSE of 0.097 standard deviations for the TC model indicates that the RMSE of its prediction accounts, when transformed into logarithms, on average for almost ten per cent of the standard deviation of logarithmic FOB values recorded in the used UN Comtrade Plus extraction at individual MoT level. The normalized bias of 0.024 indicates that the predicted logarithmic FOB value is on average 0.024 standard deviations higher than the actually observed logarithmic value.[3] As the distribution of total CIF values over MoTs is more difficult to predict than the relationship between the CIF and the FOB value, the MoT model predictions show a substantially higher normalized RMSE than the TC model predictions. The predictions of quantities generated by the QV model are also less exact than the predictions of FOB values produced by TC model. This was also revealed by figure 8 above.

Table 8. Accuracy checks outcomes

Observations analysed		Prediction method	Prediction accuracy[2]		Results of cleaning		
Variable	Level		Bias (in SD)	RMSE (in SD)	No. of cases observed	No. of cases analyzed	No. of cases adjusted
CIF value	Individual MoT	MoT model	0.140	0.787	7 316 155	4 288 812	47 515
FOB value	Individual MoT[1]	TC model	0.024	0.097	217 990	204 966	3 675
Quantity	Individual MoT[1]	QV model	-0.080	0.489	2 159 979	202 840	4 106
FOB value	MoT total	Mirroring	-0.002	0.196	296 913	1 191	8

[1] Cases with zero reported trade (CIF value) in the specific MoT are not counted in.

[2] With reference to logarithmic values.

The three columns on the right-hand side of the table show the outcomes of the accuracy checks. Out of the 7.3 million CIF values recorded in the data extraction from Comtrade Plus for individual MoTs, 4.3 million could be subjected to an accuracy check based on the MoT model.[4] Around 50 thousand of them (1.1%) have not passed that check and have been adjusted. These mostly represented cases of positive trade in a mode judged impossible to be the sole means for transporting of goods from the origin to the destination country. In slightly more than 200 thousand cases, a non-zero CIF value was accompanied by a corresponding entry for the FOB value in the source data. Most of these values could be subjected to an accuracy check based on the TC model. Around 3 700 of them (1.8%) have been found to be out of range. Quantity is reported at individual MoT level in 10 times more cases than the FOB value. However, as FOB values are required as input, also only around 200 000 of them could be checked based on the QV model. In around 4 100 cases (2.0%) these checks were not passed. Mirroring, finally, could be used to compare around 1 200 out of the around 300 thousand FOB values reported for MoT totals in UN Comtrade Plus. In eight cases, the discrepancy between the directly reported import and its corresponding mirror flow was so large that the observation was flagged as suspect.

[3] Note that the TC model is specified in a way that it yields unbiased in-sample predictions of logarithmic transport costs. This does not necessarily imply unbiased prediction of logarithmic FOB values.

[4] For instance, CIF values in non-standard modes could not be checked, as predictions were not available.

7. Gap filling and replacement of suspect values

7.1. Method applied

Expected values derived from predictions made by the methods above were used not only for the identification of suspect values but also for their treatment and to fill data gaps. As figure 6 above has shown, in some cases this exercise could rely on predictions from more than one model that could be reconciled. Specifically, at the individual MoT level, expected FOB values could be obtained from two sources: from the CIF value using the TC model, and from quantity using the QV model. Furthermore, expected FOB values at total MoT level could be compiled by both mirroring and summing up the values recorded for the individual MoTs.

The predictions of a single observation derived from several models rely on different information sets and have different levels of precision. Merging them into a single estimator enables broadening the information base used. For this, we follow two steps: first, we reconcile at the individual MoT level the predictions of the FOB value obtained from the TC and the QV model (horizontal reconciliation). Second, we reconcile the estimated and actual values across individual MoTs and with MoT totals (vertical reconciliation).

For the *horizontal reconciliation* in the first step, we construct a minimum distance estimator, the logarithm of which represents a weighted average of the logarithmic predictions obtained from the TC and the QV model, where the weights are proportional to the inverse error variance of these predictions:

(15) $$log\left(\hat{y}_{o,d,g,m}\right) = \frac{\frac{log\left(\hat{y}_{TC,o,d,g,m}\right)}{\hat{\sigma}_{log(\hat{y}_{TC}),g,m}{}^{2}} + \frac{log\left(\hat{y}_{QV,o,d,g,m}\right)}{\hat{\sigma}_{log\left(\hat{y}_{QV}\right),g,m}{}^{2}}}{\frac{1}{\hat{\sigma}_{log(\hat{y}_{TC}),g,m}{}^{2}} + \frac{1}{\hat{\sigma}_{log\left(\hat{y}_{QV}\right),g,m}{}^{2}}}$$

That way, the more imprecise prediction, evaluated based on the observed error variance, is reflected less in the reconciled estimator than the more precise prediction. It can be shown that this estimator represents the linear combination of TC-model and QV-model predictions with minimum error variance under the assumption that the errors from the two predictions are stochastically independent. The expected error standard deviation of the reconciled estimator is then given by

(16) $$\hat{\sigma}_{log(\hat{y}),g,m} = \frac{\sqrt{\frac{1}{\hat{\sigma}_{log(\hat{y}_{TC}),g,m}{}^{2}} + \frac{1}{\hat{\sigma}_{log\left(\hat{y}_{QV}\right),g,m}{}^{2}}}}{\frac{1}{\hat{\sigma}_{log(\hat{y}_{TC}),g,m}{}^{2}} + \frac{1}{\hat{\sigma}_{log\left(\hat{y}_{QV}\right),g,m}{}^{2}}}$$

(Hoffmeister 2022, forthcoming). Throughout the dataset, in most cases the predictions from the TC model dominate the horizontally reconciliated estimator, due to their low error variance compared to the QV model, as revealed by table 6 above.

The *vertical reconciliation* in the second step is carried out in a way that the sum of the squared distances from the logarithms of the horizontally reconciled estimators, in units of their standard deviation, is minimized under the constraint that all values in individual MoTs add up to their corresponding total. It can be shown that this minimum is reach when

(17) $$\frac{log\left(\hat{\Theta}^{*}_{o,d,g,k}\right) - log\left(\hat{\Theta}_{o,d,g,k}\right)}{\hat{\sigma}_{log(\Theta),g,k}{}^{2}} = \frac{log\left(\hat{\Theta}^{*}_{o,d,g,l}\right) - log\left(\hat{\Theta}_{o,d,g,l}\right)}{\hat{\sigma}_{log(\Theta),g,l}{}^{2}} = \frac{log\left(\hat{\Theta}_{o,d,g}\right) - log\left(\hat{\Theta}^{*}\right)}{\hat{\sigma}_{log(\Theta),g}{}^{2}}$$

for all *k* and *l* as representations of *m,* where $\widehat{\Theta}_{o,d,g,m}$ designates the initial estimator and $\widehat{\Theta}^*_{o,d,g,m}$ the vertically reconciled estimator in the MoT *m.* $\widehat{\Theta}_{o,d,g}$ and $\widehat{\Theta}^*_{o,d,g}$ represent their corresponding MoT totals, so that

(18) $$\widehat{\Theta}_{o,d,g} = \sum_m \widehat{\Theta}_{o,d,g,m}$$

$$\widehat{\Theta}^*_{o,d,g} = \sum_m \widehat{\Theta}^*_{o,d,g,m}$$

(Hoffmeister 2022, forthcoming). The minimization problem above cannot be solved by linear algebra, as formula 17 involves differences between logarithms and formula 18 sums of non-logarithmic values. It has therefore been solved by iterative fitting, where adjustments in the first step focused on the removal of any inequalities in weighted changes applied to the initial estimators, in accordance with formula 18, and adjustments in the second step focused on the recalculation of the sums over all feasible MoTs, in accordance with formula 19 (*ibd.*). These two steps were repeated successively 50 times. After this, all significant discrepancies between sums and corresponding totals were found to have disappeared.

A special case is represented by the replacement of values with zero when the recorded MoT was identified to be impossible to be the single MoT used. As the impossibility of any positive trade in that case appears certain, the error standard deviation was set to zero, and the reconciliation across MoTs carried out in a way that all logarithmic values recorded for feasible MoTs were increased by the same amount. Thus, $\hat{\sigma}_{log(\Theta)}$ in formula 17 was treated to be positive and to have equal value in all MoTs.[5]

7.2. Illustrations based on the data

To illustrate how the imputation process works, let us focus on a few examples from the 8 939 records, represented in figure 8 above, displaying trade in pharmaceutical products.

Example 1: Imputation of missing FOB values in imports of medical gel preparations from Turkey to Germany

Table 9 shows, on its left-hand side, the figures recorded in UN Comtrade Plus for imports of specific gel preparations made for use in human or veterinary medicine (HS code 300670) of Germany from Turkey. The CIF value and its breakdown by MoT is available, but corresponding FOB values are missing. The right-hand side of the table shows predicted FOB values obtained from the TC model, the QV model and from mirroring, as described in the previous chapters. Note that no trade is recorded in UN Comtrade Plus for the modes "railway" and "other". That means, a zero FOB value is also implicitly observed in these cases.

Table 9. Original data and model predictions

Mode	UN Comtrade data			Predictions for the FOB value (US$)					
	CIF value (US$)	FOB value (US$)	Quantity (kg)	TC model		QV model		Mirroring	
				y	$\sigma_{\log(y)}$	*y*	$\sigma_{\log(y)}$	*y*	$\sigma_{\log(y)}$
Air	40 526	/	529	37 341	0.37	20 719	1.66	/	/
Sea	31 299	/	36 841	30 469	0.16	403 211	1.40	/	/
Railway	-	-	-	-	-	-	-	/	/
Road	254 124	/	159 946	245 755	0.24	1 168 567	1.59	/	/
Other	-	-	-	-	-	-	-	/	/
Total	*325 950*	/	*197 316*	*313 565*	/	*1 592 497*	/	*277 399*	*0.33*

[5] Note that the value chosen does not matter for the result, as long as it is different from zero.

