Journal of Global Economic Analysis, Volume 4 (2019), No. 1, pp. 28-49.

Estimating International Trade Margin Shares by Mode of Transport for the GTAP Data Base

BY JOSE NUNO-LEDESMA[a] AND NELSON VILLORIA[b]

We estimate international transportation margin shares by mode of transport for the Global Trade Analysis Project (GTAP) Data Base. For each available origin-destination-GTAP sector triplet, we estimate the fractional share of the transport margin attributable to air, water, and other shipping modes. We use published relationships between ad valorem proportional changes in prices due to transportation costs and distances, weight-value ratios and fuel prices. Our final database contains 344,554 observations (origin-destination-sector-mode combinations) with transportation margin modal shares organized by 228 exporter countries, 209 importer countries, 45 traded GTAP sectors and 3 transportation modes. The main contribution of this article is to bring a more comprehensive set of information on trade by transport mode covering around 65% of global trade in 2004 and 55% in 2011. Our estimated shares contrast with those traditionally used in the GTAP Data Base, which are extrapolations based solely on the modes of transport used by US exporters. A comparison of our shares with those used in version 9.0 of the GTAP Data Base reveals that the role of water transportation services in international trade is underestimated, while that of air transportation is overestimated. Overall, we find that estimations using the modal shares in version 9.0 of the GTAP Data Base overestimate the greenhouse gas emissions associated with international transport. Our new data were used to estimate transport margins by mode in Version 9.1 of the GTAP Data Base, and it is expected that our methods will be used to update future versions of the database.

JEL codes: C68, D58, F1, Y1.

Keywords: Transportation Margin Shares; International Trade; GTAP Data Base.

---

[a] Banco de México, Avenida 5 de Mayo 18, Cuauhtémoc, Ciudad de México 06059 (e-mail: jnuno@banxico.org.mx). The views and conclusions presented in this document are exclusively the responsibility of the authors and do not necessarily reflect those of Banco de México.

[b] Kansas State University, 222 Waters Hall, Manhattan, Kansas 66506 (e-mail: nvilloria@ksu.edu).

1. Introduction

An accurate representation of the modes of transport used in international trade is important to study issues that range from trade-related greenhouse gas (GHG) emissions (Cristea et al., 2013) to the effects of substituting locally produced food for imports (Ballingall and Winchester, 2010; Avetisyan et al., 2014) to the choice of transportation modes in the wake of changes in relative prices due to increased exposure to international trade (Avetisyan and Hertel, 2015; Avetisyan et al., 2015).

Modes of international transport in the standard Global Trade Analysis Project (GTAP) model (Hertel, 1997) are handled by a global transport services industry that allocates a composite international transport good among exporting countries in direct proportion to their global export shares of a given product. The composite international transport good is a bundle of air, water (mostly vessel), and other (land, predominantly truck and rail) transportation services, coded as atp, wtp, and otp in the GTAP sectoral classification. The combination of the Free on Board (FOB) price at the country of origin with the composite international transport good generates a Cost, Insurance and Freight (CIF) price at the destination country.

Prior to Version 9.1 of the GTAP Data Base (Aguiar et al., 2016), released in May 2016), the percentage of the international transport margin along each origin-destination-product link that is paid to either air, water or other transportation services, or *modal shares*, were solely based on transport mode information for the US (Gehlhar and McDougall, 1997, 2006, 2016). Although useful, this approach required, in the words of their authors, "heroic assumptions" that justified extrapolating transport mode shares from US trade with Mexico and Canada to the rest of the world (Gehlhar and McDougall, 1997). To the extent that other sources of information were unavailable or difficult to obtain, this approach proved useful in supplementing the GTAP Data Base.

A major opportunity to improve these modal shares arose from the work of Cristea et al. (2013) quantifying trade-related GHG emissions. A critical contribution of Cristea et al. (2013) was to compile extensive data on bilateral values and quantities, covering around 65% of the global trade in merchandise, by mode of transport: vessel, rail, truck, and air. These data comprise detailed information on US exports by mode to 227 countries. They also compiled data from Eurostat, with detail on transport modes of intra-EU trade, as well as the transport modes of European exports and imports to/from 227 countries. In addition, they collected data on transportation modes for goods transported from 232 countries to 11 Latin American countries. These data are specific to each origin-destination-tariff line code.

In this context, our contribution is to bring the data from Cristea et al. (2013) to bear on the estimation of an improved set of international modal shares for the GTAP Data Base. The main difficulty for accomplishing this is that, except for the US, for each bilateral transaction, we observe either CIF or FOB values, but not both. Not having both the CIF and FOB value by mode for data other than US exports implies that we are unable to calculate the portion of the international transport margin that is paid to a given transportation mode.

To circumvent this difficulty, we use a set of equations estimated by Hummels et al. (2014) that relate the international transport margin that is paid to a given transportation mode by US exporters to the weight-to-value ratios of the shipped merchandise, the distance between the US and the importing destination, and fuel prices. We use these equations to predict, for any given product and origin-destination pair, the international transport margin accruing to vessels, airplanes, trucks and rail, as a share of the total value of the bilateral transaction. For this, we use the weight-to-value ratios of the transaction valued at either CIF or FOB values, both of which are available from Cristea et al. (2013), and bilateral distances and fuel prices, both readily available from public sources. Of course, the sum of the transport margins across modes is the overall international transport margin, and therefore, it is straightforward to express the transport margins as modal shares.

