International Transportation Costs Around the World: a New CIF/FoB rates Dataset

Guillaume Gaulier, Daniel Mirza, Sébastien Turban, Soledad Zignago

CEPII, March 2008

1 Introduction

This paper presents a new dataset on international transportation costs around the world, set-up by the CEPII. These data are based on CIF and FOB trade values delivered by UN COMTRADE, cleaned-up and adjusted for research purposes. They are set by country pairs, at a product level (HS6 nomenclature, 5000 products), and over a ten years' period (1995-2004).

The production of international transport data is important for two main reasons. First, these data can provide a better understanding of the international transport sector by itself, its evolution, its characteristics. A common thought is that transport costs have decreased over time due to new and cheaper technologies of transport. Using transport data of US import merchandizes, Hummels(2007) showed that the global trade-weighted average transport costs have declined from 6% to 4% in 30 years. However, while air freight costs have decreased, ocean shipping costs have increased during the period. Whereas Hummels work is one of the few references one can think of, we try in this paper to bring a complementary piece of evidence on the evolution of transport costs between pairs of countries around the world.

Second, transport costs data enable a better understanding of the relation with the international geography of trade volumes and prices. As a matter of fact, distance and contiguity are used in general to cover (variable) transport costs in gravity equations. However, these variables might also proxy other factors than transport alone. For instance, they can approach other variable and fixed trading costs (extent of networks across countries, cultural differences, institutional differences, etc...). Hence, the use of good CIF/FOB rates' measures should be able to capture more specifically the impact of transport on bilateral trade. Besides, another grown interest in the literature concerns the relation between transport costs and price equations (see Baldwin and Harrigan, 2007 or Hummels and Skiba, 2005)). While Hummels and Skiba can account on directly observable transport charges using the US data, this is almost impossible to consider for other authors like Baldwin and Harrigan who end up employing instead geographical distance as proxy of transport. Further, in their seminal article on trade costs, Anderson and Van Wincoop (2003) point to the fact that another way to measure the impact of trade costs is through their effect on trade prices and more generally price indexes. This also asks for a real alternative to bilateral distance in the empirical trade literature.

In this paper, we provide a new method that produces CIF/FoB rates between pairs of countries around the world, which we detail in the next section. The produced data is set for a 10 years period (1995-2004). This methods accounts for the measurement errors that arise when computing these rates. In fact, errors in the calculation of CIF/FOB rates can be easily depicted in 2 situations : 1/ when mirror quantities (in physical units) do not match across partners ; 2/ when CIF unit values are smaller than FoB ones or when CIF unit values are too high compared to FoB ones, to be solely explained by transport and insurance costs. We also show that these undesirable situations arise either because of differences in the ***methods*** of reporting across countries or because of unintentional or deliberate ***incorrect data*** that is reported.

2 Existing Data

Two types of data already exist : Direct transport charge data and indirect ones. They have both been used in the literature.

Direct transport charges data are available at disaggregated levels and are generally considered to be of good quality. However, they are often limited to a very limited of importing and/or exporting countries/localities. For example, Hummels and Skiba(2004) use freight charges given by six specific importers (Argentina, Brazil, Chile, Paraguay, Uruguay, and the United States) in the year 1994. Besides, Robert Feenstra provides transport charges at the product level, covering a large period from 1972 and available from the National Bureau of Economic Research (NBER). Unfortunately, they are based only on US merchandize imports and are not available for other countries. Besides the US case, Hummels and Lugovskyy (2006) employ transport shipping costs for New Zealand as well. Alternatively, Limão and Venables(2001) highlight the dependence of trade costs on infrastructure using shipping company reports for the cost of transporting a standard container from the port of Baltimore to selected destinations.

The disadvantages of such data is that they do not convey information on transport costs variability between all available pair of countries. Thus, unless one is directly interested in these limited number of countries, they cannot be fully used in cross-country gravity and price equations studies.

Due to limited data of transport charges around the world, researchers turn to indirect measure of trade costs based on CIF/FoB ratios. A given trade flow is counted twice through the customs' declarations of the exporting and importing countries. The exporter usually declares a Franco on Board value (FoB), which is the value at the exporter's border. On the other hand, the importer declares a CIF mirror value which includes additional Costs of Insurance and Freight. The ratio of the two values provides what we call usually a CIF/FoB ratio.

Aggregate CIF/FoB rates are publicly available for many countries and years (1948 to present), from the IMF Direction of Trade Statistics. The IMF ratios are the most widely used estimates of transport costs for international trade. After gathering disaggregate CIF/FoB rates from the COMTRADE United Nations dataset, the IMF apply aggregation procedures to produce readily operational measures of transport costs.

More disagregated data are obtainable, however, from UN COMTRADE directly. COMTRADE provides bilateral trade data in physical quantities and US dollar values at the product level (6 digit Harmonized System Nomenclature, over 5000 products) for more than 200 countries in the world. The data are currently available from 1989 to 2004. When both countries declare a bilateral flow, COMTRADE provides the declaration of the importer in CIF which is easily matched with that of the exporter in FOB, from which a CIF/FoB ratio is extracted. It turns out, however, that the implied CIF/FoB rates are of questionable quality.

Hummels and Lugovskyy (2006) studied to which extent the CIF/FoB ratios obtained from the IMF and subsequently, COMTRADE, can signal true transportation costs. Based on a comparison made between IMF aggregated data and the two US and New Zealand datasets, they argue that CIF/FoB ratios cannot be usable as such. If anything, they could only be exploited as proxy of transport costs as long as one looks at variations across exporters. However, these measures appear to be helpful when considering fitted values of CIF/FoBs against plausible correlates revealing true transportation costs such as geographic distance. Time series do not seem to bring additional information, however. Turning to commodity-level data provided by UN COMTRADE, these authors do not find encouraging results, however. First, only 10% of the CIF/FoB ratios lie in the plausible range (1-2) (i.e. 0 to 100% of ad-valorem transportation costs). Besides, they find a negative correlation with Feesntra's US imports dataset. They conclude that it would be 'unwise' to exploit cross-commodity variations to signal transport costs.