We need to reconcile the different predictions of FOB values in a way that we obtain one single estimator per MoT, and one for the MoT total, where the sum of the former should equal the latter. As outlined above, we do this in two steps. First, we reconcile the predictions obtained from different models for the same MoT (horizontal reconciliation). The outcome of that step is shown in table 10. The weights in the columns 3 and 6 determine the degree by which the predictions from the TC and the QV model are taken into account in the horizontally reconciled estimator. They represent the inverse error variances, thus the inverse of the squared values of $\sigma_{\log(y)}$ in table 9, rescaled by constant in a way that their sum equals one. Apparently, the predictions from the QV model are by far less precise than the predictions from the TC model, as reflected by the higher standard deviation of their error. Correspondingly, the impact of these estimators on the reconciliation, indicated by the weights, is much smaller than the impact of the more precise estimators of the TC model. Applying the weights in columns 3 and 6 to the corresponding logarithmic point estimators, we obtain the logarithmic value of the reconciled estimator in column 8. Column 9 shows the estimated standard deviation of its error, calculated using formula 16. In column 7, the reconciled point estimator is transformed back from its logarithmic into its non-logarithmic form.

Table 10. Reconciliation of FOB value predictions across models

Mode	TC model			QV model			Reconciled		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Estimate	Log estimate	Weight	Estimate	Log est.	Weight	Estimate	Log est.	$\sigma_{\log(y)}$
Air	37 341	10.53	0.95	20 719	9.94	0.05	36 336	10.50	0.36
Sea	30 469	10.32	0.99	403 211	12.91	0.01	31 536	10.36	0.16
Road	245 755	12.41	0.98	1 168 567	13.97	0.02	254 092	12.45	0.23

In the specific case shown, the alignment with the QV model estimator increased the TC model estimator for sea to an extent that the reconciled estimator became slightly larger than the CIF value recorded in UN Comtrade Plus. This is not possible, however, as transport costs cannot be negative. Therefore, for the further calculations, the restriction was imposed that the upper bound represented by the CIF value cannot be exceeded.

So far, the estimators above were reconciled independently from each other for the specific MoT to which they apply; information from the mirror data about their joint sum has not yet been used. According to mirror data, here represented by the exports to Germany reported by Turkey for the same commodity group in all MoTs was lower than the sum of the FOB-value horizontally reconciled estimators above by US$44 thousand. This implies a relative discrepancy of 14 per cent. The steps carried out to reconcile the MoT-level estimators from the first step with the mirror-data prediction of the total, and among each other, are illustrated in table 11. The left-hand side of the table shows the only horizontally reconciled estimators in the different MoT, after imposition of the restriction not to exceed the CIF value, and the estimator obtained from the mirror data, alongside their error variances, where the error variances are given by the squares of the error standard deviations ($\sigma_{\log(y)}$) from column 9 of table 10. For the vertical reconciliation, we reduce the logarithmic estimators at the individual MoT level and increase the estimator for the MoT total by amounts that, after deflation by the respective error variance, are equal and ensure adding up of the estimators in individual MoTs to the estimator for the MoT total. For actually observed values, as for the modes "railway" and "other", the estimated error standard deviation is zero, and no adjustment is made. The right-hand side of the table shows the results of that reconciliation. That the first condition above is met is manifested by the fact that the absolute difference between columns 5 and 2, divided by column 3, is equal across all rows with estimated data, amounting to 0.8947. That the second condition is met is manifested by the equality between the sum and the total, shown in the two rows at the bottom of the table in column 4.

Table 11. Reconciliation of estimated FOB values across modes

Mode	Before			After	
	(1)	(2)	(3)	(4)	(5)
	Estimate	Log estimate	$\sigma_{\log(y)}^2$	Estimate	Log estimate
Air	36 336	10.50	0.13	32 395	10.39
Sea	31 299	10.36	0.03	30 577	10.33
Railway	-	-	-	-	-
Road	254 092	12.45	0.05	242 086	12.40
Other	-	-	-	-	-
Sum	*321 727*	/	/	*305 058*	/
Total (mirror)	277 399	12.53	0.11	305 058	12.63

Example 2: Treatment of an out-of-range value in imports of antibiotics from the United Arab Emirates to Angola

Table 12 shows the CIF value, FOB value and quantity of imports of specific antibiotics not traded in measured doses (HS code 300320) from the United Arab Emirates to Angola. All these imports were delivered by air. The reported quantity of 550 tons appears rather high, considering that the FOB value is indicated to be only US$ 1480, implying a price per kg of less than one US$-cent.

Table 12. Original data

Mode	CIF value (US$)	FOB value (US$)	Quantity (kg)
Air	1 631	1 480	550 000
Sea	-	-	-
Railway	-	-	-
Road	-	-	-
Others	-	-	-
Total	*1 631*	*1 480*	*550 000*

In fact, based on the QV model, this case was identified as an out-of-range value to be treated. The reason becomes clear from table 13. For the given type of commodity, based on the available information on the FOB value, the MoT used and distance, the QV model predicted a quantity of only 97 kg for these imports. This would correspond with a per-kg price of US$15.3. Of course, this prediction is inflicted with an error, the distribution of which has an estimated standard deviation of 1.42 when calculated from logarithms. In accordance with the tolerance limit set up, values that, in logarithms, fall short of a range from 5 error standard deviations below to 5 error standard deviations above the logarithmic estimated value are considered to be out of range and flagged for being treated. In the present case, the so defined confidence range, in terms of non-logarithmic values, spans from 8 g to 118 tons. The value of 550 tons observed in UN Comtrade Plus falls far beyond that already broad range, and thus will be treated.

As air represents the only MoT used, the value observed for the MoT total is not influenced by trade in any other MoT. It is thus considered equally suspect as the value observed for air and subjected to the same adjustment. As an outcome, the value of quantity is changed from 550 tons to 97 kg, both for transport by air and for the MoT total.

Table 13. Prediction for quantity from the QV model and its precision

	Estimate	$\sigma_{log(q)}$	Confidence Range Lower bound	Upper bound
Value	97	/	0.08	118 243
Log value	4.57	1.42	-2.54	11.68

Example 3: Adjustment for impossible MoTs as single transport medium in the case of imports of various prophylactic or therapeutic medicaments from Canada to Germany

Table 14 shows the imports of various prophylactic or therapeutic medicaments (HS code 300490) from Canada to Germany recorded in UN Comtrade Plus. Based on the information from the Distance Matrix, it is not realistic to assume that goods were transported from Canada to Germany by road. It is more likely that they were transloaded from air or sea to road on their way. Re-allocating the entries for the goods transported by road to the MoT categories air and sea will give a more realistic picture of the primarily used mode.

Table 14. Information recorded in UN Comtrade Plus and the Distance Matrix

MoT	UN Comtrade Plus: CIF value	FOB value	Quantity	DR Matrix: Unreachable?
Air	26 865 736	/	35 955	false
Sea	14 320 232	/	270 769	false
Railway	-	-	-	true
Road	73 585 248	/	15 302	true
Other	-	-	-	/
Total	*114 771 217*	*/*	*322 026*	*/*

Therefore, the following adjustments are carried out. As shown in table 15, the value of the implausible MoT is set to zero and the resulting difference between the new and the previous sum is re-allocated in the same way as described in example 2 above, applying the formulas 17 and 18.

Table 15. Reconciliation of estimated CIF values across modes

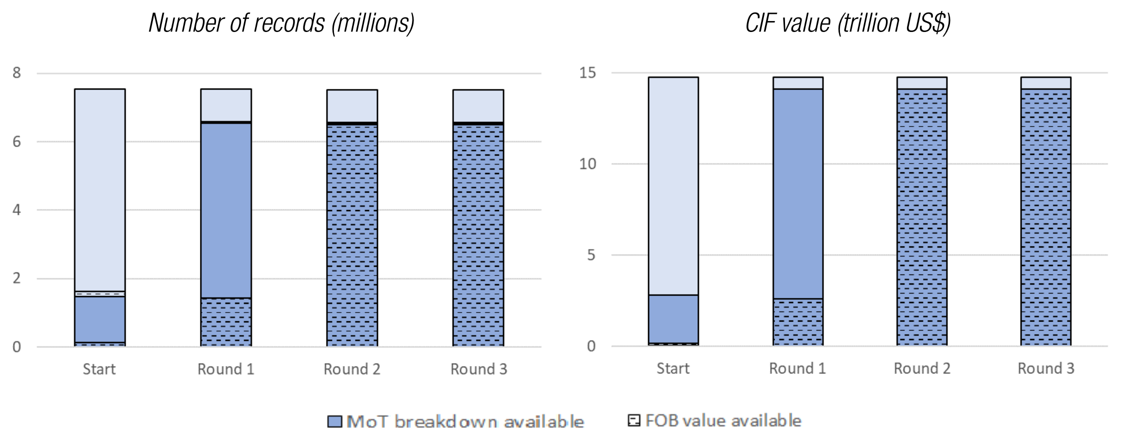
MoT	Non-reconciled				Reconciled			
	CIF value		Quantity		CIF value		Quantity	
	(1) Value	(2) Log value	(3) Value	(4) Log value	(5) Value	(6) Log value	(7) Value	(8) Log value
Air	26 865 736	17.11	35 955	10.49	74 865 624	18.13	37 749	10.54
Sea	14 320 232	16.48	270 769	12.51	39 905 593	17.50	284 277	12.56
Railway	-	/	-	/	-	/	-	/
Road	-	/	-	/	-	/	-	/
Other	-	/	-	/	-	/	-	/
Sum	*41 185 968*	*/*	*306 724*	*/*	*114 771 217*	*/*	*322 026*	*/*
Total	114 771 217	18.56	322 026	12.68	114 771 217	18.56	322 026	12.68

For this adjustment, the error standard deviation of the logarithmic entries in the other MoTs, representing actual observations, is assumed to be equal. Accordingly, the logarithmic values observed in each valid MoT is increased by an equal amount which exactly ensures that the sum of the non-logarithmic values remains the same as before the deletion of the entries for the invalid MoT. In the present example, the logarithmic CIF values were increased by 1.02 (compare columns 2 and 6) and the logarithmic quantities by 0.05 (compare columns 4 and 8). As a result, of the US$73 585 248 of trade previously recorded for "road", US$47 999 888 were re-allocated to "air" and US$25 585 361 to "sea".