The fact that these modal shares are based on a much denser matrix of international trade flows by mode suggests that they should be more accurate than modal shares inferred from US transactions only. Yet, a caveat to keep in mind is that the equations in Hummels et al. (2014) were also estimated using US data only. In this sense, the accuracy of the shares also depends on the constancy of the parameters across different regions. Hummels et al. (2014) suggest that these parameters are indeed reasonably constant as they use these equations to estimate international transport margins of Danish exports and imports. Further examining the validity of this assumption may be a useful exercise that will have to wait until available data on transportation margins by mode become available for locations other than the US. Notice also, that the fact that the data on US export transport margins by mode supports the parameter estimates used to retrieve either the CIF or FOB values of non-US countries precludes using the US data to validate our approach. In other words, such an exercise would amount to nothing else than a within-sample validation of Hummels et al. (2014) parameter estimates.

With these limitations in mind, a comparison of the improved shares with those based on US data only, as used in previous version of the GTAP Data Base, reveal that the latter tend to overestimate air and other (land) transportation modes. When converted to Carbon dioxide (CO_2) equivalents, a focus area in many studies of the environmental footprint of international trade (e.g. Cristea et al., 2013), the previous GTAP estimates of modal shares indicate that trade-related emissions amount to 4,377 metric tons (MT) of CO_2 equivalent; our modal shares indicate 3,055 MT of CO_2. The difference in emissions is explained by the larger air transportation shares under the previous methodology, which have emission factors two to three orders of magnitude larger than those of water and land transportation.

Our final database contains 344,554 observations (origin-destination-sector-mode combinations) with transportation margin shares organized by 228 exporter countries, 209 importer countries, 45 traded GTAP sectors and 3 transportation modes [air, sea, and other (mostly land)]. The new dataset, raw data used and code to reproduce our calculations are available in the supplementary files published with this paper. The data were used to estimate transport margins by mode in Version 9.1 of the GTAP Data Base, and it is expected that the methods outlined in this paper will be used to update future versions of the GTAP Data Base and other related databases such as the GTAP-Power Data Base (Peters, 2016).

This paper has four further sections. The next section outlines the conceptual foundations of our strategy. Section 3 discusses the data as well as the empirical steps followed in the estimation of the transport modal shares. Section 4 presents the results of our estimations, compares them to the margin shares in version 9.0 of the GTAP Data Base, and exhibits a brief exercise where we compare the distribution of estimated trade-related carbon dioxide emissions using both our modal shares and those in Version 9.0 of the GTAP Data Base. The last section concludes.

2. Empirical strategy

2.1 Conceptual overview

Before we formally describe how we estimate the trade margin shares by mode of transport, we present one example to illustrate what trade margins and trade margin shares are in this document. Consider the delivery of textiles via maritime modes of transport (using vessels, for example). The International Commercial

Terms (Incoterms) Free on Board (FOB), and Cost, Insurance and Freight (CIF) are pre-defined commercial terms used to specify the point at which the range of obligations and risks associated with the shipment of the textiles shifts from seller to buyer. Under an FOB contract, the seller delivers the goods on board to a vessel at a port of shipment named by the buyer, and the buyer is required to cover costs generated by marine transportation, insurance, unloading and other expenses generated during the freight transport. With a CIF contract, on the other hand, the seller is the one who covers all shipping expenses generated until the textiles are delivered to the destination port named by the buyer. Regardless of the type of contract, authorities register the total value of the shipment, which includes the value of the textiles plus additional charges related to the freight transport. Importantly, the total value of textile shipments using CIF and FOB contracts usually differ. This difference is the trade margin associated with shipments of textiles made using (in this example) water-based modes of transport. There are similar Incoterms used to classify transaction arrangements that use modes of transport other than water. For economy of space and to make our exposition clear, we refer to all delivery schemes where the seller takes on the costs and risks generated by transporting the goods as 'CIF trades', and label all shipment arrangements where the buyer absorbs these costs and risks 'FOB trades'. In a similar fashion, we can estimate trade margins for shipments of textiles made via modes of transport other than water. It follows that the margin share corresponding to shipments of textiles realized via water is the trade margin of shipments via water divided by the sum of trade margins of deliveries made via all modes of transport. Our goal is to estimate these modal shares.

2.2 Methodology

The trade margin μ for a shipment of good g originated in country o and destined to country d using mode of transport m is defined as:

$$\mu_{odgm} = CIF_{odgm} - FOB_{odgm}. \tag{1}$$

Using these margins, we can estimate margin shares κ by mode of transport as:

$$\kappa_{odgm} = \frac{\mu_{odgm}}{\sum_m \mu_{odgm}}, \tag{2}$$

where $\sum_m \kappa_{odgm} = 1$. The data sources available to us report either FOB or CIF values, but to estimate margins using equation (1), we need both. Therefore, we must first estimate the necessary either FOB or CIF value.

Under the assumption that the costs of shipping are a constant proportion of the shipped value, we can write the relationship between CIF and FOB values as:

$$\frac{CIF_{odgm}}{FOB_{odgm}} = 1 + \tau_{odgm}, \tag{3}$$

where τ is the *ad valorem* transport cost of delivering goods from sector *g* from exporter *o* to importer *d*. Provided we can recover τ, equation (3) would allow us to estimate the missing CIF or FOB value, to proceed with the computations in equations (1) and (2).