It is important to note first, that the product-level analysis of these authors is undertaken over an early period (1974-1983), and the quality of the data might have been scaled-up since then. Second, consider two countries i and j trading a given quantity q_{ij}. Let us note p^X_{ij} and p^M_{ij} to be respectively the FOB price declared by the exporter and CIF importing price delivered by the importer of the same quantity flow. Symmetrically, let us denote declared quantities by the exporter and importer respectively to be q^X_{ij} and q^M_{ij}. It comes that the usually computed CIF/FOB ratio can be expressed as $CIF/FOB = \frac{p^M_{ij} q^M_{ij}}{p^X_{ij} q^X_{ij}}$. This CIF/FOB expression has the shortcoming of reporting errors of two types, however : a/ a 'value of shipment error-type' due to differences in recorded total values by the importer and exporter ($p^M_{ij} q^M_{ij} \neq p^X_{ij} q^X_{ij}$) ; and b/ a 'physical quantity of shipment error-type' due to differences in quantities recorded ($q^M_{ij} \neq q^X_{ij}$). For instance, a given number of tons of women trousers' recorded by the Chinese authorities as exports to say, Germany, might not be actually matched by German records who could report a smaller (or higher) figure. This ends up underestimating (resp. overestimating) CIF/FoB ratios. Fortunately, quantities are observable from exporters and importers reports in COMTRADE data. One way to account for differences in quantities is thus to compute CIF/FoB ratios where each CIF and FOB valued flow is normalized by the corresponding quantity reported by the importers' and exporters' customs respectively. By doing so, one obtains an alternative ratio of CIF/FoB *based on unit values* that we can call CIFu/FoBu (i.e. $CIFu/FoBu = \frac{p^M_{ij} q^M_{ij} / q^M_{ij}}{p^X_{ij} q^X_{ij} / q^X_{ij}}$).

We have just seen that errors on quantities and values are correlated by construction, and this can be partially controlled for by using CIFu/FoBu ratios. However, one can still imagine that quantity and value-error types are correlated because people who badly register quantities could also badly report values, independently from quantity reports. Further and more plausibly, the differences in mirror quantities *and/or* values might come from cross-country differences in the methods of accounting at the borders.

In this paper, we aim at providing a method to produce consistent CIF/FoB rates. This method is undertaken in 3 steps. First, we pick only those flows where quantity declarations from the importer and the exporter are similar. In such a way, we are removing the quantity-error type. It appears that less than 11% of the dataset pass this procedure. Nevertheless, this still represents nearly 2 million observations. Hence, for the corresponding flows, we are left out with only value-error type to be treated. Second, based on these selected data, we employ an econometric method to condition out the remaining measurement errors. To do so, we use a new series of control variables informing about the method of data reporting by the partners. These variables are based on survey responses to a COMTRADE questionnaire submitted to each country's reporting authority. The survey objective was to see by how much the national authorities comply with COMTRADE's recommendations in flows registrations. Accounting for these measurement biases, the econometrics can now predict new CIF/FoB ratios, using factors such as distance, transport infrastructure, etc... which should affect true transportation costs. In a third and final step, we suppose that transport costs behave similarly for those observations that did not pass the first step procedure (i.e. where quantities from exporters and importers did not match). Under this hypothesis, we reproduce a CIF/FoB estimates for all of the out of sample data by using the insample's coefficients.

3 Crude CIF/FoB Ratios and quality of matching

Errors of CIF/FoB calculation arise in the data not simply because of mis-measurement but because of differences in registration methods across countries. For instance, if one country does not comply with UN recommendation to register the incoming flows as CIF flows (including transport and insurance costs), and still declares flows as FoB, the CIF/FoB ratios would then tend to be biased downward as they approximate unity. Pick another example : the UN recommends that each importing country declares each incoming flow as coming from the country of production of the flow (or Origin country), not from a transit country. If one country A happen to declare goods coming from a country B, but that these goods are only transiting through B (they are sourced from say, a country C), a mismatch might appear then between declarations of A and those of B on one hand, and between A and C declarations on the other hand.

Before looking at what drives differences in registration across countries, let us look at the extent of these differences.

	All observations		1<x<2		w>0.9	
	CIF/FOB	CIFu/FOBu	CIF/FOB	CIFu/FOBu	CIF/FOB	CIFu/FOBu
Median	1.007	1.002	1.23	1.19	1.02	1.03
10th percentile	0.12	0.0009	1.02	1.02	0.87	0.89
90th percentile	8.26	5.06	1.75	1.7	1.42	1.41

TAB. 1 – Comparisons between CIF/FOB and CIFu/FOBu rates in COMTRADE

Table 1 presents some descriptive statistics regarding the freight rates constructed measures. As in Hummels and Lugovskyy, we expect the whole distribution to lie in a (1-2) (i.e. the minimum rate should be higher than unity and the maximum rate should not exceed the value of the good transported). The first column gives an idea about the distribution of CIF/FoB measures, when all observations are considered. Although the median CIF/FoB is slightly higher than 1, many observations fell far below 1 and far above 2. As one can deduce from the figures, 80% of the data lie in a (0.12- 8.26) range where the extremes far exceed the "normal" range one could think about. Without any correction, obviously these data are not usable as such.