7.3. Results

Figure 10 shows how data gaps were successively filled in each iteration round in which the data editing and imputation framework was run (see section 4.5). The main change in the *first round* consisted of the generation of new MoT breakdowns: for around 5.0 million records (representing US$6.2 trillion of imports, measured in terms of CIF) observations of CIF values were newly broken down by MoT with the help of the MoT model; and for almost all 1.3 million records (representing imports of US$2.6 trillion) for which MoT breakdowns were available in the source data without FOB values, these missing FOB values were generated. The main change in the *second round* consisted in the compilation of FOB values for the around 5 million newly generated MoT breakdowns of CIF values. In the *third round*, imputations for missing data on quantity were carried out, using the newly generated FOB values from the second round as input, while no further MoT breakdowns or observations in FOB values were added.

Figure 10. Data coverage after each round of editing and imputation



As a result of this process, most gaps in the primary data outlined in table 5 above were filled. Out of the 5.9 million records from UN Comtrade Plus in which breakdowns by MoT and FOB values had been missing (coverage type A), 938 thousand remained empty, accounting for US$659 billion, thus 5 per cent of the value of international merchandise trade, measured as CIF, recorded in the source data. This remaining gap is primarily attributable to the fact that for 30 relatively small reporting countries the MoT breakdown could not be generated.[6] Furthermore, out of the 1.3 million records for which the breakdown by MoT, but not the FOB value, was available in the source data (coverage type C), 16 thousand cases, accounting for US$8 billion of international trade, remained without FOB value after the imputations. And out of the 163 thousand records for which a FOB value but no MoT breakdown was initially available (coverage type B), 50 thousand cases, accounting for another US$8 billion of international trade, remained without MoT breakdown. The number of records containing both MoT breakdown and FOB value grew from 134 thousand to 6.5 million as a result of

[6] These countries comprise: Andorra; Bahrain; Bermuda; Botswana; Belize; Solomon Islands; Burundi; Cabo Verde; Costa Rica; Benin; El Salvador; Fiji; the State of Palestine; Kiribati; Greenland; Israel; Luxembourg; China, Macao SAR; Malawi; Mauritania; Mauritius; Nepal; Aruba; Niger; Palau; Rwanda; Saint Lucia; Swaziland; Burkina Faso; Samoa

the imputation procedures described above. They cover US$14.1 trillion, in terms of CIF value, thus 95 per cent of the international merchandise trade recorded in the UN Comtrade extraction and 87 per cent of global merchandise trade, according to the UNCTAD-WTO Merchandise Trade Dataset (UNCTAD, 2021b).

8. Concluding remarks

GTCDIT represents the first dataset which records transport costs alongside the corresponding bilateral trade data, detailed by commodity group and MoT. By deriving transport costs from aggregated data, as the difference between the CIF and the FOB value recorded in UN Comtrade Plus, we have broken new ground in the measurement of international transport costs. Our approach enables us to avoid aggregation problems, as they incur in the summing up transaction-level data in a bottom-up approach, and to ensure consistency with UN Comtrade, the standard global source for detailed merchandise trade statistics. This endeavour, however, has been faced with challenges, not all of which could be overcome during the first year of the project in which the beta version of the dataset was built. A main remaining challenge consists, for example, of increasing the accuracy of the breakdown of trade by MoT, taking into account transloading. Another consists of a more differentiated measurement of distance than in the current version. More imperfections may be discovered during the first uses of the beta version of GTCDIT in practice. Users are therefore advised to interpret the released data with care and take the mentioned deficiencies into account when drawing conclusions.

Despite these challenges, the database contains much potential to be enhanced further. The following activities, in particular, seem suited to considerably improve its coverage and accuracy:

a) The time coverage, currently limited to the year 2016, could be extended by adding more years, especially recent years, using available UN Comtrade Plus data. This would require only slight adjustments in the data editing and imputation framework presented above. An extended time coverage would not only increase the relevance of the data for research and policy analysis, it would also widen the opportunities for accurate imputations and data quality checks throughout the whole dataset.

b) The coverage with primary data can be increased by updating the database with new data released in UN Comtrade Plus. Over time, more and more countries are expected to report FOB values jointly with CIF values and/or provide breakdowns by MoT. The primary data coverage in these variables has been found to be much higher already for 2017 and subsequent years than for 2016.

c) The algorithms for breaking down trade by MoT could be refined, especially by making them capable of taking transloading into account. Transloading should be reflected in the UN Comtrade Plus data by a discrepancy between the trading partner and the country of consignment, both for exports and imports, and by specific customs procedure codes. These variables have not been used for the present version of GTCDIT, but the information they contain could easily be exploited in the future.

d) More use could be made of mirror data for the filling of gaps and accuracy checks, if cross-country discrepancies in the measurement and recording of trade were resolved, for instance, by comparison and reconciliation of the imports and exports reported among groups of trading partners connected by strong trade links.

e) The compilation of the distance variable could be improved by sourcing more information about the routes which goods actually travelled, for example, by analyzing microdata on the travels of individual airplanes, vessels, trains and trucks. These types of data could be sourced from international organisations specialized in transport, such as the International Maritime Organisation, the International Civil Airline Organization, the International Union of Railways, and from haulage companies. These microdata may also support the modelling of transhipment to enhance the MoT model, as mentioned under point (c) above.

f) The accuracy checks and imputations could be further enhanced by refining the underlying econometric models, and by running more iterations of data editing and imputation so that the quality of predictions would gain from successively enhanced input data.

g) Finally, the feedback obtained by users of the published beta version of GTCDIT could be reviewed and used to clean the dataset from any detected inaccuracies and to adapt the dataset according to users' needs.

The authors hope that these enhancements can be realized in the future, helping GTCDIT to be become a recognized, highly robust and reliable information source for studies on international transport and trade.

References

Anderson JE and Van Wincoop E (2004). Trade costs. *Journal of Economic Literature*. 42(3):691–751.

Berthelon ME and Freund C (2007). On the conservation of distance in international trade. *Journal of International Economics*. 75(2):310–320.

Bierlaire M (1998). Discrete Choice Models. *M Labbé, G Laporte, K Tanczos. and P Toint (Eds) Operations Research and Decision Aid Methodologies in Traffic and Transportation Management*. Springer. Berlin and Heidelberg.

Carrère C and Christopher G (2014). Can mirror data help to capture informal international trade?

Cascetta E (2009). Random Utility Theory. *Transport Systems Analysis*. Springer. Boston.

Deb K, Agrawal S, Pratap A and Meyarivan T (2000). A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II.

Fortanier F (2016). Towards merchandise trade statistics without asymmetries. *The Statistics Newsletter - OECD*. 64:7–11.

Gaulier G, Mirza D, Turban S and Zignago Z (2008). International Transportation Costs Around the World: a New CIF/FoB rates Dataset. CEPII.

Gaullier G and Zignago S (2010). BACI: International trade database at the product level. The 1994-2007 version. Document de Travail No. 23. CEPII.

Halim RA, Kirstein L, Merk O and Martinez L (2018). Decarbonization pathways for international maritime transport: A model-based policy impact assessment. *Sustainability*. 10(7):2243.

Hoffmann J, Pérez G and Wilmsmeier G (2002). International trade and transport profiles of Latin American countries, year 2000. Manuales No. 19. United Nations publication. Sales No. E.02.II.G.139, Santiago (Chile).

Hummels D (2001). *Toward a Geography of Trade Costs*. Purdue University.

Hummels D (2007). Transportation costs and international trade in the second era of globalization. *Journal of Economic Perspectives*. 21(3):131–154.

Hummels D and Lugovsky V (2006). Are Matched Partner Trade Statistics a Usable Measure of Transportation Costs? *Review of International Economics*. 14(1):69–86.

Hummels D, Lugovsky V and Skiba AM (2009). The trade reducing effects of market power in international shipping. *Journal of Development Economics*. 89(1):84–97.

International Chamber of Commerce (2019). *Incoterms 2020*.

Korinek J (2011). Clarifying trade costs in maritime transport.

Korinek J and Sourdin P (2009). Maritime transport costs and their impact on trade. Available at http://www.etsg.org/ETSG2009/papers/korinek.pdf (accessed 22 December 2021).

Krugman P (1980). Scale economies, product differentiation and the pattern of trade. *The American Economic Review*. 70(5):950–959.

Miao G and Fortanier F (2017). Estimating transport and insurance costs of international trade. OECD Statistics Working Papers No. STD/DOC(2017)4. OECD.

Redding S and Venables AJ (2004). Economic geography and international inequality. *Journal of International Economics*. 62(1):53–82.

Sanchez RJ et al. (2003). Port efficiency and international trade: Port efficiency as a determinant of maritime transport costs. *Maritime Economics and Logistics*. 5199–218.

Streicher G and Stehrer R (2013). Whither Panama? Constructing a consistent and balanced world SUT system including international trade and transport margins. WIIW Working Papers No. 94.

UNCTAD (2017). *Review of Maritime Transport 2017*. United Nations publication. Sales No. E.17.II.D.10E.19.II.D.20. New York and Geneva.

UNCTAD (2021a). *Review of Maritime Transport 2021*. United Nations publication. Sales No. E.21.11.D.21. New York and Geneva.

UNCTAD (2021b). UNCTADStat. Available at https://unctadstat.unctad.org/EN/Index.html (accessed 11 January 2021).

UNCTAD (2021c). UNCTADStat: Classifications. Available at https://unctadstat.unctad.org/EN/Classifications.html (accessed 10 June 2021).

UNCTAD (2022). *SDG Pulse 2021. UNCTAD Takes the Pulse of the SDGs*. United Nations publication. Sales No. E.21.II.D.19.

UNECE (2019). Generic Statistical Data Editing Model. GSDEM (Version 2.0, June 2019). Available at https://statswiki.unece.org/display/sde/GSDEM (accessed 22 December 2021).

UNSD (2011). International Merchandise Trade Statistics: Concepts and definitions 2010. ST/ESA/STAT/SER.M/52/Rev.3. United Nations publication. Sales No. E.10.XVII.13.

UNSD (2020). Standard country or area codes for statistical use (M49). Available at https://unstats.un.org/unsd/methodology/m49/ (accessed 28 May 2020).