To estimate τ, we use the transport-cost equations estimated by Hummels et al. (2014). These authors estimate transportation costs as functions of a weight-value ratio, oil or jet fuel prices, and distance between exporters and importers. The specific functions as estimated by Hummels et al. (2014) estimate the logarithm of τ for shipments made via water (equation 4), air (5), road (6), and rail (7).

$$\ln(\tau_{odg}) = -4.16 + 0.39\ln(WV_{og}) + 0.35\ln(oil) + 0.03DIS_{od} + 0.01\ln(oil)\,DIS_{od} \tag{4}$$

$$\ln(\tau_{odg}) = -3.80 + 0.44\ln(WV_{og}) + 0.21\ln(jet) + 0.03DIS_{od} + 0.02\ln(jet)\,DIS_{od} \tag{5}$$

$$\ln(\tau_{odg}) = -8.18 + 0.23\ln(WV_{og}) + 0.86\ln(oil) + 1.21\ln(DIS_{od}) - 0.37\ln(oil)\ln(DIS_{od}) \tag{6}$$

$$\ln(\tau_{odg}) = -4.37 + 0.54\ln(WV_{og}) + 0.08\ln(oil) - 0.90\ln(DIS_{od}) - 0.22\ln(oil)\ln(DIS_{od}). \tag{7}$$

where *jet* refers to the jet fuel price (in USD per barrel), *oil* refers to the crude oil price (in USD per gallon), DIS denotes the distance in thousands of kilometers between exporter and importer, and WV_{og} is a proxy for bulkiness that varies only across exporter-sector combinations. This proxy is referred to as the weight-value ratio, and is defined as:

$$WV_{og} = \frac{\sum_d \sum_{h\in g} w_{odh}}{\sum_d \sum_{h\in g} v_{odh}}, \tag{8}$$

where *w* and *v* are the weight and value of the shipment, respectively; while the sub index *h* refers to a commodity.[1] With help from equations 4 to 7,

[1] Commodities are elements of GTAP sectors. For example, the commodities barley and oats are members of the GTAP sector Other Grains.

we retrieve τ, which is then used to recover the needed FOB or CIF value to calculate μ_{odgm}.

A caveat to keep in mind is that the GTAP Data Base contains modes *air*, *water* and *other*. The GTAP *other* mode includes road and rail transport, pipelines, auxiliary transport activities, and travel agencies. For our purposes, our own *other* mode will include the sum of road and rail transport, as we do not have information on pipelines, auxiliary transport activities, and travel agencies.

3. Data and Methods

In this section we describe the data we used, the steps we followed, and other assumptions we relied on to operationalize our procedure to estimate trade margin shares by transportation mode. Our three main sources of data are: imports from Latin American countries which are members of the Latin American Integration Association (ALADI); imports from both the United States Bureau of the Census (henceforth the Census) and the North American Transborder Freight Data (NATFD); and bilateral trade from the European Statistical Office (Eurostat) containing information on exports and imports between 15 European nations and 226 countries[2].

3.1 Data from Latin American

The information reported by ALADI contains data on the imports of 11 Latin American countries from 232 trade partners.[3] For all Latin American countries, except Mexico, these data were reported in 2004. Whenever possible, we work with the most recent data (2004), this is possible for most cases, except for Mexico. We used Mexican data from 2000, because the 2004 data for this country did not include information on weight of imports. Relevant variables in the dataset we utilize are: The 6-digit Harmonized System (HS-6) product code; mode of transportation (air, water, rail, road, and others); shipment values in dollars, and shipment weights in hundreds of kilograms. Because equations (4) to (7) can be

---

[2] Most of the data in this paper is not easily retrievable from websites, or are simply unavailable, as in in the case of the data from ALADI. We are indebted to Laura Puzzello, Anca Cristea, and David Hummels who graciously shared the Latin American, US, and European data they used in Cristea et al. (2013). The US data is also available on DVD's from the US Bureau of the Census. To complement the information on nomenclature, we relied on the Census' own reference lists available in https://www.census.gov/foreign-trade/reference/index.html, and in reference lists from the online appendix for Hummels (2017).

3 The Latin American countries in the ALADI database include Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Mexico, Paraguay, Peru, Uruguay, and Venezuela.

only used to estimate transport costs for air, water, rail and road, we exclude transportation methods coded in the ALADI database as "others".

We use the HS to GTAP sectors concordance reported by the World Bank's World Integrated Trade Solution (WITS, World Integrated Trade Solution, 2011). Out of the 4,864 unique HS-6 codes included in the database, three can be mapped to two GTAP sector codes. In these cases, we assign half the value and weight to one GTAP sector and the other half to the second possible GTAP sector concordance. For example, the HS-6 code 121299 is an HS-6 code classifying miscellaneous vegetable products including locust beans, seaweeds and other algae. According to WITS, this HS code can be assigned to either GTAP sector 4 (vegetables, fruits, and nuts) or 8 (crops other). Thus, in every instance were the code 121299 appears, we divide the reported value and weight by two and impute the result to both GTAP sectors 4 and 8.

3.2 Data from the United States

Data regarding imports made by the United States come from the Census and the NATFD. From the Census, we obtain information about imports delivered via air and water modes. From the NATFD we retrieve data on merchandise imported via rail and road transport.

The Census database contains observations reported in 2004 of imports coming from 227 different trading partners. Key variables include: commodity classified by a 10-digit Harmonized Tariff Schedule (HTS-10) code; the value in dollars of the delivered shipment and its weight in kilograms, and the total value of imports for consumption in dollars.

The 10-digit number used by the Census to codify commodities is based on the HS-6 classification. To add granularity to the categorization of products, the United States International Trade Commission attaches four extra digits to the HS-6 code to generate their HTS-10 classifier. The Census uses the HTS-10 code. For our purposes, when mapping products to GTAP sectors, we take into consideration only the first 6 digits of the classification code reported by the Census. Again, in three occasions, a single HS-6 number could be assigned to two possible GTAP sectors. We split values and weights as we did for the Latin American database.

