On the relative importance of iceberg and additive transport costs in international trade

Note: Second-stage regression

In this second stage regression, the key question is: Do the proxies frequently used as measures for trade costs, typically distance, correspond more to additive cots or to multiplicative trade costs? Or total? Usually, people use a measure of distance as a way to capture iceberg trade costs. Using our decomposition of transport costs, we want to assess the contributions of such proxies to each component (additive / iceberg). What is the size of the error we make (what is the size of the approximation) when we suppose that distance is a good proxy of (total) transport costs?

This question drives our second-step analysis. Consider first our "raw" estimate of trade costs (meaning, only modeling a multiplicative form); we estimate the role of distance (as quite frequent) and found that this first reasonably well, the estimated coefficient stands in the average. We then apply the same regression on each multiplicative / additive component of trade costs. Suppose that we find that we find a high correlation with the additive cost, almost 0 with the multiplicative cost. This means that, using this proxy to capture the size of iceberg trade costs amounts making a sizeable approximation error.

To investigate this point, we thus regress 1) the "overal trade costs" (i.e.), modeled as an iceberg transport cost (denoted τ_{iy}^{nlI} hereafter (i for the origin country, y for the estimation year and nlI for the "non-linear iceberg" estimation method) and 2) the two additive and multiplicative transport costs, τ_{iy} and t_{iy} respectively, on some common determinants.

Precisely, what are these dependent variables? We use our first-stage regression results, from which we extract the fixed effect of transport costs associated to the country dimension (recall that we assumed that the iceberg trade cost from country i, good k, year y, $\tau_{iky} = \tau_{iy}\tau_{ky}$, and similarly for the additive transport cost $t_{ik} = t_{iy} + t_{ky}$. Both are expressed in percentage (of the fob price).

Careful: This supposes some re-treatment of our first-stage estimation results (recalling our general estimated equation):

$$p_{iky}^{cif} = [\widehat{\tau}_{iy}\widehat{\tau}_{ky}]\, p_{iky}^{fob} + \left[\widehat{t}_{iy}\widehat{t}_{ky}\right]$$

From this, we re-treat the iceberg component (both estimated alone and with the additive part) to consider $\tau_{iy} = 100(\widehat{\tau}_{iy} - 1)$ (leaving aside the "product" dimension as unrelated to the question here), to have the (country-specific) iceberg trade cost expressed in percentage of the fob price.

As for the additive cost, we have obtained the value of (country-specific) trade cost as a fraction of the fob price, that we re-scale to consider it in percentage, ie considering $t_{iy} = 100t_{iy}$, so that we have the additive cost in percentage of the fob price to explain.

We use data from 2004 to 2013, because we have information provided by Doing Business about exports formalities on an annual basis on this period.

Extracted from the Doing Business website:

Doing Business measures the time and cost (excluding tariffs) associated with exporting and importing a standardized cargo of goods by sea transport. The time and cost necessary to complete

4 predefined stages (document preparation; customs clearance and inspections; inland transport and handling; and port and terminal handling) for exporting and importing the goods are recorded; however, the time and cost for sea transport are not included. All documents needed by the trader to export or import the goods across the border are also recorded. The most recent round of data collection for the project was completed in June 2014.

In particular, we are interested in the variable "Cost to export" (US$ per container). From DB: It is the "cost associated with all procedures required to export goods. Includes the costs for documents, administrative fees for customs clearance and technical control, customs broker fees, terminal handling charges and inland transport." **Careful**: inland transport, transport between the warehouse (assumed to be in the main economic city of the country) and the port. So, this part of the export formality cost is not in the cif-fob price, because the fob price already includes the cost incurred to bring the shipment from the initial warehouse to the boat (even includes handling at the port to bring the merchandises on the boat itself). So, there is a measurement error, the export formality cost is higher than it should be; a priori, bias the associated coefficient downwards.

What do we call "fob" price exactly? In the US import data we use, the cif price is based on applying import charges on the custom value. What is defined as import charges? As detailed in Appendix A, "import charges represent the aggregate cost of all freight, insurance, and other charges (excluding U.S. import duties) incurred in bringing the merchandise from alongside the carrier". So in fact what we call "fob price" is "fas price", for "free alongside". This means that the seller has paid to bring the goods to the port, but not to put them on the boat. In this respect, all hauling charges are excluded from what we call the "fob" price. Our "export formality cost" measure is thus adequate as potential explanation of the gap between the cif price and the fob (in fact, fas) price.