UNSD (2021a). Methodology guide for UN Comtrade user on UN Comtrade Upgrade 2019. Available at https://comtrade.un.org/data/doc/UpgradePlan (accessed 22 December 2021).

UNSD (2021b). UN Comtrade Database. Available at https://comtrade.un.org/ (accessed 1 January 2021).

UNSD (2021c). UN Comtrade Plus. International Trade Statistics Database. Available at https://comtrade.un.org/ (accessed 22 October 2021).

World Customs Organization (2022). HS Nomenclature previous editions. Available at http://www.wcoomd.org/en/topics/nomenclature/instrument-and-tools/hs_nomenclature_previous_editions.aspx (accessed 28 February 2022).

Yeatz A, Amjadi A and Reincke U (1996). Did external barriers cause the marginalization of Sub-Saharan Africa in world trade?

Yi K-M (2003). Can vertical specialization explain the growth of world trade? *Journal of Political Economy*. 111(1):52–102.

Annex

Table 16. Data coverage in the UN Comtrade Plus extraction, by reporting country (importer)

		FOB value	
		No	Yes
MoT break-down	No	*Coverage type A:* Albania; Algeria; Andorra; Armenia; Aruba; Australia; Austria; Bahrain; Belarus; Belgium; Belize; Benin; Bermuda; Bosnia and Herzegovina; Botswana; Brazil; Brunei Darussalam; Burkina Faso; Burundi; Cabo Verde; Cambodia; Cameroon; Canada; Central African Republic; Chile; China (excluding Hong Kong, Macao and Taiwan); China, Macao SAR; Colombia; Costa Rica; Côte d'Ivoire; Croatia; Czechia; Denmark; Dominican Republic; Ecuador; El Salvador; Fiji; Finland including Åland Islands; France (including French Guiana, Guadeloupe, Martinique, Mayotte, Monaco and Reunion); Ghana; Greece; Greenland; Guatemala; Honduras; Iceland; India including Sikkim; Indonesia; Iran (Islamic Republic of); Ireland; Israel; Italy; Jamaica; Japan; Jordan; Kazakhstan; Kiribati; Korea, Republic of; Kyrgyzstan; Lao People's Democratic Republic; Lebanon; Lithuania; Luxembourg; Malawi; Malaysia; Mali; Malta; Mauritania; Mongolia; Morocco; Namibia; Nepal; Netherlands; New Zealand; Niger; Nigeria; Norway including Svalbard and Jan Mayen Islands excluding Bouvet Island; Pakistan; Palau; Panama; Paraguay; Poland; Qatar; Republic of Moldova; Romania; Russian Federation; Rwanda; Saint Lucia; Samoa; Saudi Arabia; Senegal; Serbia including Kosovo; Singapore; Solomon Islands; South Africa; Spain; Sri Lanka; State of Palestine; Swaziland; Switzerland including Liechtenstein; Tunisia; Turkey; Uganda; Ukraine; United Arab Emirates; United Kingdom (including Channel Islands and Isle of Man); United Republic of Tanzania; United States of America (including Puerto Rico and United States Virgin Islands); Uruguay; Viet Nam; Zimbabwe	*Coverage type B:* Argentina; Chile; Mauritius
	Yes	*Coverage type C:* Azerbaijan; Bulgaria; China, Hong Kong SAR; Cyprus; Estonia; Germany; Hungary; Kuwait; Latvia; Madagascar; Mexico; Montenegro; Nicaragua; North Macedonia; Oman; Portugal; Slovakia; Slovenia; Sweden; Thailand; Togo	*Coverage type D:* Angola; Madagascar; Peru

Table 17. Mode of transport model estimation results

Commodity group (HS code)	Coefficients								R^2	N
	Air	Sea	Rail	Distance (1000 km)	Time (10 h)	Cont_RL	Cont_RD	Trade		
01	-2.290	-2.310	0.040	-0.200	-0.030	3.500	4.430	4.000	0.988	1500
02	-2.910	-0.250	0.050	-0.043	-0.013	3.460	0.260	3.270	0.983	1692
03	-1.580	0.100	-4.950	-0.012	-0.003	2.930	4.980	2.310	0.934	2860
04	-1.580	0.420	-5.000	-0.015	-0.001	4.980	1.410	3.620	0.951	2220
05	-0.420	0.480	-4.990	0.000	0.000	0.000	3.480	3.050	0.943	1796
06	-4.130	-3.710	-4.230	0.000	0.000	2.670	4.370	2.460	0.938	2064
07	0.240	1.910	0.140	-0.086	0.000	4.580	1.090	1.720	0.940	2820
08	-2.130	0.250	0.040	-0.103	-0.001	0.490	3.250	1.480	0.950	3272
09	0.010	2.320	-1.970	-0.045	-0.008	2.400	4.950	3.920	0.947	3392
10	-5.000	3.230	1.000	-0.189	-0.010	0.870	3.700	1.380	0.980	2228
11	-4.210	3.250	1.000	-0.043	-0.008	2.420	3.270	3.990	0.985	2308
12	-0.480	0.800	-4.370	0.000	-0.005	2.040	3.030	3.460	0.874	3220
13	-2.930	-1.390	-4.330	-0.171	-0.006	4.610	1.510	1.810	0.950	1864
14	0.700	4.820	-4.020	-0.104	-0.073	0.270	5.000	2.940	0.977	1124
15	-2.130	-0.060	-0.980	-0.070	-0.004	2.770	2.890	1.290	0.942	2928
16	-2.490	0.180	-3.110	-0.102	-0.025	2.110	3.740	2.850	0.969	2316
17	-2.140	0.430	0.300	0.000	-0.017	3.380	3.870	2.390	0.942	2744
18	-2.600	-0.670	-0.290	-0.061	0.000	0.140	4.930	1.290	0.950	2692
19	-4.420	3.370	-1.860	-0.045	-0.005	3.670	3.850	2.630	0.973	3120
20	-0.420	4.510	0.810	-0.169	-0.006	2.170	4.330	2.280	0.967	3280
21	-1.700	-0.140	-4.960	-0.021	-0.002	2.040	4.970	3.650	0.906	3700
22	-0.300	3.130	-4.170	-0.121	-0.013	0.010	4.910	2.260	0.951	3756
23	-3.540	-0.600	-4.200	-0.047	-0.011	3.970	4.160	3.540	0.965	2284
24	-2.120	-0.320	-4.320	-0.026	-0.005	0.890	3.510	1.830	0.944	2000
25	-3.210	4.370	0.950	-0.007	-0.007	4.900	3.830	-0.610	0.930	3080
26	-1.390	4.750	-3.030	-0.082	-0.042	4.660	0.070	2.980	0.972	1500
27	-4.860	4.720	-1.830	-0.140	-0.028	2.270	2.310	2.660	0.894	3136
28	-3.800	-1.780	-3.400	-0.030	-0.008	1.610	4.900	1.010	0.910	3320
29	0.250	1.320	0.540	-0.020	-0.001	1.230	4.160	2.150	0.840	3356
30	-4.340	-4.630	-4.920	0.000	-0.004	0.790	0.860	-0.980	0.869	3968
31	-3.020	1.780	-0.630	-0.176	-0.004	1.960	4.870	1.120	0.955	2092
32	-2.740	-0.820	-3.750	-0.003	-0.024	1.930	3.320	2.260	0.905	3568
33	-3.150	-1.760	-4.970	0.000	-0.022	0.970	3.240	2.770	0.883	4224
34	-1.760	-0.060	-3.940	-0.002	-0.012	2.370	0.340	3.250	0.907	3820
35	-2.750	-1.300	-2.540	-0.006	-0.017	2.030	1.430	-0.220	0.899	2836
36	-1.320	-0.350	-3.420	-0.001	-0.002	2.840	4.230	3.380	0.927	1236
37	-2.590	-1.500	-4.860	-0.017	-0.003	1.710	4.060	2.130	0.944	1576
38	-2.250	-0.870	-3.650	-0.064	-0.010	1.040	3.060	2.630	0.865	3980
39	-2.200	-0.710	-2.790	-0.003	-0.005	2.170	2.410	2.170	0.879	5584
40	-3.140	-1.800	-3.220	-0.008	-0.008	1.760	3.390	2.550	0.853	4672