The Census does not report details regarding imports made via transport modes other than air and vessel. We retrieve information about imports made via road and rail using the NATFD database. This data include US imports acquired from Canada and Mexico in 2004. However, imports coming from either Mexico or

Canada are registered as coming from a single trade partner. Key variables include: commodity categorized by a 2-digit Harmonized System code; values in dollars imported by rail and road, the total value of goods imported from Mexico and Canada, weight in kilograms of imports shipped via rail and road, and total weight of imports shipped from the Mexico-Canada trade partner. Details about our estimation of weight shipped by rail and road for the US database can be found in Appendix A.

3.3 Data from Europe

The database from Eurostat we use includes information on exports and imports among 15 European countries and 226 partners in 1999. [4] The main variables in the database are: values in thousands of euros, weights in tons, and modes of transportation for goods traded among the 15 countries, as specified by the 3-digit Revised Standard Classification of Goods by the Transport Statistics (NSTR, by its acronym in French),

We converted values from euros to US dollars using the exchange rate on June 15, 1999 (1.042 USD per euro). Eurostat reports the following modes of transportation: water, rail, road, air, post, fixed mechanisms, inland waterway, self-propulsion, and unknown. For our purposes, we only consider data on shipments made via water, rail, road and air. This is because the equations we use to estimate *ad-valorem* transport costs (equations 4 to 7) can predict costs for shipments made via water, air, rail and road only. To map commodities codified with the NSTR nomenclature to GTAP sectors, we employ the same concordance table used by Cristea et al. (2013).[5]

3.4 Other data and procedures

To fully operationalize our procedure to estimate trade costs (equations 4 to 7), we require oil and jet fuel prices, bilateral distances between exporters and importers,

[4] The European countries are: Austria, Belgium, France, Denmark, Finland, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden and the United Kingdom.

[5] The concordance table used in Crstea et al. (2013) does not have GTAP correspondence for four NSTR codes. We assigned mappings for these goods as follows: the NSTR code 325 (Distillate fuels) was assigned to the GTAP sector 32; the NSTR code 546 (Steel hoop and strip) was assigned to the sector 35; the NSTR code 721 (basic slag) was classified as GTAP 33, while the NSTR code 713 was assigned to sector 18. Various NSTR codes can match to more than one GTAP sector; when the aggregation is performed, we divided the value and weight equally among the corresponding GTAP sectors.

and estimates of the weight-value ratio for each commodity classification group, as indicated above (equation 8).

For oil prices, we use USD per barrel of Brent-Europe barrel published by the United States Energy Information Administration (2018). For jet fuel prices we employ the spot price in USD per gallon of Gulf Coast Kerosene-Type Jet Fuel as reported annually by the United States Energy Information Administration (2018). To use fuel prices from the same year as the source datasets, we use 1999 oil and jet fuel prices when working with the Eurostat database, and we utilize the 2004 observations when working with data from US and Latin America.

Regarding data on origin-destination distances, we used two sources depending on the mode of transportation used for a given trade flow. For trades realized in modes other than water, we rely on information from the GeoDist database (Mayer and Zignago, 2011) published by the Center for Prospective Studies and International Information (CEPII, by its French acronym). GeoDist distances are calculated using the great circle formula, which relies on latitudes and longitudes of the most populated cities or the capitals of the countries involved (Mayer and Zignago, 2011). For shipments sent via water, we rely on maritime distances published by the Center for Studies and Research on International Development (CERDI, by its acronym in French) in the CERDI-seadistance database (Bertoli et al., 2016).[6] The CERDI-seadistance database reports the length of the shortest existing sea route between the origin and destination ports (Bertoli et al., 2016).

We construct our weight-value ratio (WV), as defined in equation 8. Recall that the weight-value ratio is constant for a given origin-good combination. The sub index h will refer to whether we use an NSTR or an HS-6 commodity classification code depending on whether we are working with European data or not.

We generate a single database merging the data from Latin America, United States, and Europe. We omit duplicate origin-destination combinations present in two or more of the databases. If an origin-destination pair is present in both the ALADI and the Eurostat databases, we exclude the observations from the European data because the information from Latin America is more recent. Likewise, if an origin-

[6] In addition to more accurately measuring shipping distances, by using the CERDI-seadistance database instead of the GeoDist data to measure bilateral distances for maritime shipments, we preserve, respectively, 200, 2109, and 39 more observations in the Latin American, European, and US datasets. This is because some origin-destinations present in CERDI-seadistance database are absent in the GeoDist database. We thank an anonymous referee for bringing the CERDI-seadistance dataset to our attention.

destination combination is present in both the Eurostat base and the data from the United States, we exclude the observations from Eurostat; also, because the US information is more recent.

4. Results

The output of the exercise described above is a database of modal shares for 344,554 unique product-level, bilateral trade flows. The main innovation in this dataset relative to current practice is to bring comprehensive data on modal transport services to share out the trade margins in the GTAP Data Base. With this in mind, it is interesting to explore the extent to which the existing GTAP modal shares differ from ours. For this, we compare our shares with those in version 9.0 of the GTAP Data Base, which has a 2011 reference year. The country-pair-product combinations available to us cover 55% of the world trade reported in this version of the GTAP Data Base.