One way of explaining our two-stage reasoning is to start recalling the equation at the root of the analysis:

$$p_{iy}^{cif} q_{iy} = p_{iy}^{fob} q_{iy} \tau_{iy} + t_{iy} q_{iy} \tag{1}$$

This equation states that the value of an import by the US, from country i, in year y (we neglect the product dimension here) ($p_{iy}^{cif} q_{iy}$), is equal to the value of the good exported ($p_{iy}^{fob} q_{iy}$), raised by a given multiplicative cost τ_{iy}, plus an additive cost component, that depends on the quantity of the good exported $t_{iy} q_{iy}$. Manipulating the equation, we get:

$$\frac{p_{iy}^{cif}}{p_{iy}^{fob}} = \tau_{iy} + \frac{t_{iy}}{p_{iy}^{fob}} \tag{2}$$

In the first step, we decompose the gap between the cif and the fob price between its two multiplicative and additive components. In the second stage, we raise the question, what is behind the gap between the fob and the cif prices? As clear from Equation (1), the multiplicative cost is tied to the price of the good exported (the fob price), while the additive cost is rather applied on the quantity exported. This drives us to refine the above question: What is behind each component of the cif-fob gap? In particular, do the gravity variables, primarily distance, that are commonly used as a proxy for iceberg trade costs, really matter on iceberg (i.e., multiplicative) transport costs, or rather on the additive part?

To answer the question, it is worth recalling what the "cif" acronym mean, i.e. what is at the root of the gap between the cif-fob prices. In comparison with the fob price, the cif price is "cost, insurance, freight" included. Our second-stage estimation is then guided by two questions. For each dimension of the three dimension, 1) how can we proxy this dimension, and 2) do we expect this dimension to play rather on the additive or the multiplicative dimension?

From OECD stats: The c.i.f. price (i.e. cost, insurance and freight price) is the price of a good delivered at the frontier of the importing country, including any insurance and freight charges incurred to that point, or the price of a service delivered to a resident, before the payment of any import duties or other taxes on imports or trade and transport margins within the country.

- Freight. 1) how to measure it: the distance from the origin country, as well the energy cost implied by this distance. We accordingly use the "distance" variable, and the oil price per kilometer covered. 2) additive or multiplicative? We expect that this cost depends more on the quantity of goods exported, rather than their value. In terms of the estimated equation, we expect β_2 and α to be non significantly different from 0 when Equation (3) is estimated on the iceberg transport cost τ_{ij}, while significantly positive when considering the additive component t_{ij} (in percentage of the fob price).

- Costs (for handling costs). 1) how to measure it: Our first measure is the "export formality cost" provided by the Doing Business database. Our second measure is the oil price (per USD exported) involved in the shipment, that we view as measuring the maintenance cost of the shipment (independent of the distance covered). 2) additive or multiplicative? We expect this dimension to be additive (related to the quantities of goods that are exported, but not to their value (ie, their price)). In terms of the estimated equation, we expect γ and β_1 to be non significantly different from 0 when Equation (3) is estimated on the iceberg transport cost τ_{ij}, while significantly positive when considering the additive component t_{ij} (in percentage of the fob price).

- Insurance: 1) how to measure it: We have no specific variable to measure the insurance cost. Accordingly, it is likely to be captured by the constant term. You may also argue that insurance cost is related to the distance covered. 2) additive or multiplicative? By contrast to the two other dimensions of the cif-fob price, we expect insurance costs to be proportional to the unit price; accordingly, it is likely to imply a coefficient α significantly different from 0 when estimation is on the iceberg part.

Careful: Again, challenging the role of distance (as we ambition) may receive limited support, in that the counter-argument we should expect: "ok it has no role on the iceberg part of transport costs, and all on additive component of transport costs, but you don't show that it does not play on the other dimensions of trade costs you don't cover (on top of the cif-fob difference) multiplicatively. So, you don't demonstrate what you argue, that distance is not a good proxy for trade costs." Even if we carefully explain that indeed, we are aware of that, the (second) counter-argument could be, ok but then what's the point...