Commodity group (HS code)	Coefficients								R^2	N
	Air	Sea	Rail	Distance (1000 km)	Time (10 h)	Cont_RL	Cont_RD	Trade		
41	-2.810	-1.860	-4.820	-0.098	-0.007	1.630	3.040	1.010	0.911	2356
42	-4.470	-4.360	-4.990	-0.004	-0.011	0.170	0.300	0.580	0.843	4480
43	-0.720	-0.410	-2.350	-0.027	-0.016	3.230	5.000	0.400	0.907	1500
44	0.020	2.180	0.700	-0.028	-0.010	2.050	4.490	1.210	0.917	4244
45	-0.790	0.700	-1.210	0.000	-0.023	0.060	4.040	1.520	0.924	1288
46	0.070	1.490	-2.560	0.000	-0.007	0.990	4.430	3.160	0.960	1612
47	-3.870	-0.310	-0.070	-0.044	-0.012	0.800	4.920	-0.310	0.962	1324
48	-1.420	0.230	-1.740	-0.005	-0.011	0.550	4.770	1.690	0.869	4508
49	-1.220	-1.110	-3.230	-0.014	-0.007	1.090	4.850	-0.200	0.826	4632
50	-0.640	-0.570	-3.070	0.000	0.000	1.150	4.340	-0.400	0.951	944
51	-2.760	-1.720	-4.680	-0.002	-0.002	1.920	0.850	0.600	0.931	1732
52	-1.470	0.040	-4.730	-0.022	-0.007	0.530	4.720	2.160	0.911	2852
53	0.120	1.070	-0.030	0.000	-0.008	1.250	4.040	2.120	0.927	1608
54	-4.040	-2.440	-4.680	-0.007	-0.013	1.540	4.820	0.280	0.908	2760
55	-4.220	-2.190	-4.250	-0.001	-0.013	1.110	2.820	1.070	0.920	2580
56	-3.270	-1.550	-4.780	-0.027	-0.019	2.230	3.260	1.570	0.900	3180
57	-0.780	0.360	-3.850	-0.049	-0.011	0.680	4.680	2.410	0.903	2820
58	-0.610	-0.370	-3.690	-0.004	-0.005	0.240	3.650	1.900	0.882	2880
59	-2.330	-1.360	-2.910	-0.002	-0.009	0.830	3.860	2.640	0.887	2920
60	-2.910	-1.540	-3.100	-0.003	-0.015	0.040	4.770	0.840	0.918	2248
61	-2.650	-2.500	-4.180	-0.003	-0.010	0.270	4.910	-0.090	0.842	5216
62	-0.320	-0.300	-3.110	-0.003	-0.006	1.780	1.830	1.650	0.838	5364
63	-0.230	0.620	-4.670	-0.007	-0.007	0.690	3.960	3.540	0.835	4736
64	-2.920	-2.160	-4.980	-0.012	-0.011	1.510	2.470	1.590	0.871	4084
65	-2.640	-2.340	-4.800	-0.002	-0.004	2.190	4.900	-0.470	0.879	3400
66	0.600	2.180	-1.350	0.000	-0.032	1.680	2.300	1.900	0.937	1792
67	-2.780	-2.350	-4.780	-0.001	-0.018	0.040	4.340	0.150	0.912	1848
68	-3.900	-2.570	-4.910	-0.004	-0.005	1.590	3.180	0.250	0.878	3676
69	-3.760	-2.420	-4.900	-0.015	-0.014	0.360	0.090	2.290	0.884	3804
70	-2.100	-0.970	-4.310	-0.003	-0.009	1.290	3.270	3.860	0.863	4208
71	0.410	-0.440	-4.470	-0.022	-0.008	1.720	3.130	0.620	0.879	3676
72	-4.820	3.330	-0.070	-0.074	-0.016	2.580	3.040	3.080	0.910	3276
73	-2.800	-1.520	-3.440	-0.020	-0.013	1.980	3.710	1.870	0.841	5172
74	-3.370	-1.760	-3.730	0.000	-0.021	0.100	3.140	1.700	0.878	3372
75	-1.880	-0.820	-4.180	0.000	-0.025	2.740	4.830	3.000	0.910	1520
76	-2.600	-1.220	-3.200	-0.024	-0.009	1.800	2.690	2.980	0.884	3912
78	-4.990	-3.780	-4.980	-0.072	-0.004	1.290	4.270	3.530	0.969	1252
79	-3.220	-1.500	0.190	-0.001	-0.013	0.030	2.380	-0.890	0.954	1600
80	-3.890	-2.710	-0.070	-0.058	-0.016	1.420	4.740	2.470	0.926	1032
81	-3.590	-3.230	-4.640	0.000	-0.004	0.270	3.860	1.310	0.879	1692
82	-2.190	-1.810	-4.920	-0.006	-0.008	0.090	3.970	2.090	0.806	4420
83	-1.760	-0.740	-3.330	-0.001	-0.015	4.650	3.250	-0.780	0.853	4256
84	0.390	1.090	-1.660	-0.005	-0.007	2.300	2.430	0.600	0.818	6092

Commodity group (HS code)	Coefficients								R^2	N
	Air	Sea	Rail	Distance (1000 km)	Time (10 h)	Cont_RL	Cont_RD	Trade		
85	-2.040	-1.950	-4.070	-0.018	-0.009	0.490	2.680	1.960	0.800	6232
86	-3.810	-2.050	-4.740	-0.074	-0.014	3.950	4.980	3.850	0.907	1880
87	-3.970	-2.180	-4.730	-0.002	-0.012	2.060	3.880	2.160	0.878	4592
88	0.440	-0.070	-3.350	-0.032	-0.017	2.100	2.600	2.590	0.891	2184
89	-1.030	2.440	-1.690	0.000	-0.048	1.170	0.340	2.560	0.964	1600
90	-1.160	-1.170	-3.020	-0.013	-0.034	1.320	4.910	0.540	0.822	5148
91	-1.640	-2.460	-4.200	-0.025	-0.007	0.110	4.860	-0.580	0.907	3208
92	-1.020	-0.440	-4.840	-0.002	-0.006	0.260	2.550	2.420	0.876	2228
93	-1.190	-0.640	-4.680	-0.002	-0.015	2.340	0.660	3.360	0.885	1240
94	-0.440	0.640	-3.320	-0.001	-0.003	4.270	2.870	3.000	0.876	4676
95	-0.070	1.170	-0.240	-0.012	-0.019	1.740	4.660	2.420	0.858	3884
96	-1.140	0.460	-3.770	-0.004	-0.027	0.380	0.790	2.610	0.872	4104
97	-3.770	-4.100	-4.920	0.000	0.000	2.440	4.890	2.170	0.887	1792

Table 18. Transport cost model, estimation results

Commodity group (HS 2012)	Mode of transport	Coefficients Distance (km)	CIF value (US$)	N	R^2
01	Air	-	0.82 (0.03)	80	0.92
01	Sea	-	0.98 (0.05)	83	0.84
01	Railway	-	0.83 (0.08)	19	0.86
01	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
02	Air	-	0.88 (0.03)	102	0.91
02	Sea	0.24 (0.11)	1.02 (0.02)	409	0.91
02	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
03	Air	0.6 (0.06)	0.93 (0.01)	159	0.97
03	Sea	-	0.98 (0.02)	366	0.88
03	Road	1.53 (0.43)	1.04 (0.05)	59	0.92
04	Air	0.24 (0.09)	0.96 (0.01)	156	0.97
04	Sea	0.32 (0.09)	0.94 (0.01)	414	0.93
04	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
05	Air	-	0.92 (0.03)	142	0.84
05	Sea	-	0.95 (0.03)	47	0.95
05	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
06	Air	0.3 (0.13)	0.92 (0.03)	99	0.92
06	Sea	-	0.93 (0.03)	42	0.95
06	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
07	Air	-0.24 (0.11)	0.98 (0.02)	173	0.95
07	Sea	-0.3 (0.09)	1.02 (0.01)	417	0.93
07	Railway	-	0.83 (0.08)	19	0.86
07	Road	-	0.97 (0.03)	59	0.94
08	Air	-	1.04 (0.02)	164	0.97
08	Sea	-	1 (0.02)	369	0.9
08	Road	0.73 (0.29)	0.96 (0.04)	57	0.94
09	Air	-	0.86 (0.03)	132	0.88
09	Sea	0.21 (0.09)	0.94 (0.02)	478	0.89
09	Road	0.83 (0.35)	1.09 (0.04)	42	0.95
10	Air	-	0.76 (0.06)	43	0.77
10	Sea	-	0.99 (0.02)	140	0.96
10	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
11	Air	-	0.97 (0.04)	65	0.9
11	Sea	0.2 (0.1)	1.03 (0.01)	336	0.96
11	Road	-	1.09 (0.04)	31	0.96
12	Air	-	0.75 (0.02)	168	0.87
12	Sea	-	0.96 (0.02)	252	0.92
12	Road	-	0.9 (0.03)	42	0.94
13	Air	0.53 (0.19)	0.9 (0.05)	62	0.86
13	Sea	-	0.92 (0.03)	110	0.88

Commodity group (HS 2012)	Mode of transport	Coefficients			
		Distance (km)	CIF value (US$)	N	R^2
13	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
14	Air	-	0.92 (0.03)	142	0.84
14	Sea	-	0.98 (0.05)	83	0.84
14	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
15	Air	-0.42 (0.18)	0.95 (0.03)	136	0.88
15	Sea	-	0.99 (0.01)	439	0.95
15	Road	-	1.01 (0.07)	35	0.87
16	Air	-	0.88 (0.02)	124	0.93
16	Sea	-	0.95 (0.01)	323	0.93
16	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
17	Air	-	0.91 (0.04)	101	0.87
17	Sea	0.25 (0.09)	0.99 (0.01)	295	0.95
17	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
18	Air	0.4 (0.13)	0.94 (0.03)	88	0.94
18	Sea	0.35 (0.1)	0.9 (0.02)	227	0.9
18	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
19	Air	0.34 (0.09)	0.94 (0.01)	180	0.96
19	Sea	0.23 (0.07)	0.98 (0.01)	589	0.94
19	Road	-	0.9 (0.03)	55	0.93
20	Air	0.2 (0.09)	0.96 (0.02)	204	0.93
20	Sea	0.12 (0.05)	0.98 (0.01)	904	0.92
20	Road	0.28 (0.11)	1.04 (0.02)	76	0.96
21	Air	-	0.85 (0.02)	223	0.92
21	Sea	0.24 (0.07)	0.94 (0.01)	511	0.94
21	Road	0.63 (0.25)	0.96 (0.05)	45	0.92
22	Air	-	0.93 (0.03)	170	0.84
22	Sea	-	0.99 (0.01)	517	0.94
22	Road	-	1 (0.04)	45	0.94
23	Air	-	0.83 (0.04)	51	0.9
23	Sea	-	0.93 (0.02)	167	0.94
23	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
24	Air	-	0.92 (0.03)	142	0.84
24	Sea	-	0.96 (0.06)	56	0.83
24	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
25	Air	-	1.08 (0.03)	151	0.87
25	Sea	-	1.09 (0.01)	628	0.94
25	Road	-	1.06 (0.04)	53	0.93
26	Air	-0.57 (0.26)	1.21 (0.11)	32	0.83
26	Sea	1.19 (0.23)	0.97 (0.06)	60	0.84
26	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
27	Air	-	1.04 (0.03)	204	0.87
27	Sea	-	1.02 (0.01)	427	0.95
27	Road	0.52 (0.14)	1.03 (0.04)	53	0.94