At the most aggregate level (across commodities and trade flows), Table 1 shows that, relative to our results, the modal shares calculated by Gehlhar and McDougall (1997, 2006, 2016) overestimate the share of the international transport margins attributed to air transportation, and underestimate the shares paid to water and other (land) transportation services, at least for the 55% of the world trade represented in our data. In our estimates, 59% of total trade margins are attributed to water transport services, 30% is paid to land (other) transportation services, and only 11% is paid to air transportation services. Decomposing the aggregate numbers in Table 1, as illustrated in Figure 1, payments to air transport estimated by Gehlhar and McDougall (1997, 2006, 2016) are higher than our calculations in all broad sectoral groups considered. Our method also produces higher estimates for land and water transport shares than Gehlhar and McDougall (1997, 2006, 2016) for all sectoral groups.

Table 1. Overall modal shares across all GTAP merchandize sectors. Gehlhar and McDougall (1997, 2006, 2016, GTAP V 9.0) vs. our estimates

Type	Air	Other	Water	Total
GTAP V 9.0	0.17	0.42	0.40	1.00
Our estimates	0.11	0.30	0.59	1.00

Note: Transport shares by mode are aggregated using trade weights using bilateral trade flows in year 2011 from the GTAP Data Base V 9.0. The comparison is limited to the importer-exporter-product triplets for which we have actual trade data by mode of transport; these data comprise 55% of total global trade in 2011.

Source: Authors' calculations with data from GTAP Data Base Version. 9.0 and Gehlhar and McDougall (1997, 2006, 2016).

It is tempting to ask which shares are more accurate. In this sense, it is important to recognize that the new shares reflect a much denser matrix of trade flows by transport mode than those available to Gehlhar and McDougall (1997, 2006, 2016), and therefore, they are presumably more accurate. It is less apparent, however, why this could matter. After all, in typical applications of the GTAP Data Base, transportation services are aggregated into a single margin commodity, with the value of this margin reflecting difference between CIF and FOB prices.— Whenever the focus is on trade taxes, the changes in the price of the margin commodity are generally small and treated as a nuance. In other applications, however, the margins by mode are the focus of the exercise. In particular, analysis of the changes of GHG emissions due to international trade, such as Cristea et al. (2013), focus squarely on the composition of the international transport margins. The composition of the margins is important because the emission factors by mode of transport differ by orders of magnitude. Cristea et al. (2013) report that transporting one metric ton (MT) of merchandise over one kilometer using airplanes produces between 476 and 1,020 grams of CO_2. In contrast, transporting one MT of merchandise over one kilometer using rail or a maritime container produces 22.7 and 12.1 grams of CO_2, correspondingly.

We contrast trade-related CO_2 emissions calculations using first our resulting modal margin shares, and then with those in Version 9.0 of the GTAP Data Base, which use Gehlhar and McDougall (1997, 2006, 2016) transport model estimates. We calculate trade related emissions by multiplying the weight (in MT) of each bilateral trade flow times the distance (in km) between trading partners times the emission factor of each mode in grams of CO_2/MT-km. Emission factors used are the low scenario in Cristea et al. (2013): 552 grams of CO_2/MT-km for air transportation, 119.7 for land transportation and 16.3 for water transportation. As shown in Figure 2, relatively high air transport shares in Version 9.0 of the GTAP Data Base result in higher estimates for transport-related CO_2 emissions in all sectors than when our transport mode estimates are used.