Questions raised

- Which variable to consider for export formality costs? In the DB database, we also have information about the *1)* number of documents to fulfill ("The total number of documents required per shipment to export goods. Documents required for clearance by government ministries, customs authorities, port and container terminal authorities, health and technical control agencies and banks are taken into account), *2)* the time to export (in days) ("The time necessary to comply with all procedures required to export goods. If a procedure can be accelerated for an additional cost, the fastest legal procedure is chosen.") and *3)* the cost to export (deflated US$). Ie, "the Cost to export for each year is divided by GDP deflator, to take the general price level into account when benchmarking this absolute-cost indicator across economies with different inflation trends. In Doing Business 2015 the deflated costs are identical to the non-deflated (base year for the deflator)."

⇒ For now we have chosen the "cost of export" . Deflated or not? The answer is no, since we have on the left, the transport costs in percentage of the fob price, not deflated. Consider also the "administrative costs" (ie nb of documents and time delay)?, but pb of compatibility of measures? Answer: No. The reason is that the custom delay and the workload required to fill the documents are not included in the declaration fob-caf, which is what we use as measure of transport costs. Even if such time delay or administrative burden are likely to affect the magnitude of trade flows.

- Which other variables to consider? The literature commonly uses gravity variables (such as distance, common language, colonial linkages, common border, etc.) as a proxy for trade costs. In the second-step estimation, we only consider distance. This choice relies on the fact that we explain the transport gap measured by the gap between the fob price and the caf price. Our measure of transport cost is only part of the larger size of trade costs, andthere is no a priori reason why the other gravity variables except distance might intervene in the diffence between the price at the export departure point and the import arrival point.

- Question of dimension comparability. On the leftwards of the equation, we have a percentage (iceberg or additive transport costs, in % of the for price), in the country-specific dimension. As for the export formalities costs, in the raw DB database, expressed in US$ per container transported. We want to have it expressed in terms of percentage of the fob price, to have the same dimension on both sides of the equation. We thus calculate a (weighted) mean fob price per country/year, and we divide the formality cost by the mean fob price, so that this variable represents the formality cost per US$ exported.

- Oil price and distance Raw data for oil prices are from BP website (crude oil prices, US$ per barrel).[1] The oil price is expressed in US$ per barrel. To ensure comparability, we transform this variable to express it as percentage of the mean fob price (per country/year). We consider this variable alone (the cost of oil per USD exported) and in interaction with distance (the oil cost burden per kilometer made to export). We come back later on the interpretation of the associated estimated coefficients below.

[1] See more details in Appendix A.

- Time dimension. How to treat it? Two options, run the regression year by year, but not many point in the database (160 to 210 countries depending on the year considered). So, in panel, way to increase the number of observations. Besides, also exploit the idea that technological improvements in aircraft and vessel might have changed the relation between oil prices and trade cost. To capture this dimension, we interact a year fixed effect with the "oil cost per kilometer exported" variable.

- Make a difference between the estimated additive trade cost and the multiplicative trade costs. Ie, should we suppose that the explicative variables enter additively when considering the additive trade costs, and multiplicatively regarding the iceberg component? No a priori reason. In both cases, the dependent variable is expressed as a percentage (of the fob price). We decompose this percentage in various dimension (the cost of oil, the cost of distance, etc.) , which all added, compose the overall trade cost component. This justifies an additive specification, for both the additive and the iceberg trade costs components.

- The specification of error. If we estimate the equation by setting the error term in addition to the other variables, it means that the error term is expressed as a percentage of the fob price (ie, it is independent of the other dimensions, the residual term can be interpreted in percentage points). If we specify the error term in a multiplicative manner, it means that we estimate an error term in percentage of each explicative variable. See below with the estimated equation.

- The transport mode. We have run estimation separating air from vessel. Question, how to treat the "mode" dimension? If we consider them jointly (on a yearly basis), and distinguish between air and vessel by including a dummy for, say, air, it amounts assuming that the only difference between trade costs associated to air and to vessel is in the constant term. This seems to us a quite restrictive assumption. The transport mode is rather to interact with the distance dimension, the oil price dimension, etc. with a specific manner. So, if all explicative variables are interacted with the mode dummy, there is no gain in having a single equation. This drives us to estimate the equation by transport mode. Then, we will compare the results (of distance, of oil price) between both modes, to explore whether there is (or not), a significant difference (regarding the roles of the various explicative variables, as well as in the constant terms) between both modes.