In a second column, we present the alternative CIFu/FOBu measure. The 9th decile decreases, but the 1st as well. However, both tails of the distribution are still rather extreme. Columns 3 and 4 retain only those values within the (1-2) range (i.e. 30% of all observations). However, although rates lie in the expected range, there is no reason to pretend that the related transport costs are usable in this range. To understand why, suppose that the true freight rate of a good shipped from one country to another is 10%. If, for some reason, the CIF price is reported as being 90% higher than the FoB price however, the ratio would still be in the 1-2 range of ad-valorem transport costs (CIF/FoB=1.9) while obviously the freight cost would be reported with an error of 80 percentage points higher than the true rate.

Columns 5 and 6 report finally the same freight rates statistics for those goods which physical quantity from the exporter declaration better matches that of the importer. More precisely, we allow for differences in measurement between the two declarations not to exceed 10%. We have defined for that purpose an indicator of quality of reporting quantities across partners which is $w = Max(QM, QX)/Min(QM, QX) > 0.90$. The implied freight rates appear then to be more in line with our beliefs. Although still relatively low, the median values of the implied freight rates are now around 2 to 3%. Besides, the distribution around this value is much smaller with 80% of the values roughly lying within the range 0.87-1.42. Although, the lower-extreme is still below 1, the range is far more satisfactory than previous columns, where the quality of reports was not yet controlled for. Besides, the 90th percentile is now around 1.40 which is also a plausible figure for those merchandizes with the highest cost of freight.

4 Comparison with Feenstra's US imports database

Next, we compare the crude UN COMTRADE CIF/FOB rates with those that can be obtained from Feenstra's dataset (also available from the NBER). Feenstra's data reports bilateral values of merchandises reaching the United States customs along with 'true' freight charges, directly reported by the customs, at the HS 6-digit product level. In order to compare both datasets, we restrain COMTRADE to those registered flows heading to the US market. After adjusting for ISO codes (i.e. France's code is 250 for instance in Feenstra and 251 in Comtrade), and removing Luxembourg and Panama along with 7 HS products, we found that more than 95% of the registered flows in COMTRADE match Feenstra's, while more than 90% of the registered flows in the latter matches Comtrade's. We have then compared CIF and FOB values across the two datasets. CIF values on one hand and FOB values on the other hand were very similar across the datasets (correlation of almost 0.999 and 0.987 respectively).

Finally, we compared the CIF/FOB ratio (i.e. ad valorem transport costs) from COMTRADE with its corresponding measure in Feenstra (i.e. 1+(Freight/FOB)). Transformed into logs, the correlation is around 0.45. But the more weight is given to flows with low discrepancies between the quantity reported by the exporter and that reported by the importer in COMTRADE, the higher the coefficient is. In particular, for those goods where measurement error between the flow reported by the exporter and its mirror is less than 10%, the correlation coefficient reaches 0.85. As it seems, implied CIF/FOB measurement errors might be conditioned out, at least partly, by using those observations where physical quantities match.

5 Sources of reporting differences

In order to investigate the extent of compliance with UN recommendations, UN COMTRADE have asked each reporting country to respond to a series of questions that have been uploaded on their website. Among these questions, the following question codes concerned directly or indirectly the procedure of registration regarding incoming and outgoing flows.

– Q106 : Do you use customs declarations as a source ?

– Q117 : Is the exchange rate used for currency conversion that which is in effect at the time of exporting or importing ?

– Q143 : Do you use a standard unit of weight for quantity measurement of all commodities where applicable ?

– Q148 : Do you use units of weight on a net basis (e.g. excluding packing) ?

– Q61-64(Qexp) : As an exporter, do you declare the importer as Last Known Destination ?

– Q58-60(Qimp) : As an importer, do you declare the exporter as Origin ?

The UN recommendation is to answer 'YES' to all of these questions.

Table 7 shows descriptive statistics on these questions. Percentages given are percentages of flows where the corresponding UN recommendation is followed. Nearly all countries appear to use customs' declaration as a source. More problematic is the cross-country heterogeneity in the date of application of the exchange rate, which serves to convert all incoming and outgoing product values to and from a given country into US dollars. This is not without introducing differences in mirror values, which has nothing to do with personnel skills in registering flows at the borders.

TAB. 2 – COMTRADE Questionnaire

Country	COMTRADE Questionnaire	code	freq
exporter	Use Customs as a source of data	Q106i	98,84%
importer	Use Customs as a source of data	Q106j	97,85%
exporter	Exchange rate in effect at date of exports	Q117i	63,50%
importer	Exchange rate in effect at date of imports	Q117j	64,23%
exporter	Declares a Standard unit of Weight	Q143i	67,09%
importer	Declares a Standard unit of Weight	Q143j	70,36%
exporter	Net weight declaration (no packaging included)	Q148i	87,30%
importer	Net weight declaration (no packaging included)	Q148j	86,50%
Exporter	Declares importer as Last Known Destination	Q dec export	77,05%
Importer	Declares exporter as Origin producer	Q dec import	90,45%

TAB. 3 – Frequency of countries meeting UN Recommendation

Standard units of weights where applicable are not effectively applied everywhere neither. Around 60% of the respondents answer positively to this question. Hence, two partners reporting different units of weight bias mechanically the CIF/FoB ratio. Besides, net weight declarations (excluding packages) are not always followed. Only 87% of the sample data comply with this recommendation. This heterogeneity of weight declarations (in the standards and inclusion/exclusion of packages) increases the importance of using 'matched quantities' data when estimating new CIF/FoB ratios.

Finally, another relevant result to be noted is that though most importers declare the exporter to be the origin of the product (90%), a significant part of exporters seem to declare the importer not to be necessarily the last country of destination, but might be a transit country (23%). This again should affect registered quantities and values of traded goods.

6 Estimated equation

As already mentioned, because CIF/FoB rates directly computed from UN COMTRADE data appear to be problematic because of measurement errors and differences in the methods and timing of reporting across countries, we apply an econometric technique that tries to handle all these problems. In a first step, econometrics are applied on the "representative" 2 million observations sub-sample[1]. In a second step, the estimated coefficients are then used to produce consistent CIF/Fob rates for the data out of the sub-sample.

