UNITED NATIONS CONFERENCE ON TRADE AND DEVELOPMENT

June 2022

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

Developing a global transport costs dataset for international trade

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

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.

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	