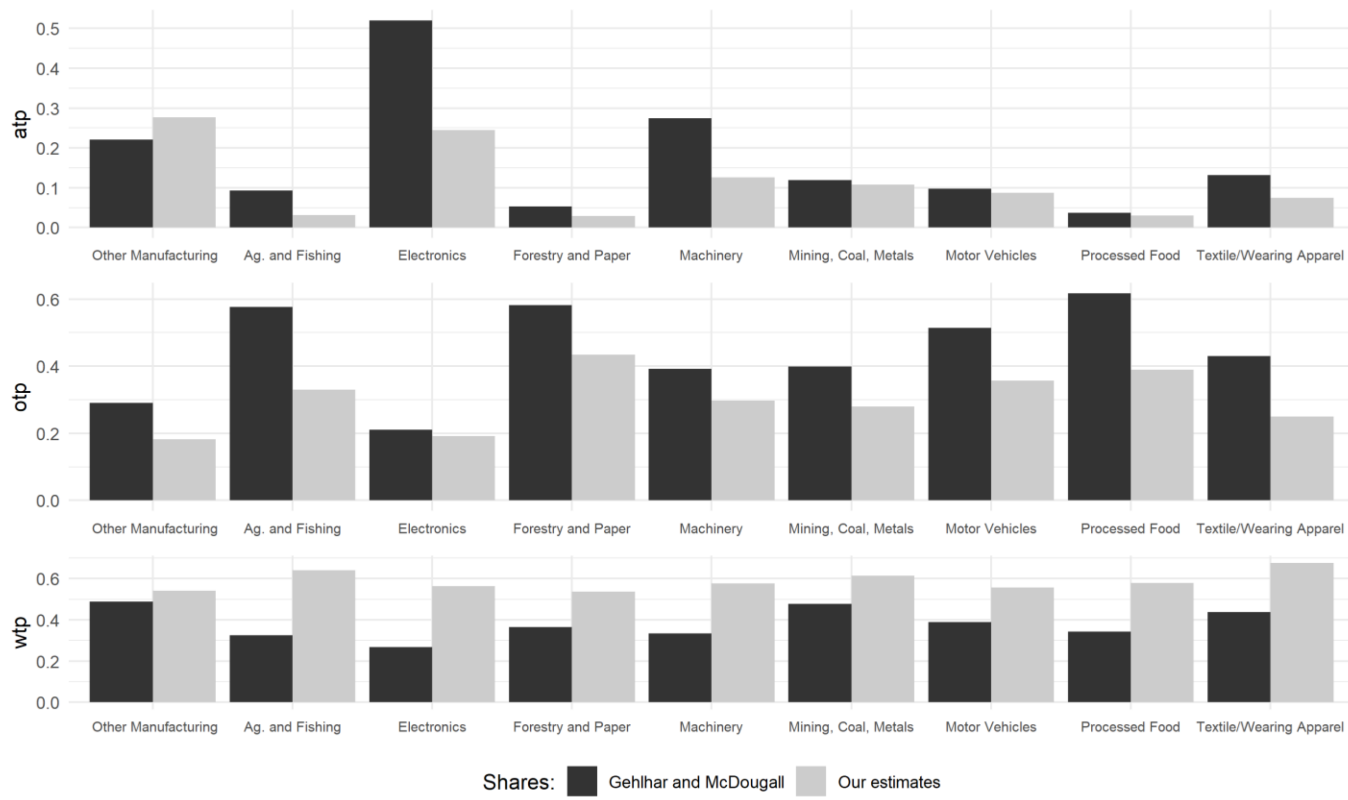

Figure 1: Transport modal shares by aggregated GTAP merchandize sectors. Gehlhar and McDougall (1997, 2006, 2016) vs. our estimates.

Notes: wtp, otp, and atp are water, other (land) and air transportation services. Transport shares by mode are aggregated using trade weights using bilateral trade flows in year 2011 from the GTAP Data Base V9, reference year 2009.

Source: Authors' calculations with data from GTAP Data Base Versoin. 9.0 and Gehlhar and McDougall (1997, 2006, 2016).

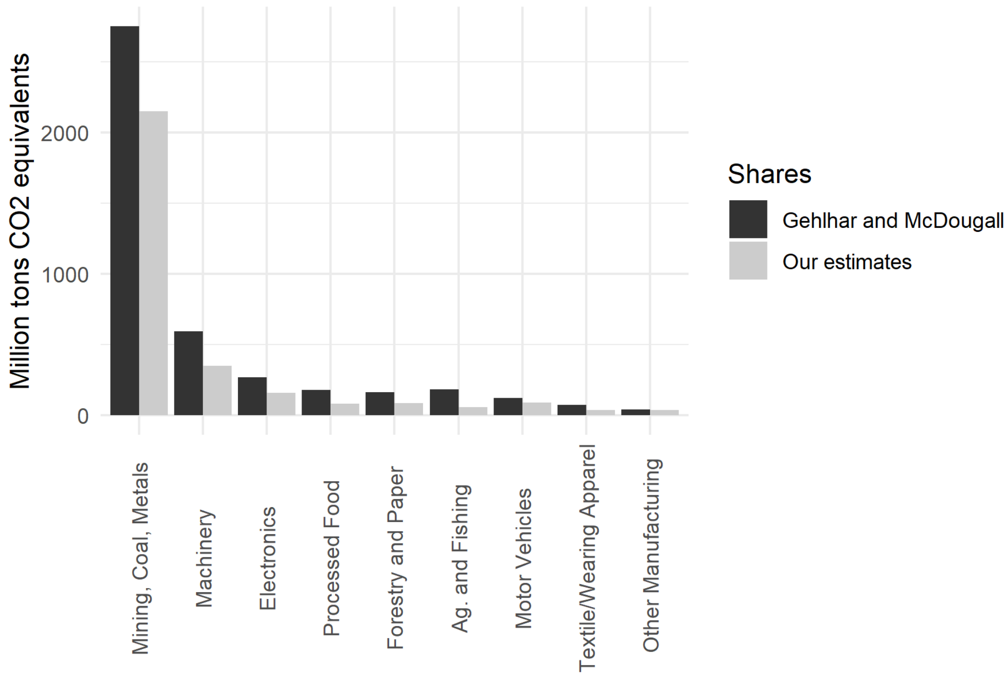

Figure 2: Trade related CO2 emissions. Gehlhar and McDougall (1997, 2006, 2016, GTAP V. 9, year 2011) vs. our estimates.

Source: Authors' calculations with data from GTAP Data Base Versoin. 9.0 and Gehlhar and McDougall (1997, 2006, 2016).

5. Conclusions

We use a comprehensive dataset on bilateral trade flows detailing modes of transport in conjunction with a set of estimated equations that transform trade flows into the dollar amount paid to each mode of transportation to build a bilateral dataset of modal transport shares that matches the country and product aggregation of the GTAP Data Base. Using such a comprehensive dataset is an improvement over previous versions that use transport modes from the US to other partners to infer these shares. Bringing new data to bear into these shares is not innocuous: we find that the older shares underestimate water transportation by a factor of two, and overestimate air transport by a factor of three. A simple conversion of these shares into CO_2 equivalents reveals that the higher incidence of air transportation services in the older dataset overestimate the CO_2 emissions corresponding to cross border transportation. The new transport mode estimates outlined in this paper were included in Version 9.1 of the GTAP Data Base, and it is expected that the methods outlined in this paper will be used to update future versions of the GTAP Data Base.

We envision future work along three dimensions. First, we think it will be fruitful to the project of estimating our own functions to update equations 4 to 7. Since, these equations were estimated using data from the United States only, and our estimates rely on the implicit assumption that the equations parameters do not vary across regions; relaxing this assumption would be a natural step to take in future research. In future endeavors, different functional forms and explanatory variables may be introduced in the estimation of the shares, for example it seems natural to expect that whether two nations share a border or are part of the same continent may affect the choice of mode of transportation and the associated transport cost. Second, given the difference in emissions between rail and road, future editions of the GTAP Data Base could consider separating rail from road so users can perform more granular analysis of transportation modes. Finally, as the data on modes becomes more accurate, one can envision a more sophisticated modeling of the market for transportation services, including opportunities to substitute across modes of transport. Although it is probably sensible to assume that such substitutability is low, the application of carbon taxes could change the structure of incentives enough to induce changes in the intensity with which transportation services are used.