Here, we use a linear regression model of CIF/FoB rates to clean errors and differences in reporting methods while conserving that part only corresponding to real transport costs. A simple way to do so is to run a transport cost regression where right hand side variables convey information about the true transport costs, while leaving in the residuals the rest of the variance. Based on the results, one can easily re-construct a vector of (estimated) transport costs. We have run the following regression :

$$\begin{aligned}\left(\frac{CIFu}{FoBu}\right)_{ijkt} = \; & \beta_0 + \beta_1 log(dist_{ij}) + \beta_2 log(dist_{ij}^2) + \beta_3 contig_{ij} + \beta_4 comlang + \beta_5 colony \\ & + \delta log(UV)_k + \theta_1 GDP_i + \theta_2 log(GDP_j) + \theta_3 log(GDPpc_i) + \theta_4 log(GDPpc_j) \\ & + \phi_1 log(infra_i) + \phi_2 log(infra_j) + \phi_3 landlock_i + \phi_4 landlock_j + \lambda_t + \varepsilon_{ij,k,t} \quad (1)\end{aligned}$$

This equation follows more or less that of Limao and Venables(2000), with variables having bilateral characteristics (like distance or contiguity), variables with characteristics specific either to the exporting or the importing country and variables informing about cross-product variation.

- Bilateral characteristics variables : We use distance and squared distance, along with contiguity, common language and colony dummies. Because we want to test for a more general relationship between transport costs and geographical distance, we have introduced the latter together with its square. Besides, the first two reported variables are directly related to transport while it is indirectly the case for language and colony variables. As a matter of fact, the latter are more related to networks. Networks grease the wheels of trading goods and thus transporting those goods. In fact, one can imagine transport networks to be more developed between countries with higher propensity to communicate due to common historical links. With the colony and common language variables we are then capturing 'bilateral' transportation networks.
- Product-type variable : Product related variables are hard to obtain since we are looking for data available at a highly disaggregated levels. We follow some authors by defining a ratio of unit value to weight variable UV_k, in order to capture product specific features in our equation. Hummels(2007) computes a weight-to-value ratio for each flow and uses it as an explanatory variable for the flow's implicit freight rate. Anderson and Van Wincoop(2004) explain that "High value-to-weight" goods are less penalized by transport costs. Indeed, if the FoB price is higher, we can assume that transport costs represent a smaller part of the global CIF price. Clark, Dollar and Micco(2004) use value per weight as a proxy for insurance costs. They justify the use of product-variables by the fact that the content of trade may explain freight rates' differences across countries.

  CIF/FOB measures deliver information about both types of costs. Unless very strategic, rare or precious goods are transported, one can easily imagine that true

[1]The whole dataset at hand includes more than 18 Millions of observations.

CIF/FoB measures are more closely related to the value of transport than that of insurance, however. Hence, following Anderson and Van Wincoop prediction and Hummels results, we expect our UV_k variable to be negatively related to CIF/FoB ratio.

- Economies of scale and congestion variables : We also include GDP variables to account for economies of scale and congestion factors reported in the literature. Clark, Dollar and Micco(2004), Bloningen and Wilson (forthcoming), and others, address the issue of economies of scale and/or congestion as determinants of transport costs. Two opposite effects can affect transport costs. Countries where activity is large enough may have lower transport costs due to economies of scale due to filling-up the containers. As a matter of fact, if fixed costs exist and are substantial in transport costs, we can expect economies of scale, since full containers would be obviously more profitable than an empty one. Increasing returns to scale are typical of maritime transport. Clark, Dollar and Micco (2004) quotes Alfred Marshall : "A ship's carrying power varies as the cube of her dimensions, while the resistance offered by the water increases only a little faster than the square of her dimensions". Kymer(1999) showed that a ship owner operating a vessel carrying 200 Twenty-foot Equivalent Unit (TEU) containers will pay 70$ per container for the channel access which leads to the port of Buenos Aires ; however, if he operates a 1000 11 TEU vessel, the rate is only 14$ per container. But high activity increases time costs induced by congestion. This, in turn may also raise final transport costs. Bloningen and Wilson address the congestion issue using the difference between exports and imports in order to take into account the emptiness of containers in one of the directions. They also use traded volumes to seize congestion and economies of scale effects. Bloningen and Wilson find that congestion effects are slightly stronger than economies of scale.
- Infrastructure variable : we follow closely here Limao and Venables(2000) by introducing in our equation some road infrastructure measures taken from the Worldbank. These authors show that infrastructure may explain 40% of transport costs for coastal countries and 60% for landlocked countries.
- Level of development : GDP per capita has been widely used as a proxy of transport costs, but often for different reasons. Limao and Venables(2000) uses both GDP per capita and infrastructure and find that both reduce transport costs when their quality improves. Inversely, Clark, Dollar and Micco(2004) use GDP per capita as a proxy of infrastructure, and more precisely of port efficiency. We use it here to complement infrastructure variables hoping to increase the fit of the regression in order to obtain eventually better predictions of CIF/FoB ratios.

Now, some can correctly argue that many of these variable might not only deliver information about transport but also about differences in reporting treatments across countries along with measurement errors. For instance, the more distant from a importing port is an exporter, the higher the probability that the the merchandize crosses other countries' ports before arriving to final destination. This means that distance, although correlated to the cost of transport can also be correlated to the propensity of reexports. To the extent that some countries register some flows coming from the last country receiving the merchandize but not from the origin country of the latter, distance can then be correlated to the residual of our equation. Let us give another example. The quality of transport infrastructure should reduce transport cost on one hand but can also be correlated to the quality of reporting flows at the custom borders. Then, infrastructure might not only capture transport cost but also measurement errors.