Commodity group (HS 2012)	Mode of transport	Coefficients Distance (km)	CIF value (US$)	N	R^2
28	Air	-0.16 (0.07)	0.98 (0.01)	999	0.82
28	Sea	-	1.05 (0.01)	1616	0.9
28	Road	-	1.11 (0.05)	88	0.87
29	Air	-0.34 (0.06)	0.93 (0.01)	1786	0.81
29	Sea	-	1 (0.01)	2291	0.89
29	Road	-	1.06 (0.05)	61	0.88
30	Air	-	0.86 (0.01)	842	0.91
30	Sea	0.21 (0.07)	0.9 (0.01)	644	0.92
30	Road	0.6 (0.16)	0.94 (0.05)	46	0.91
31	Air	-	0.99 (0.05)	75	0.86
31	Sea	0.34 (0.09)	1.04 (0.02)	303	0.94
31	Road	-	1.12 (0.07)	37	0.89
32	Air	-	0.93 (0.01)	830	0.85
32	Sea	-	0.95 (0.01)	1363	0.9
32	Road	0.34 (0.15)	0.95 (0.06)	99	0.76
33	Air	-	0.89 (0.01)	659	0.93
33	Sea	-	0.92 (0.01)	1071	0.91
33	Road	-	0.85 (0.04)	99	0.8
34	Air	-0.18 (0.09)	1 (0.02)	552	0.87
34	Sea	-	0.97 (0.01)	1045	0.93
34	Railway	-	0.83 (0.08)	19	0.86
34	Road	-	1.02 (0.05)	85	0.85
35	Air	-	0.99 (0.02)	313	0.87
35	Sea	0.18 (0.09)	0.93 (0.01)	400	0.91
35	Road	-	0.98 (0.08)	36	0.82
36	Air	-	0.93 (0.04)	39	0.95
36	Sea	-	1 (0.03)	76	0.93
36	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
37	Air	-0.37 (0.18)	1 (0.04)	131	0.83
37	Sea	-	0.89 (0.02)	167	0.91
37	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
38	Air	-	0.9 (0.01)	850	0.87
38	Sea	0.18 (0.05)	0.96 (0.01)	1397	0.92
38	Road	-	0.96 (0.04)	104	0.87
39	Air	-	0.99 (0.01)	2792	0.88
39	Sea	-	0.97 (0)	4336	0.93
39	Road	0.12 (0.06)	0.99 (0.02)	337	0.86
40	Air	-	0.95 (0.01)	2140	0.89
40	Sea	-	0.98 (0.01)	2568	0.92
40	Railway	-	0.83 (0.08)	19	0.86
40	Road	-	0.97 (0.03)	203	0.88
41	Air	-	0.92 (0.04)	93	0.86
41	Sea	-	0.79 (0.04)	74	0.86

Commodity group (HS 2012)	Mode of transport	Coefficients			
		Distance (km)	CIF value (US$)	N	R^2
41	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
42	Air	-	0.95 (0.01)	757	0.86
42	Sea	-	0.94 (0.01)	787	0.87
42	Railway	-	0.83 (0.08)	19	0.86
42	Road	-	1.12 (0.06)	74	0.85
43	Air	-	0.92 (0.03)	142	0.84
43	Sea	-	0.98 (0.05)	83	0.84
43	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
44	Air	-	1.04 (0.02)	324	0.86
44	Sea	0.23 (0.06)	0.98 (0.01)	1008	0.92
44	Road	-	1.07 (0.03)	94	0.91
45	Air	-	1.09 (0.06)	55	0.88
45	Sea	-	1.11 (0.05)	56	0.91
45	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
46	Air	-	0.95 (0.07)	40	0.83
46	Sea	-	1 (0.03)	116	0.91
46	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
47	Air	-	0.92 (0.03)	142	0.84
47	Sea	-	0.98 (0.03)	44	0.95
48	Air	-	1.02 (0.01)	1493	0.88
48	Sea	0.24 (0.03)	1 (0.01)	2639	0.94
48	Railway	-	0.83 (0.08)	19	0.86
48	Road	-	0.99 (0.03)	176	0.88
49	Air	-	0.96 (0.01)	684	0.89
49	Sea	0.16 (0.07)	0.91 (0.01)	706	0.91
49	Road	-	1.03 (0.07)	45	0.83
50	Air	-	0.93 (0.05)	45	0.88
50	Sea	-	0.98 (0.05)	83	0.84
50	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
51	Air	-	0.9 (0.03)	98	0.91
51	Sea	0.51 (0.23)	0.82 (0.03)	88	0.88
51	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
52	Air	-	0.95 (0.02)	456	0.86
52	Sea	0.22 (0.06)	0.9 (0.01)	767	0.89
52	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
53	Air	-	0.99 (0.04)	68	0.9
53	Sea	-	0.91 (0.05)	101	0.78
54	Air	0.22 (0.1)	0.95 (0.02)	358	0.86
54	Sea	0.23 (0.07)	0.96 (0.01)	594	0.9
54	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
55	Air	-	0.97 (0.02)	303	0.87
55	Sea	0.23 (0.08)	0.92 (0.01)	591	0.91
55	Road	0.27 (0.06)	0.99 (0.02)	557	0.89

Commodity group (HS 2012)	Mode of transport	Coefficients Distance (km)	CIF value (US$)	N	R^2
56	Air	-	1.03 (0.02)	419	0.88
56	Sea	0.15 (0.06)	0.99 (0.01)	661	0.92
56	Road	0.32 (0.14)	0.96 (0.05)	61	0.87
57	Air	-	0.93 (0.03)	167	0.88
57	Sea	-	0.89 (0.02)	375	0.89
57	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
58	Air	-	0.92 (0.02)	380	0.86
58	Sea	0.23 (0.12)	0.89 (0.02)	379	0.85
58	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
59	Air	-	0.97 (0.02)	357	0.89
59	Sea	-	0.9 (0.02)	420	0.89
59	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
60	Air	-	1.02 (0.03)	144	0.88
60	Sea	-	0.95 (0.02)	268	0.89
60	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
61	Air	0.16 (0.03)	0.93 (0.01)	2198	0.88
61	Sea	-	0.88 (0.01)	2382	0.86
61	Road	0.36 (0.08)	1 (0.03)	229	0.87
62	Air	0.15 (0.03)	0.93 (0.01)	2502	0.87
62	Sea	-0.2 (0.04)	0.91 (0.01)	2800	0.83
62	Road	0.49 (0.07)	0.92 (0.03)	232	0.86
63	Air	-	0.97 (0.01)	893	0.87
63	Sea	-	0.94 (0.01)	1465	0.9
63	Road	0.61 (0.1)	1.03 (0.03)	145	0.87
64	Air	-	0.92 (0.02)	570	0.86
64	Sea	0.19 (0.08)	0.91 (0.01)	674	0.88
64	Road	0.53 (0.15)	0.92 (0.04)	64	0.89
65	Air	-	0.96 (0.02)	235	0.87
65	Sea	-	0.98 (0.02)	285	0.91
65	Road	0.47 (0.22)	1.03 (0.08)	33	0.85
66	Air	-	0.92 (0.06)	53	0.83
66	Sea	-	0.94 (0.03)	127	0.9
66	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
67	Air	-	0.91 (0.04)	87	0.88
67	Sea	-	1 (0.03)	105	0.94
67	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
68	Air	-	0.96 (0.02)	637	0.82
68	Sea	-	1.04 (0.01)	1069	0.91
68	Railway	-	0.83 (0.08)	19	0.86
68	Road	-	0.99 (0.06)	88	0.78
69	Air	-	0.94 (0.02)	384	0.81
69	Sea	0.15 (0.07)	0.93 (0.01)	784	0.9
69	Road	0.3 (0.11)	1.06 (0.04)	55	0.93

Commodity group (HS 2012)	Mode of transport	Coefficients Distance (km)	CIF value (US$)	N	R^2
70	Air	-	1.01 (0.01)	1094	0.85
70	Sea	0.11 (0.05)	0.99 (0.01)	1580	0.91
70	Road	-	1.08 (0.04)	100	0.86
71	Air	-	0.82 (0.02)	231	0.83
71	Sea	-	0.93 (0.03)	189	0.84
71	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
72	Air	-	1.09 (0.02)	502	0.83
72	Sea	0.14 (0.04)	1.01 (0.01)	1861	0.92
72	Road	-	0.99 (0.04)	124	0.83
73	Air	-0.1 (0.03)	0.98 (0.01)	3840	0.88
73	Sea	0.1 (0.03)	0.98 (0)	4850	0.91
73	Railway	-	0.83 (0.08)	19	0.86
73	Road	0.17 (0.05)	0.97 (0.02)	520	0.85
74	Air	-	0.99 (0.02)	543	0.83
74	Sea	0.17 (0.09)	0.94 (0.01)	698	0.86
74	Road	-	0.86 (0.06)	56	0.81
75	Air	-	0.79 (0.05)	83	0.76
75	Sea	-	0.94 (0.04)	78	0.89
75	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
76	Air	-	1.02 (0.02)	611	0.86
76	Sea	0.24 (0.06)	0.95 (0.01)	967	0.9
76	Road	-	1.05 (0.06)	69	0.82
78	Air	-	0.92 (0.03)	142	0.84
78	Sea	-	0.88 (0.03)	55	0.94
78	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
79	Air	-	1.02 (0.06)	54	0.83
79	Sea	-	0.96 (0.04)	93	0.86
79	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
80	Air	-	0.92 (0.03)	142	0.84
80	Sea	-	0.98 (0.04)	35	0.95
80	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
81	Air	-	0.93 (0.05)	83	0.83
81	Sea	-	0.95 (0.05)	72	0.86
81	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
82	Air	-	0.93 (0.01)	2034	0.87
82	Sea	0.13 (0.04)	0.91 (0.01)	2639	0.89
82	Road	-0.19 (0.08)	0.98 (0.03)	304	0.74
83	Air	-	0.95 (0.01)	1430	0.87
83	Sea	0.13 (0.05)	0.95 (0.01)	1584	0.91
83	Road	-	1.01 (0.05)	115	0.76
84	Air	-0.07 (0.02)	0.9 (0)	12028	0.86
84	Sea	0 (0)	0.93 (0)	14398	0.89
84	Railway	-	0.83 (0.08)	19	0.86