If we adopt the specification that the error term enters additively, on top of the other explicative variables, the estimated equation is the following (conditional on the transport mode, vessel or air, and the type of transport costs, iceberg or additive):

$$\begin{aligned} TC_{iy} \;=\;& a + \alpha dist_i + \beta_1 \frac{oil_y}{p\bar{f}ob_{iy}} + \beta_2 \frac{oil_y}{p\bar{f}ob_{iy}} \times dist_i + \beta_3 \frac{oil_y}{p\bar{f}ob_{iy}} \times dist_i \times FE_y \\ &+ FE_y + \gamma \frac{formalities_{iy}}{p\bar{f}ob_{iy}} + \varepsilon_{iy} \end{aligned} \tag{3}$$

with the subscript i referring to the origin country, y the year of estimation. Under this specification,

- TC_{iy} is the transport cost, expressed as a fraction of the fob price, that varies over time and with the origin country. That is, for 1 US\$ exported by country i to the US in year y, it costs TC_{iy} in trade costs (specific to the transport mode x (x = air, vessel) and the trade cost type z (iceberg, additive or total[2], denoted τ_{iy}^{nlI})).

- $dist_i$ is the distance (in km) from the US,
- oil_y is the oil price in US\$, per year
- $formalities_{iy}$ is the export formality costs in US\$, for country i in year y (from Doing Business database).

As explained above, we transform the last two variables to have them expressed in percentage of the mean fob price observed for the pair (origin country, year). Reasoning: oil_y is the price of the barrel (in US\$, per year). Then $\frac{oil_y}{p\bar{f}ob_{iy}}$ is the price of a barrel per US\$ exported. The variable $\frac{formalities_{iy}}{p\bar{f}ob_{iy}}$ represents the export formalities cost (in US\$) per USD\$ exported (by year-country). FE_y are year fixed effects (the reference year being 2004), ε_{iy} is the residual, and a is a constant term.

Under this specification, the explicative variables and the associated coefficients can be interpreted as follows.

- α measures the cost for each kilometer made, in percentage of the fob price, i.e. for each US\$ exported, that is involved in the transport cost of mode x, type z. Such that $\alpha dist$ represents the share of transport cost (x, z) (in percentage of the fob price) that is attributable to distance (alone). It is expressed in percentage points.
- β_1 represents the number of barrels involved per US\$ exported (by country-year, conditional on the transport mode and the transport cost type), such that $\beta_1 \frac{oil_y}{p\bar{f}ob_{iy}}$ measures the cost in oil per US\$ exported (the number of oil barrels per \$ exported (β_1), times the oil price of one barrel per \$ exported ($\frac{oil_y}{p\bar{f}ob_{iy}}$)). In a sense, this measures the energy cost of freight involved in the export procedure.
- The variable $\frac{oil_y}{p\bar{f}ob_{iy}} \times dist_i$ represents the oil price cost per barrel, of a kilometer covered by the shipment. β_2 can thus be interpreted as the number of barrels associated with a kilometer "exported", such that $\beta_2 \frac{oil_y}{p\bar{f}ob_{iy}} \times dist_i$ measures the cost in oil involved in the shipment (the number of barrels, times the cost of an oil barrel par USD exported, times the number of kilometers crossed.)
- $\beta_3 \frac{oil_y}{p\bar{f}ob_{iy}} \times dist_i \times FE_y$. Accordingly, β_3 in a vector of dimension (1,11) (2005-2013). The idea here is to capture the possibility that the oil cost of international trade might have changed over the years, because aircrafts and vessels have gained in energy needs, so that everything else equal the cost in oil per km covered for exports is lower in 2013 than in 2004 because the machines are less demanding. If this intuition is correct, we expect (the vector of) β_3 to be negative (and all the more negative as we are on recebt years).
- Given that $formalities_{iy}$ represents the cost of export formality per container (in USD), and $\frac{formalities_{iy}}{p\bar{f}ob_{iy}}$ the export formalities of a container per USD exported, γ can be interpreted as the

[2]Meaning, retaining an iceberg formulation without modeling an additive component.

number of containers per USD exported, such that $\gamma\frac{formalities_{iy}}{p\tilde{fob}_{iy}}$ scales the overall cost of export formality, per USD exported.