Hence, in order to avoid such critics, we add to the CIF/FoB relationship a series of dummy variables to account for measurement errors and differences in methods of reporting trade data. These dummies come from the results to the questionnaire mentioned above.

7 CIF/FOB ratios : regression results

The results of the regressions are given in the following table. The dependent variable is the ratio of unit values. As mentioned, estimations are run on observations where bilateral quantities match (with a 10% error margin).

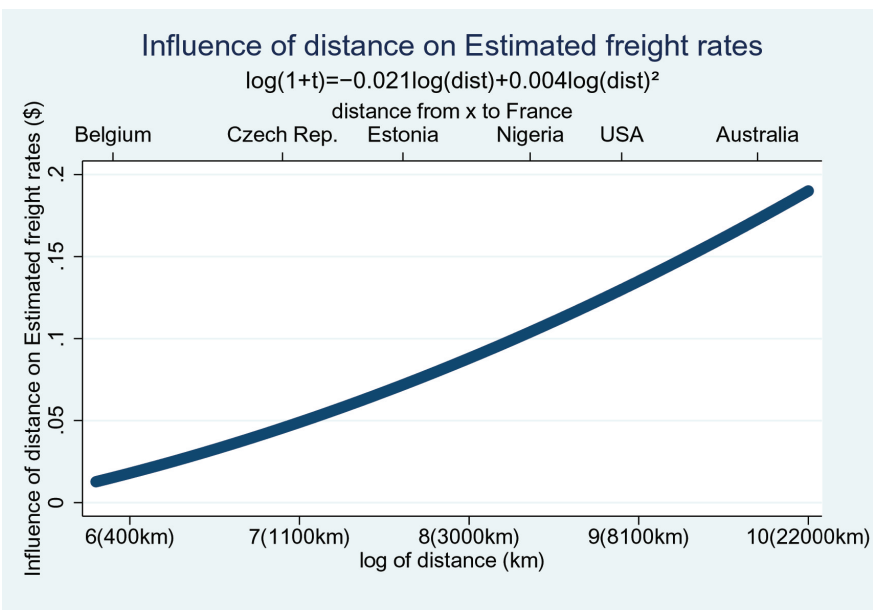
As a benchmark, we run first a regression similar to the one served to harmonize mirror trade data in BACI's working paper (BACI WP), except that variables for common language and colonial ties are added. All parameters are significant at 1% and have the expected sign. It is important to note however that the threshold distance above which the relationship with transport costs becomes positive is 860 km, which is intuitively a high figure. At small distances however (less than 860 km), the relationship is negative ! This surprising result might be due to a misspecification of the econometric equation (or an omitting variables problem), however. For instance, at small distances transport infrastructures might be more developed which ends up reducing transport. Not accounting for infrastructure in the equation biases the coefficients on the two distance variables.

In the second and third columns, we add variables of infrastructure, GDP and GDP per capita. Once accounting for GDP per capita (column 3), higher infrastructure in both countries reduces transport costs between those countries although by small amounts (elasticities of .002 and .009 respectively). Besides, all things being equal, a higher activity in the export or import country reduces transport costs between them. Here, higher activities seem to bring about lower costs of shipping goods due to economies of scale in the transport sector. Note also, that the importer GDP per capita increases transport costs, possibly due to higher demand effects. However, the fact that GDP per capita of the exporter reduces them is more related to a supply side story : the level of development and the quality of transport supply can go hand in hand explaining the negative effect on the cost of transporting goods. Finally, notice that after controlling for infrastructure and GDP type variables, the coefficient on the two distance variables have changed, reducing by half the coefficient on the log of distance and that of its square. As a result, the impact of distance becomes now positive on transport, only after a distance of 180 km.

Column 4 presents the results of an augmented specification where we control for errors and differences in methods of reporting, using the UN questionnaire. First, the questions are all highly significant and complying with the UN recommendation either reduce or increase the value of transport costs. As a matter of fact, most coefficients values and signs are not interpretable as such because there is no expectation about the direction and extent of the bias when complying with the UN recommendation. One exception however arises for the responses to the last two questions about declaring the importer as the last country of destination or the exporter as the origin producer country. In fact, one expects that when the importer declares the merchandize coming from a country farer away than a transit country (usually located between the two true trading countries), transport value should be higher. One also expects transport costs to be higher when the exporter declares goods to be shipped to the last destination country. And indeed, these two expectations are verified in our econometrics. The two correposnding declarations that comply with the UN are increasing the value of the CIF/FoB ratios.

Second, the threshold distance above which transport is increasing with the latter is now almost 0 (17 km only...). Below, we present a graph illustrating the impact of distance from France on transport costs. While distance from Belgium increases transport costs by 0.02 percentage point, distance from Australia to France increases 10 times more (around 0.2 percentage point).

In the last column (column 5), we run the same regression on the full sample. The results are qualitatively similar to those for the good quality subsample data. However, the value of the parameters differs. As we are more confident about the coefficient values of the good quality matching sample where $w > 0.9$, we prefer using the coefficients of column 4 to reproduce a new vector of CIF/FoB for all of the data at hand.