Commodity group (HS 2012)	Mode of transport	Coefficients Distance (km)	CIF value (US$)	N	R^2
84	Road	-	0.92 (0.01)	1365	0.81
85	Air	-0.1 (0.02)	0.9 (0)	10258	0.86
85	Sea	-	0.93 (0)	9674	0.89
85	Railway	-	0.83 (0.08)	19	0.86
85	Road	-	0.92 (0.02)	881	0.74
86	Air	-	0.87 (0.06)	72	0.78
86	Sea	-	0.98 (0.04)	187	0.81
86	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
87	Air	0.25 (0.04)	0.97 (0.01)	1630	0.9
87	Sea	-	0.99 (0.01)	2536	0.92
87	Railway	-	0.83 (0.08)	19	0.86
87	Road	-	0.95 (0.03)	221	0.87
88	Air	-	0.8 (0.03)	124	0.82
88	Sea	-	0.89 (0.05)	40	0.91
88	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
89	Air	-	0.99 (0.06)	65	0.83
89	Sea	-	0.88 (0.03)	181	0.83
89	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
90	Air	-0.07 (0.03)	0.86 (0)	4704	0.86
90	Sea	-	0.91 (0.01)	3425	0.88
90	Railway	-	0.83 (0.08)	19	0.86
90	Road	-	0.87 (0.03)	292	0.74
91	Air	-	0.86 (0.02)	422	0.85
91	Sea	-	0.89 (0.02)	403	0.85
91	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
92	Air	-	0.95 (0.03)	136	0.86
92	Sea	0.25 (0.1)	0.85 (0.02)	299	0.87
92	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
93	Air	0.53 (0.19)	0.84 (0.03)	96	0.88
93	Sea	-	1.01 (0.03)	65	0.95
93	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
94	Air	-	0.94 (0.01)	1075	0.86
94	Sea	0.31 (0.04)	0.93 (0.01)	2210	0.91
94	Railway	-	0.83 (0.08)	19	0.86
94	Road	0.29 (0.09)	1.01 (0.03)	168	0.89
95	Air	-	0.89 (0.01)	506	0.88
95	Sea	0.16 (0.07)	0.92 (0.01)	836	0.9
95	Road	-	0.97 (0.04)	60	0.92
96	Air	-	0.93 (0.01)	1005	0.88
96	Sea	-	0.94 (0.01)	1435	0.89
96	Railway	-	0.83 (0.08)	19	0.86
96	Road	-	1.08 (0.04)	135	0.84
97	Air	0.31 (0.16)	0.83 (0.05)	51	0.83

Commodity group (HS 2012)	Mode of transport	Coefficients		N	R^2
		Distance (km)	CIF value (US$)		
97	Sea	-	0.9 (0.04)	89	0.88
97	Road	0.27 (0.06)	0.99 (0.02)	557	0.89
99	Air	-	0.75 (0.04)	103	0.79
99	Sea	-	0.9 (0.02)	130	0.92
99	Road	0.27 (0.06)	0.99 (0.02)	557	0.89

Note: Standard error in parentheses.

Table 19. Quantity-value model, estimation results

Commodity group (HS 2012)	Mode of transport	Coefficients			
		Distance	CIF value	N	R^2
01	Air		0.48 (0.07)	74	0.41
01	Sea		0.74 (0.06)	83	0.68
01	Railway			2	
01	Road		0.71 (0.02)	528	0.68
02	Air		0.91 (0.03)	102	0.87
02	Sea	-0.33 (0.12)	0.77 (0.02)	407	0.86
02	Road		0.71 (0.02)	528	0.68
03	Air	-0.36 (0.16)	0.88 (0.03)	158	0.85
03	Sea		0.78 (0.02)	372	0.88
03	Road		0.93 (0.04)	62	0.88
04	Air		0.94 (0.03)	156	0.9
04	Sea		0.89 (0.01)	412	0.91
04	Road	-0.49 (0.22)	0.88 (0.05)	34	0.92
05	Air		0.36 (0.06)	169	0.2
05	Sea		0.83 (0.04)	47	0.9
05	Road		0.71 (0.02)	528	0.68
06	Air	-0.65 (0.26)	0.69 (0.05)	99	0.72
06	Sea		0.72 (0.06)	42	0.85
06	Road		0.71 (0.02)	528	0.68
07	Air		0.84 (0.03)	173	0.79
07	Sea		0.85 (0.01)	420	0.89
07	Railway			1	
07	Road	-0.77 (0.26)	0.86 (0.04)	64	0.9
08	Air		0.77 (0.02)	164	0.87
08	Sea		0.82 (0.02)	371	0.88
08	Road		0.94 (0.03)	58	0.93
09	Air		0.97 (0.05)	132	0.77
09	Sea		0.89 (0.02)	480	0.84
09	Road	-1.31 (0.52)	0.81 (0.05)	42	0.86
10	Air		0.86 (0.09)	42	0.71
10	Sea		0.82 (0.02)	141	0.93
10	Road		0.71 (0.02)	528	0.68
11	Air		0.8 (0.05)	62	0.82
11	Sea		0.86 (0.01)	332	0.95
11	Road		0.89 (0.09)	31	0.77
12	Air		0.94 (0.06)	168	0.59
12	Sea	-0.38 (0.18)	0.74 (0.03)	252	0.77
12	Road		0.81 (0.05)	43	0.89
13	Air		0.8 (0.05)	62	0.81
13	Sea		0.87 (0.04)	110	0.8

Commodity group (HS 2012)	Mode of transport	Coefficients			
		Distance	CIF value	N	R^2
13	Road		0.71 (0.02)	528	0.68
14	Air		0.36 (0.06)	169	0.2
14	Sea		0.74 (0.06)	83	0.68
14	Road		0.71 (0.02)	528	0.68
15	Air	0.63 (0.21)	0.91 (0.03)	133	0.84
15	Sea		0.82 (0.01)	441	0.91
15	Road		0.83 (0.05)	35	0.89
16	Air		1.01 (0.03)	124	0.9
16	Sea		0.88 (0.01)	325	0.93
16	Road		0.71 (0.02)	528	0.68
17	Air		0.81 (0.05)	99	0.76
17	Sea		0.82 (0.01)	294	0.93
17	Road		0.71 (0.02)	528	0.68
18	Air		1.01 (0.03)	88	0.92
18	Sea		0.92 (0.02)	226	0.89
18	Road		0.71 (0.02)	528	0.68
19	Air		0.99 (0.03)	180	0.89
19	Sea		0.91 (0.01)	589	0.92
19	Road	-0.74 (0.19)	0.94 (0.04)	57	0.92
20	Air		0.95 (0.03)	203	0.82
20	Sea		0.84 (0.01)	910	0.89
20	Road	-0.67 (0.17)	0.83 (0.04)	76	0.89
21	Air		1.01 (0.03)	223	0.79
21	Sea		0.93 (0.01)	511	0.9
21	Road	-1.71 (0.26)	0.91 (0.05)	45	0.91
22	Air		0.78 (0.04)	167	0.66
22	Sea		0.83 (0.01)	517	0.88
22	Road		0.8 (0.05)	45	0.86
23	Air		0.93 (0.06)	50	0.81
23	Sea	0.34 (0.15)	0.82 (0.03)	164	0.87
23	Road		0.71 (0.02)	528	0.68
24	Air		0.36 (0.06)	169	0.2
24	Sea		0.83 (0.06)	57	0.77
24	Road		0.71 (0.02)	528	0.68
25	Air		0.63 (0.06)	145	0.45
25	Sea		0.69 (0.01)	625	0.87
25	Road		0.58 (0.04)	53	0.78
26	Air		0.36 (0.06)	169	0.2
26	Sea	-0.77 (0.27)	0.64 (0.05)	59	0.75
26	Road		0.71 (0.02)	528	0.68
27	Air	0.71 (0.26)	0.66 (0.04)	194	0.61
27	Sea	0.22 (0.11)	0.78 (0.01)	419	0.92
27	Road		0.74 (0.06)	53	0.78

Commodity group (HS 2012)	Mode of transport	Coefficients: Distance	Coefficients: CIF value	N	R^2
28	Air	0.4 (0.12)	0.54 (0.02)	920	0.5
28	Sea		0.63 (0.01)	1563	0.8
28	Road		0.6 (0.04)	84	0.75
29	Air	1.21 (0.1)	0.56 (0.01)	1667	0.53
29	Sea	0.39 (0.05)	0.7 (0.01)	2272	0.81
29	Road		0.75 (0.07)	60	0.68
30	Air	0.43 (0.1)	0.92 (0.02)	842	0.75
30	Sea		0.9 (0.02)	644	0.82
30	Road		0.97 (0.09)	46	0.75
31	Air		0.57 (0.07)	69	0.52
31	Sea		0.76 (0.01)	297	0.91
31	Road		0.66 (0.05)	34	0.86
32	Air	0.3 (0.09)	0.78 (0.02)	828	0.73
32	Sea	0 (0)	0.82 (0.01)	1367	0.86
32	Road		0.78 (0.04)	99	0.77
33	Air	0.41 (0.11)	0.98 (0.02)	659	0.82
33	Sea	0 (0)	0.87 (0.01)	1075	0.82
33	Road	-0.76 (0.16)	0.85 (0.04)	99	0.81
34	Air	0.52 (0.12)	0.78 (0.02)	552	0.72
34	Sea	0 (0)	0.82 (0.01)	1046	0.89
34	Railway			1	
34	Road		0.77 (0.03)	85	0.85
35	Air		0.7 (0.03)	311	0.69
35	Sea		0.81 (0.02)	400	0.85
35	Road		0.65 (0.08)	36	0.67
36	Air		0.82 (0.1)	39	0.64
36	Sea		0.83 (0.04)	77	0.82
36	Road		0.71 (0.02)	528	0.68
37	Air		0.61 (0.05)	113	0.62
37	Sea		0.8 (0.04)	146	0.77
37	Road		0.71 (0.02)	528	0.68
38	Air	0.37 (0.11)	0.83 (0.02)	845	0.7
38	Sea		0.79 (0.01)	1397	0.84
38	Road		0.71 (0.04)	103	0.77
39	Air	0.31 (0.05)	0.78 (0.01)	2783	0.73
39	Sea	-0.01 (0)	0.82 (0)	4318	0.89
39	Road		0.69 (0.02)	337	0.72
40	Air	0.2 (0.06)	0.87 (0.01)	2123	0.79
40	Sea	0.12 (0.06)	0.81 (0.01)	2551	0.79
40	Railway			1	
40	Road		0.69 (0.04)	204	0.65
41	Air	0.69 (0.28)	0.82 (0.04)	93	0.83
41	Sea	-0.72 (0.23)	0.87 (0.05)	74	0.86