References

Aguiar, Angel, Badri Narayanan, and Robert McDougall. 2016. "An overview of the GTAP 9 data base." Journal of Global Economic Analysis 1.1: 181-208. doi: 10.21642/JGEA.010103AF

Avetisyan, Misak, Nathaniel Heatwole, Adam Rose, and Bryan Roberts. 2015. "Competitiveness and Macroeconomic Impacts of Reduced Wait Times at US Land Freight Border Crossings." *Transportation Research Part A: Policy and Practice* 78 (August): 84–101. doi: 10.1016/j.tra.2015.04.027

Avetisyan, Misak, Thomas Hertel, and Gregory Sampson. 2014. "Is Local Food More Environmentally Friendly? The GHG Emissions Impacts of Consuming Imported versus Domestically Produced Food." *Environmental and Resource Economics* 58 (3): 415–62. doi: 10.1007/s10640-013-9706-3.

Avetisyan, Misak, and Thomas W. Hertel. 2015. "Impacts of Trade Facilitation on Modal Choice in International Trade." In , 1–35. 2015 Agricultural & Applied Economics Association Annual Meetings San Francisco, CA. https://www.aeaweb.org/conference/2016/retrieve.php?pdfid=1211.

Ballingall, John, and Niven Winchester. 2010. "Food Miles: Starving the Poor?" *The World Economy* 33 (10): 1201–17. doi: 10.1111/j.1467-9701.2010.01270.x.

Bertoli, S., Goujon, M., Santoni, O. (2016). "The CERDI-seadistance database", *Études et Documents* no 7, CERDI. http://cerdi.org/production/show/id/1791/type_production_id/1

Cristea, Anca, David Hummels, Laura Puzzello, and Misak Avetisyan. 2013. “Trade and the Greenhouse Gas Emissions from International Freight Transport.” *Journal of Environmental Economics and Management* 65 (1): 153–73. doi: 10.1016/j.jeem.2012.06.002.

Gehlhar, M. (1998). “Reconciling Merchandise Trade”, Chapter 11 in R. McDougall, A. Elbehri, and T. Truong (eds.) Global Trade Assistance and Protection: The GTAP 4 Database. Center for Global Trade Analysis, Purdue University.

Gehlhar, Mark, and Robert McDougall. 1997. “Chapter 15.D: Transport Margins and Mode.” In *GTAP 5 Data Base Documentation*, 15-B-1–11. GTAP Center.

———. 2006. “Chapter 15.D: Transport Margins and Mode.” In *GTAP 6 Data Base Documentation*, 15-B-1–11. GTAP Center.

———. 2016. “Chapter 9.C: Transport Margins and Mode.” In *GTAP 9 Data Base Documentation*, 15-B-1–11. GTAP Center.

Gehlhar, M., D. Gray, T. Hertel, K. Huff, E. Ianchovichina, B. McDonald, and R. Wigle. (1997). “Overview of the GTAP Data Base”. *Global Trade Analysis: Modeling and Applications*, 74-123.

Hertel, Thomas W. 1997. *Global Trade Analysis: Modeling and Applications*. Cambridge, MA: Cambridge University Press.

Hummels, David, Rasmus Jørgensen, Jakob Munch, and Chong Xiang. 2014. “The Wage Effects of Offshoring: Evidence from Danish Matched Worker-Firm Data.” *American Economic Review* 104 (6): 1597–1629. doi: 10.1257/aer.104.6.1597.

Hummels, D. (2007). “Transportation costs and international trade in the second era of globalization”. *Journal of Economic Perspectives*, 21(3), 131-154. doi: 10.1257/jep.21.3.131

Hummels, D., R. Jørgensen, J. Munch, and C. Xiang. (2014). “The wage effects of offshoring: Evidence from Danish matched worker-firm data”. *American Economic Review*, 104(6), 1597-1629.

Mayer, T. and S. Zignago. (2011). “Notes on CEPII’s distances measures: the GeoDist Database”. CEPII Working Paper 2011-25.

Peters, Jeffrey C. "The GTAP-Power Data Base: disaggregating the electricity sector in the GTAP Data Base." Journal of Global Economic Analysis 1.1 (2016): 209-250. doi: 10.21642/JGEA.010104AF

United States Energy Information Administration. (2018). Retrieved October 10 2018, from http://www.eia.gov

World Integrated Trade Solution. (2011). Retrieved July 22, 2015, from http://wits.worldbank.org/product_concordance.html