TAB. 4 – Regression Estimates

	Dep. Variable : Ratio of unit values				
	$w = Min(qx, qm)/Max(qx, qm) > 0,9$				All Observations
Name	*Baseline*	*infra&GDP*	*&GDPpc*	*with questionnaire*	
Intercept	0.535***	0.623***	0.499***	0.108***	0.009
dist(log)	-0.138***	-0.072***	-0.066***	-0.021***	0.04***
dist (log squared)	0.01***	0.007***	0.006***	0.004***	0
Median Unit value	-0.036***	-0.033***	-0.032***	-0.032***	-0.047***
contiguity	-0.035***	-0.031***	-0.03***	-0.019***	**0.002**
landlocked_exp	0.01***	-0.005***	0.004***	0.015***	0.001
landlocked_imp	0.003***	0.012***	0.005***	0.003**	0.039***
comlang_off	0.016***	0.013***	0.017***	0.006***	0.038***
colony	-0.03***	-0.03***	-0.028***	-0.015***	-0.062***
infra_exp		**-0.023*** **	**-0.009*** **	**-0.009*** **	-0.036***
infra_imp		**0.027*** **	**-0.002*** **	**-0.007*** **	0.023***
GDPpc_exp			**-0.043*** **	-0.047***	-0.095***
GDPpc_imp			**0.069*** **	0.07***	0.117***
GDP_exp		-0.01***	-0.006***	-0.005***	0.002***
GDP_imp		-0.004***	-0.012***	-0.011***	-0.029***
EXP use customs as data-source				-0.012***	0.063***
IMP use customs as data-source				0.219***	0.287***
EXP use Exch rate at time of exp				-0.013***	-0.024***
IMP use Exch rate at time of imp				0.013***	0.058***
EXP Use std unit of wght				0.021***	0.071***
IMP Use std unit of wght				-0.023***	-0.033***
EXP Use *Net* units of weight (no pack)				-0.044***	-0.181***
IMP Use *Net* units of weight (no pack)				0.009***	-0.026***
EXP Declares last country of dest				0.009***	0.145***
IMP Declares country Origin				0.003***	-0.012***
Year effect	yes	yes	yes	yes	yes
N	1718904	1533224	1501726	1501726	11429133
R^2	0.013	0.021	0.024	0.029	0.02
Threshhold dist(km)	**862.842**	**252.825**	178.218	**17.291**	0

8 Illustrative graphs

In this section, using our 'new' measure of CIF/FoB we present some evidence about the behavior of these values across sectors and countries. We expect now the constructed CIF/FOB rates to be higher than those shown in table 1 with a lower variance. Ideally, one would expect the whole distribution to be higher than 1. Tha table below show some statistics for the Insample and outsample distribution. Although the median for the outsample remains low, the insample distribution has a median of CIF/FoB rates around 2%, which is a significant improvement compared to what have been shown in table 1. Besides, the variance of the insample but also the outsample distributions is now much more plausible with 80% of the data lying around the range (0.91-1.14).

Also, we expect transport costs to be higher for Mining and Quarrying sectors than for manufacturing sectors, in particular due to higher weights and more specialized vessels for the former than the latter. Besides, we expect countries that are more remote from the rest of world markets to have on average higher transport costs than countries that are more central to world markets.

The graphs below illustrate very well these expectations. We do find that non-container shipping that prevails in most quarrying and mining industries are more than two times higher than those other goods (average CIF/FoB ratios around 10%) . In turn, fresh goods and other commodities (agricultural, fishing) appear to have 2 percentage points higher transport costs than those of manufacturing, perhaps because of the costs of processing and refrigerating these goods as they are transported.

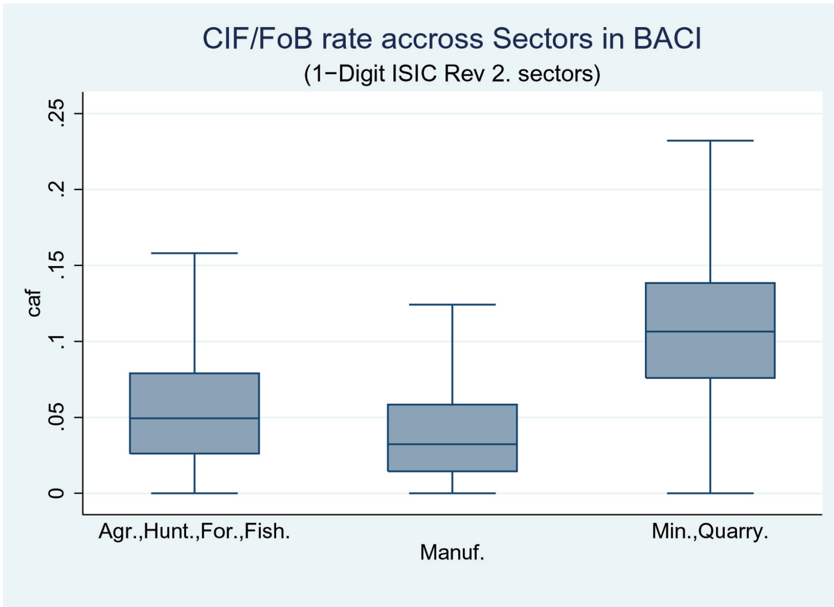
Besides, one graph gives the average CIF/FoB ratios for some exporting countries of interest. Again, as expected, centrally located and rich exporting countries bear less transport costs than poor and/or remotely located countries.

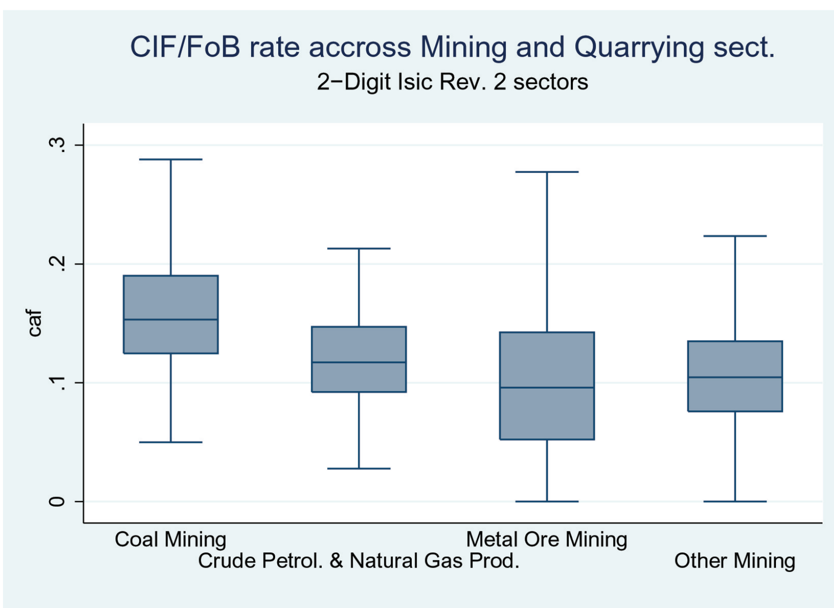
	Insample	**Out**sample
	CIF/FOB	CIF/FOB
Median	1.02	1.003
10th percentile	0.93	0.91
90th percentile	1.14	1.12