Commodity group (HS 2012)	Mode of transport	Coefficients			
		Distance	CIF value	N	R^2
41	Road		0.71 (0.02)	528	0.68
42	Air	0.21 (0.09)	0.8 (0.02)	726	0.7
42	Sea	0.27 (0.1)	0.79 (0.02)	716	0.77
42	Railway			1	
42	Road	0.41 (0.14)	0.68 (0.04)	76	0.79
43	Air		0.36 (0.06)	169	0.2
43	Sea		0.74 (0.06)	83	0.68
43	Road		0.71 (0.02)	528	0.68
44	Air		0.74 (0.03)	319	0.64
44	Sea		0.78 (0.01)	1001	0.85
44	Road		0.68 (0.04)	96	0.72
45	Air		0.86 (0.08)	55	0.71
45	Sea		0.77 (0.05)	56	0.83
45	Road		0.71 (0.02)	528	0.68
46	Air		0.99 (0.1)	40	0.71
46	Sea		0.84 (0.04)	118	0.81
46	Road		0.71 (0.02)	528	0.68
47	Air		0.36 (0.06)	169	0.2
47	Sea		0.83 (0.03)	44	0.93
48	Air	0.19 (0.08)	0.73 (0.01)	1483	0.69
48	Sea		0.79 (0.01)	2632	0.89
48	Railway			1	
48	Road		0.67 (0.03)	176	0.76
49	Air	0.35 (0.12)	0.78 (0.02)	681	0.69
49	Sea		0.88 (0.02)	704	0.82
49	Road		0.61 (0.1)	45	0.48
50	Air		0.97 (0.08)	45	0.79
50	Sea		0.74 (0.06)	83	0.68
50	Road		0.71 (0.02)	528	0.68
51	Air		1.07 (0.04)	98	0.88
51	Sea		0.99 (0.05)	88	0.79
51	Road		0.71 (0.02)	528	0.68
52	Air	0.55 (0.14)	0.93 (0.03)	456	0.74
52	Sea	0.12 (0.06)	0.84 (0.01)	772	0.89
52	Road		0.71 (0.02)	528	0.68
53	Air		0.9 (0.07)	68	0.73
53	Sea		0.81 (0.04)	103	0.78
54	Air		0.86 (0.02)	358	0.78
54	Sea		0.8 (0.01)	594	0.87
54	Road		0.71 (0.02)	528	0.68
55	Air	0.33 (0.16)	0.88 (0.03)	301	0.8
55	Sea		0.84 (0.01)	593	0.87
55	Road		0.71 (0.02)	528	0.68

Commodity group (HS 2012)	Mode of transport	Coefficients			
		Distance	CIF value	N	R^2
56	Air		0.82 (0.03)	418	0.72
56	Sea		0.81 (0.01)	662	0.85
56	Road		0.69 (0.05)	63	0.72
57	Air		0.75 (0.05)	162	0.59
57	Sea		0.75 (0.02)	365	0.73
57	Road		0.71 (0.02)	528	0.68
58	Air		0.9 (0.03)	379	0.74
58	Sea		0.83 (0.02)	379	0.8
58	Road		0.71 (0.02)	528	0.68
59	Air		0.87 (0.03)	356	0.77
59	Sea		0.85 (0.02)	422	0.79
59	Road		0.71 (0.02)	528	0.68
60	Air		0.93 (0.03)	144	0.85
60	Sea	0 (0)	0.89 (0.02)	270	0.88
60	Road		0.71 (0.02)	528	0.68
61	Air		0.85 (0.01)	2184	0.8
61	Sea	0.13 (0.05)	0.88 (0.01)	2369	0.82
61	Road		0.74 (0.02)	246	0.83
62	Air		0.84 (0.01)	2485	0.75
62	Sea	0.24 (0.05)	0.83 (0.01)	2756	0.8
62	Road		0.72 (0.02)	239	0.79
63	Air	0.25 (0.08)	0.83 (0.02)	893	0.72
63	Sea		0.82 (0.01)	1463	0.79
63	Road		0.71 (0.04)	148	0.73
64	Air	0.3 (0.09)	0.83 (0.02)	561	0.8
64	Sea		0.84 (0.01)	655	0.87
64	Road		0.79 (0.05)	64	0.83
65	Air		0.81 (0.03)	224	0.75
65	Sea		0.81 (0.03)	267	0.79
65	Road		0.71 (0.07)	33	0.81
66	Air		0.46 (0.11)	52	0.29
66	Sea		0.77 (0.05)	116	0.66
66	Road		0.71 (0.02)	528	0.68
67	Air	0.92 (0.32)	0.73 (0.06)	87	0.64
67	Sea		0.81 (0.03)	106	0.87
67	Road		0.71 (0.02)	528	0.68
68	Air	0.54 (0.12)	0.78 (0.02)	637	0.69
68	Sea	0.17 (0.08)	0.73 (0.01)	1071	0.83
68	Railway			1	
68	Road		0.59 (0.05)	89	0.65
69	Air		0.72 (0.03)	385	0.61
69	Sea		0.79 (0.01)	785	0.82
69	Road		0.75 (0.05)	55	0.83

Commodity group (HS 2012)	Mode of transport	Coefficients			
		Distance	CIF value	N	R^2
70	Air	0.22 (0.09)	0.73 (0.02)	1084	0.65
70	Sea		0.75 (0.01)	1566	0.81
70	Road		0.68 (0.04)	101	0.72
71	Air		0.64 (0.05)	226	0.43
71	Sea		0.64 (0.04)	185	0.61
71	Road		0.71 (0.02)	528	0.68
72	Air	0.28 (0.13)	0.65 (0.02)	487	0.67
72	Sea		0.73 (0.01)	1826	0.88
72	Road		0.62 (0.03)	124	0.74
73	Air	0.25 (0.04)	0.82 (0.01)	3821	0.78
73	Sea		0.8 (0.01)	4815	0.83
73	Railway			1	
73	Road	0.17 (0.07)	0.68 (0.02)	526	0.66
74	Air		0.74 (0.02)	541	0.67
74	Sea		0.82 (0.02)	698	0.79
74	Road	1.17 (0.25)	0.87 (0.07)	56	0.73
75	Air		0.74 (0.08)	82	0.51
75	Sea		0.83 (0.05)	78	0.8
75	Road		0.71 (0.02)	528	0.68
76	Air		0.77 (0.02)	609	0.73
76	Sea		0.8 (0.01)	965	0.84
76	Road		0.72 (0.05)	70	0.73
78	Air		0.36 (0.06)	169	0.2
78	Sea		0.84 (0.05)	55	0.84
78	Road		0.71 (0.02)	528	0.68
79	Air		0.78 (0.09)	54	0.61
79	Sea		0.73 (0.04)	94	0.8
79	Road		0.71 (0.02)	528	0.68
80	Air		0.36 (0.06)	169	0.2
80	Sea		0.83 (0.06)	35	0.85
80	Road		0.71 (0.02)	528	0.68
81	Air		0.71 (0.06)	78	0.63
81	Sea		0.69 (0.05)	70	0.77
81	Road		0.71 (0.02)	528	0.68
82	Air	0.21 (0.07)	0.89 (0.01)	2007	0.75
82	Sea		0.84 (0.01)	2596	0.79
82	Road	0.36 (0.09)	0.68 (0.03)	308	0.65
83	Air	0.19 (0.07)	0.87 (0.01)	1427	0.79
83	Sea		0.83 (0.01)	1583	0.85
83	Road		0.62 (0.05)	119	0.58
84	Air		0.74 (0.01)	10945	0.56
84	Sea		0.64 (0.01)	12983	0.44
84	Railway			3	

Commodity group (HS 2012)	Mode of transport	Coefficients Distance	CIF value	N	R^2
84	Road		0.58 (0.02)	1363	0.34
85	Air	0.12 (0.04)	0.71 (0.01)	9400	0.56
85	Sea		0.69 (0.01)	8973	0.56
85	Railway			1	
85	Road		0.55 (0.02)	888	0.37
86	Air		0.82 (0.09)	69	0.54
86	Sea		0.67 (0.06)	166	0.46
86	Road		0.71 (0.02)	528	0.68
87	Air		0.88 (0.01)	1612	0.8
87	Sea		0.81 (0.01)	2448	0.73
87	Railway			2	
87	Road		0.8 (0.06)	222	0.56
88	Air		0.72 (0.07)	117	0.52
88	Sea		0.46 (0.13)	35	0.37
88	Road		0.71 (0.02)	528	0.68
89	Air		0.57 (0.11)	55	0.34
89	Sea				
89	Road		0.71 (0.02)	528	0.68
90	Air		0.63 (0.01)	4160	0.45
90	Sea		0.59 (0.01)	3000	0.45
90	Railway			2	
90	Road	0.57 (0.11)	0.53 (0.05)	294	0.29
91	Air		0.76 (0.04)	393	0.56
91	Sea		0.67 (0.03)	349	0.6
91	Road		0.71 (0.02)	528	0.68
92	Air		0.64 (0.07)	132	0.39
92	Sea		0.74 (0.04)	280	0.62
92	Road		0.71 (0.02)	528	0.68
93	Air		0.88 (0.06)	93	0.78
93	Sea		0.92 (0.05)	63	0.86
93	Road		0.71 (0.02)	528	0.684486
94	Air	0.27 (0.09)	0.86 (0.02)	1057	0.675909
94	Sea	-0.21 (0.06)	0.89 (0.01)	2122	0.801682
94	Railway			1	
94	Road		0.7 (0.04)	172	0.637698
95	Air		0.77 (0.03)	472	0.615561
95	Sea		0.79 (0.02)	762	0.692771
95	Road	-0.64 (0.32)	0.63 (0.07)	63	0.595174
96	Air		0.77 (0.02)	936	0.656976
96	Sea		0.81 (0.01)	1343	0.784073
96	Railway			1	
96	Road		0.67 (0.03)	137	0.787616
97	Air		0.4 (0.18)	50	0.105945

Commodity group (HS 2012)	Mode of transport	Coefficients		N	R^2
		Distance	CIF value		
97	Sea	-1.17 (0.37)	0.87 (0.08)	80	0.595949
97	Road		0.71 (0.02)	528	0.684486

Note: Standard error in parentheses.