Appendix A. Obtaining weight shipped by rail and road for the US data

Our first task is to obtain estimates of shipments sent through rail and road for the US data. From the census database, we derive total imported value by other (i.e. rail or road) modes subtracting the sum of imported value by air and vessels from the total trade value:

$$(value\ by\ other)_{cus}^{hs6} = (total\ value)_{cus}^{hs6} - [\,(value\ by\ air)_{cus}^{hs6} + (value\ by\ sea)_{cus}^{hs6}] \tag{A1}$$

where the superscript $hs6$ indicates that the information is at the HS6 level, while the subscript cus denotes that we are working with data from the census. Where ground trade is not an option (inter-continental trade, for example), we set the values shipped by rail and road to zero. In a few instances within the database, the added value of shipments sent by air and water is larger than the total reported trade; in these cases we set the values of trade shipped by rail and road to zero. To obtain the value of US imports shipped by rail and road modes in the Census data, we integrate information from NATFD:

$$(value\ by\ rail)^{hs6} = \frac{(\text{value by rail})_{NATFD}^{hs2}}{(\text{total value})_{NATFD}^{hs2}}\ (value\ by\ other)_{cus}^{hs6} \tag{A2}$$

$$(value\ by\ road)^{hs6} = \frac{(\text{value by road})_{NATFD}^{hs2}}{(\text{total value})_{NATFD}^{hs2}}\ (value\ by\ other)_{cus}^{hs6} \tag{A3}$$

where the superscript indicates the HS level while the subscript denotes the dataset of origin. Note that the applied ratios vary only across chapters (HS2 level). The following assumption underlies our calculations:

$$\frac{(\text{value by road})_{NATFD}^{hs2}}{(\text{total value})_{NATFD}^{hs2}} + \frac{(\text{value by rail})_{NATFD}^{hs2}}{(\text{total value})_{NATFD}^{hs2}} = 1 \tag{A4}$$

We estimate the weight imported by rail and road as follows:

$$(weight\ by\ rail)^{hs2} = \frac{(\text{weight by rail})_{NATFD}^{hs2}}{(\text{total weight})_{NATFD}^{hs2}}\ (value\ by\ other)_{cus}^{hs2} \tag{A5}$$

$$(weight\ by\ road)^{hs2} = \frac{(\text{weight by road})_{NATFD}^{hs2}}{(\text{total weight})_{NATFD}^{hs2}}\ (value\ by\ other)_{cus}^{hs2} \tag{A6}$$

where the superscripts indicate the HS level, the subscript cus denotes that the variable comes from the census and de subscript $NATFD$ indicates that the variable was taken from the NATFD database.

Appendix B. HS-GTAP correspondence

To map data from the HS6 classification to GTAP sectors, we used the correspondence published by World Bank's "World Integrated Trade Solution" (WITS) database.[1] Table A1 presents the GTAP sector codes and their description. Some issues arose during this aggregation procedure. There are three HS6 codes that map to more than one GTAP sector: the code 121299 (other vegetable products) maps to GTAP sectors 4 and 8; the HS code 140390 (other vegetable materials of a kind used in brooms or in brushes) corresponds to GTAP sectors 8 and 13; the code 590699 (other rubberised textile fabrics) maps to GTAP 27 and 33. The weight and values associated with each of these HS codes were split equally among the corresponding GTAP sectors. The HS code 310000 does not have a correspondence in the WITS database. We assigned this code to the GTAP sector 33 because most of the codes belonging to HS chapter 31 correspond to sector 33. During the sectoral aggregation, we did not include codes from the HS chapters 98 (Special classification provisions) and 99 (Temporary modifications proclaimed).

Table A1. GTAP Sectors

Sector	Code	Description
1	PDR	Paddy rice
2	WHT	Wheat
3	GRO	Cereal grains nec*
4	V_F	Vegetables, fruit, nuts
5	OSD	Oil seeds
6	C_B	Sugar cane, sugar beet
7	PFB	Plant-based fibers
8	OCR	Crops nec*
9	CTL	Bovine cattle, sheep and goats, horses
10	OAP	Animal products nec*

Notes :nec* stands for "not elsewhere classified".

Source: With data from http://www.gtap.agecon.purdue.edu

[1] http://wits.worldbank.org/product_concordance.html, the page was retrieved for the last time in May 3rd 2019.

Table A1. GTAP Sectors (continued)

Sector	Code	Description
11	RMK	Raw milk
12	WOL	Wool, silk-worm cocoons
13	FRS	Forestry
14	FSH	Fishing
15	COA	Coal
16	OIL	Oil
17	GAS	Gas
18	OMN	Minerals nec*
19	CMT	Bovine meat products
20	OMT	Meat products nec*
21	VOL	Vegetable oils and fats
22	MIL	Dairy products
23	PCR	Processed rice
24	SGR	Sugar
25	OFD	Food products nec*
26	B_T	Beverages and tobacco products
27	TEX	Textiles
28	WAP	Wearing apparel
29	LEA	Leather products
30	LUM	Wood products
31	PPP	Paper products, publishing
32	P_C	Petroleum, coal products
33	CRP	Chemical, rubber, plastic products
34	NMM	Mineral products nec*
35	I_S	Ferrous metals
36	NFM	Metals nec*
37	FMP	Metal products

Notes :nec* stands for "not elsewhere classified".

Source: With data from http://www.gtap.agecon.purdue.edu

Table A3. GTAP Sector (continued)

Sector	Code	Description
38	MVH	Motor vehicles and parts
39	OTN	Transport equipment nec*
40	ELE	Electronic equipment
41	OME	Machinery and equipment nec*
42	OMF	Manufactures nec*
43	ELY	Electricity
44	GDT	Gas manufacture, distribution
45	**WTR**	**Water**
46	**CNS**	**Construction**
47	**TRD**	**Trade**
48	**OTP**	**Transport nec***
49	**WTP**	**Water transport**
50	**ATP**	**Air transport**
51	**CMN**	**Communication**
52	**OFI**	**Financial services nec***
53	**ISR**	**Insurance**
54	OBS	Business services nec*
55	**ROS**	**Recreational and other services**
56	**OSG**	**Public Administration, Defense, Education, Health**
57	**DWE**	**Dwellings**

Notes :nec* stands for “not elsewhere classified”.

Source: With data from http://www.gtap.agecon.purdue.edu