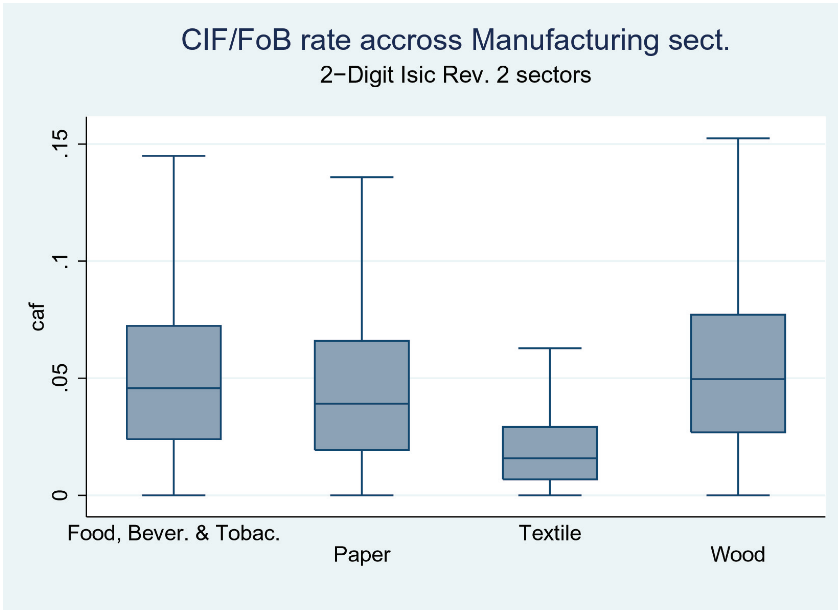
TAB. 5 – Estimated CIF/FOB rates **In** and **Out** of sample

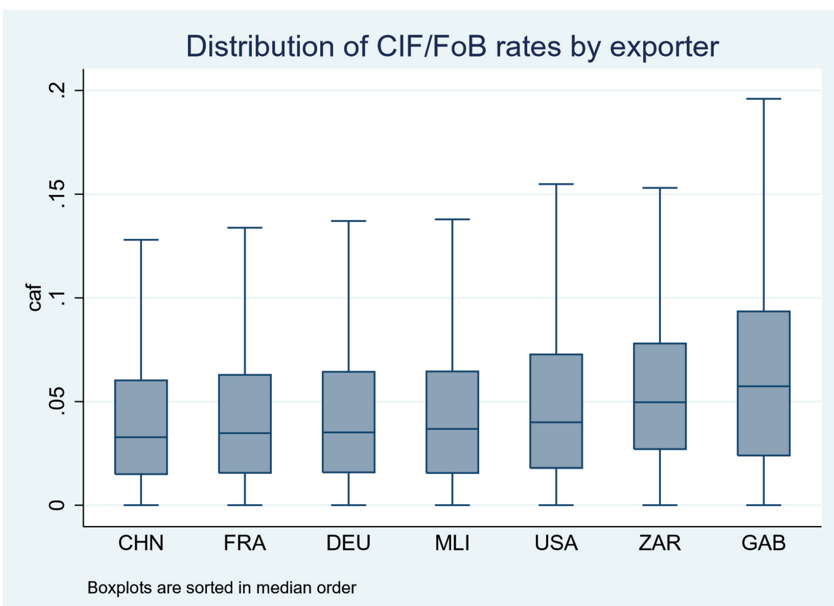
9 CIF/FoB ratios in gravity regressions

Another way to verify whether or not our CIF/FoB constructed measure is good enough to be trusted is to include it in gravity equations of trade. We want to see first whether its coefficient come closer to the theory coefficient. As a matter of fact, theory predict that the coefficient on transport costs represents the elasticity of substitution. In the literature, the elasticity of substitution has been estimated in many alternative ways to approach values between 5 and 20 (Hummels, 2001 ; Head and Ries, 2003, Erkel and Mirza, 2001, Hanson (2001), Trefler and Lai, 2002, etc...).


CIF/FoB rate accross Sectors in BACI
(1−Digit ISIC Rev 2. sectors)
caf
0
.05
.1
.15
.2
.25
Agr.,Hunt.,For.,Fish.
Manuf.
Min.,Quarry.


CIF/FoB rate accross Mining and Quarrying sect.
2−Digit Isic Rev. 2 sectors
caf
0
.1
.2
.3
Coal Mining
Crude Petrol. & Natural Gas Prod.
Metal Ore Mining
Other Mining


CIF/FoB rate accross Manufacturing sect.
2−Digit Isic Rev. 2 sectors
caf
0
.05
.1
.15
Food, Bever. & Tobac.
Paper
Textile
Wood


Distribution of CIF/FoB rates by exporter
caf
0
.05
.1
.15
.2
CHN
FRA
DEU
MLI
USA
ZAR
GAB
Boxplots are sorted in median order

Next, we want to see whether including it instead of including distance and contiguity, brings at least as much information as the latter to explain trade flows.