- ε_{iy} is the trade cost per USD exported that is left unexplained (for instance, if on the left-hand side we have a value for a given pair (country-year) of 0.008 (for this counry-year pair, trade costs is 0.8% of the fob price, or 0.8 USD for 1 USD exported), and that the sum of all explicative terms (from a to $\gamma\frac{formalities_{iy}}{p\tilde{fob}_{iy}}$) amounts to 0.007, it means that the error term is 0.001, equal to 0.1 percentage point of the trade cost (in % of the fob price).

rediger totpo sur forme de l'erreur It is not impossible that the variance of the error term depends on the size of the observed costs. E.g., because costs and their measures are bounded by zero, the error has to be small when costs are small and it might be higher when costs are high. To correct for the potential heteroskedasticity, we report Huber?White standard errors.

A Data Appendix

A.1 Fob-cif prices

From the US Census website:

The Customs value is the value of imports as appraised by the U.S. Customs and Border Protection in accordance with the legal requirements of the Tariff Act of 1930, as amended. This value is generally defined as the price actually paid or payable for merchandise when sold for exportation to the United States, excluding U.S. import duties, freight, insurance, and other charges incurred in bringing the merchandise to the United States. The term "price actually paid or payable" means the total payment (whether direct or indirect, and exclusive of any costs, charges, or expenses incurred for transportation, insurance, and related services incident to the international shipment of the merchandise from the country of exportation to the place of importation in the United States) made, or to be made, for imported merchandise by the buyer to, or for the benefit, of the seller. In the case of transactions between related parties, the relationship between buyer and seller should not influence the Customs value.

In those instances where assistance was furnished to a foreign manufacturer for use in producing an article which is imported into the United States, the value of the assistance is required to be included in the value reported for the merchandise. Such "assists" include both tangible and intangible assistance, such as machinery, tools, dies and molds, blue prints, copyrights, research and development, and engineering and consulting services. If the value of these "assists" is identified and separately reported, it is subtracted from the value during statistical processing. However, where it is not possible to isolate the value of "assists", they are included. In these cases the unit values may be increased due to the inclusion of such "assists". Import Charges

The import charges represent the aggregate cost of all freight, insurance, and other charges (excluding U.S. import duties) incurred in bringing the merchandise from alongside the carrier at the port of exportation in the country of exportation and placing it alongside the carrier at the first port of entry in the United States. In the case of overland shipments originating in Canada or Mexico, such costs include freight, insurance, and all other charges, costs and expenses incurred in bringing the merchandise from the point of origin (where the merchandise begins its journey to the United States) in Canada or Mexico to the first port of entry. C.I.F. Import Value

The C.I.F. (cost, insurance, and freight) value represents the landed value of the merchandise at the first port of arrival in the United States. It is computed by adding “Import Charges” to the “Customs Value” (see definitions above) and therefore excludes U.S. import duties.

A.2 The second-stage explicative variables

Oil prices It may come from two sources, and for two measures.

- From the BP Statistical Review of World Energy, we get an Oil Price series from 1974 (annual data, in dollars per barrel).

- From the Federal reserve of Saint Louis (`https://www.stlouisfed.org/`), we get the “Crude Oil Prices: West Texas Intermediate (WTI) - Cushing, Oklahoma” series, expressed in dollars per Barrel (Annual, Not Seasonally Adjusted data). It is available from 1986 to 2013. We could also use the “Crude Oil Prices: Brent - Europe” series (still in dollars per Barrel, from 1987 to 2013).

We check that these 3 series are in fact very close to one another over 1987-2013. Because we don’t want to limit ourselves to data post 2004 (the availability of the “export formality cost” from Doing Business, in order to exploit all information over the entire period 1974-2013, we retain the BP series as benchmark (very frequently retained in the literature), as it is the sole which is available over the whole period.