The next table presents a flavor of the results. We ran first a series of regression based on a random sample (250,000 observations). We have chosen random samples for computation purposes : as we include product fixed effects, SAS software could not perform regressions on the whole sample. In a second series of regressions, we could work though on the whole sample by running regressions where we transform the variables in deviations from mean exports by HS6 product.

Column (a) is a benchmark gravity equation where we include the observed CIF/FOB ratio (or the directly computed ratio from COMTRADE). The impact of this measure is negative on trade with a coefficient of -0.53. When we introduce our constructed measure of CIF/FoB ratio in column (b) instead, the coefficient is 6 times higher in absolute values and reaches -3.65. This is much more in line with the literature that estimate elasticities of substitutions. Column (c) considers only the data on which was primarily estimated the CIF/FoB ratios. These are the data where the matching of quantities between the two declaring partners is of good quality. We call this subsample the insample data. Here, the effect is even higher. The coefficient reaches -5. The performance of our variable on the outsample however, proves to be relatively good as column (c) mentions a coefficient around 3.4.

In the second series of regressions, we transform variables in a way that we can now use all the dataset at hand and control implicitely or explicitely in the equation for product, exporter and importer fixed effects. The value of the coefficient on the CIF/FoB variable climb to around -15, whether all the sample, the insample or the outsample are considered respectively. These values, although pointing to high elasticities of substitution, can actually still be compared to some of those provided in the literature.

Next, in order to compare the information conveyed by CIF/FOB ratio to explain trade flows with that of the other variables usually considered in gravity equations (Distance, contiguity), we run alternative regressions by inserting the latter while removing our CIF/FoB measure (see columns (h) and (i)). The obtained R2 appear to be similar whether using our CIF/Fob measure or its alternatives together. This is encouraging as it mention that using our measure of transport or distance and contiguity together brings about the same additional information to explain trade. One of the advantages of our measure however is that it is varying over time and products and thus can be easily introduced together with including dyadic effects in a gravity equation.

Bilateral FOB export regressions using alternative measures of CIF/FOB ratios

	Random Sample (250,000 observations)				*ALL Sample (Deviation from average bilateral exports per product)*				
	(a)	(b)	(c) insample	(d) outsample	(e)	(f) insample	(g) outsample	(h) insample	(i) outsample
Intercept	-59.18 (2.02)***	-53.77 (2.2)***	-63.73 (2.12)***	-49.38 (2.19)***	-0.97 (0.06)***	-1.71 (0.2)***	-0.86 (0.07)***	-3.00 (0.2)***	-2.07 (0.07)***
Log obs. CIF/FOB	**-0.53 (0)*** **								
Log estim. CIF/FOB		**-3.64 (0.07)*** **	**-5.02 (0.07)*** **	**-3.41 (0.07)*** **	**-14.14 (0.02)*** **	**-15.62 (0.05)*** **	**-13.61 (0.02)*** **		
Log GDP exporter	8.15 (0.07)***	6.71 (0.08)***	7.33 (0.08)***	6.87 (0.08)***					
Log GDP importer	10.69 (0.05)***	10.92 (0.06)***	13.38 (0.06)***	10.52 (0.06)***					
Log Distance								-0.54 (0)***	-0.53 (0)***
Log contiguity								0.51 (0)***	0.48 (0)***
Log Common Language								0.16 (0.01)***	0.12 (0)***
Log Colony								0.02 (0.01)***	0.14 (0)***
time effect	yes	yes	yes	yes	yes	yes	yes	yes	yes
product effect	yes	yes	yes	yes	yes	yes	yes	yes	yes
exporter effect	no	no	no	no	yes	yes	yes	yes	yes
importer effect	no	no	no	no	yes	yes	yes	yes	yes
Observations	250,000	250,000	250,000	250,000	7.59E+07	8.47E+06	6.62E+07	8.35E+06	6.51E+07
R-Square	0.37	0.26	0.36	0.24	0.20	0.29	0.20	0.29	0.20
Root MSE	2.00	2.17	2.10	2.16	2.10	2.03	2.10	2.02	2.08

*Note: figures between parentheses are standard errors. *** designates 1% significance*
(insample) designates the sample on which 1st stage CIF/FOB regressions, have been performed. Basically, these are observations where declarations of quantities among bilateral partners are matched
(outsmaple)= designates sample where CIF/FOB where inferred using regressions results from the insample

Bibliography

Anderson, J. and Eric van Wincoop (2003) : "Gravity with Gravitas : A Solution to the Border Puzzle", *American Economic Review*, vol. 93, No. 1 (Mar., 2003), pp. 170-192

James E. Anderson and E. van Wincoop (2004), "Trade Costs", *Journal of Economic Literature*, vol. 42(3), pages 691-751, September

Baldwin, R. and J. Harrigan (2007) : "Zeros, Quality and Space : Trade Theory and Trade Evidence", *NBER Working Papers 13214*

Blonigen, Bruce A., and Wesley W. Wilson : "Port Efficiency and Trade Flows." Review of International Economics, forthcoming, *Review of International Economics*

Clark, X., D. Dollar, and A. Micco. 2004. "Port Efficiency, Maritime Transport Costs and Bilateral Trade" *NBER Working Papers 10353*

D. Hummels(2007) : "Have International Transportation Costs Declined ?", *Journal of Economic Perspectives*, 21 (3)

D. Hummels and V. Lugovskyy(2006) : "Are Matched Partner Trade Statistics a Usable Measure of Transportation Costs", *Review of International Economics*, Feburary

D. Hummels and A. Skiba(2004) :"Shipping the Good Apples Out ? An Empirical Confirmation of the Alchian-Allen Conjecture", *Journal of Political Economy*, vol.112, 1384-1402

N. Limao and Anthony J. Venables (2000) : "Infrastructure, Geographical Disadvantage and Transport Costs", Columbia University, London School of Economics working paper