

OPENING THE TRANSPORT COSTS BLACK BOX

NOTE: SECOND-STAGE REGRESSION

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

$$p_{ikt}^{cif} q_{ikt} = p_{ikt}^{fob} q_{ikt} \tau_{ikt} + t_{ikt} q_{ikt} \quad (1)$$

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

$$\frac{p_{ikt}^{cif}}{p_{ikt}^{fob}} = \tau_{ikt} + \frac{t_{ikt}}{p_{ikt}^{fob}} \quad (2)$$

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?

In this second stage of the paper, we thus keep on questioning the way the literature models the trade costs determinants. Do the proxies frequently used as measures for trade costs, typically distance, correspond more to additive costs or to multiplicative trade costs? Or total? Usually, people use a measure of distance as a way to capture iceberg trade costs. 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? Using our decomposition of transport costs obtained in the first stage, we can assess the contributions of typical gravity variables to each component of transport costs (additive / iceberg). To put this simply, consider first our “raw” estimate of trade costs (meaning, only modeling a multiplicative form); we estimate the role of distance (as quite frequent) and find that this fits 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 a high correlation with the additive cost, and 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.

Data treatment Before presenting the estimated equation, let us mention how we deal with the data. We regress 1) the “overall trade costs” (i.e.), modeled as an iceberg transport cost (denoted τ_{ik}^{nli} hereafter (i for the origin country, k for the product and nli for the “non-linear iceberg” estimation method) and 2) the two additive and multiplicative transport costs, τ_{ik} and t_{ik} respectively, on some

common determinants. Precisely, we re-treat the iceberg component (both τ_{ik}^{nII} and τ_{ik}) to have it expressed in percentage of the fob price, ie we build the variable $\tau_{ik} = 100(\hat{\tau}_{ik} - 1)$ (with $\hat{\tau}_{ik}$ the estimated variable obtained at stage 1). As for the additive cost, from stage 1 of the paper, we have obtained the value of (country-specific) trade cost as a fraction of the fob price (denoted \hat{t}_{ik}); we multiply it to have it expressed as a percentage of the fob price (considering $t_{ik} = 100\hat{t}_{ik}$). Further, and in line with our first-step estimation method, we estimate the determinants of trade costs on a yearly basis, and conditional to the transport mode (air or vessel).

In the three cases, we want to explain trade costs expressed as a percentage of the fob price. All our three measures of trade costs are bounded by 0 as minimal value, and we drop outliers by only keeping the 95 first percentile of each distribution.

Functional form To derive the estimated equation, we start recalling what is behind the gap between cif-fob prices. 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. In comparison with the fas price, the cif price is “cost, insurance, freight” included.¹

Accordingly, if we denote TC_{ik} the total transport cost paid to import the product k from country i (on a given year, conditional to a certain transport mode), expressed in USD, we can decompose it in three elements:

$$TC_{ik} = \text{Freight} + \text{Insurance} + \text{Handling costs}$$

Our second-stage estimation is then guided by two questions. For each of these three dimensions (cost, insurance, freight) at the root of the cif-fas prices gap, 1) what are its determinants, and 2) do we expect this dimension to play rather on the additive or the multiplicative dimension?

- Insurance: Based on information collected about insurance grids (**source???**)², we infer that insurance costs depend on the value of the shipment, conditional on the type of product transported (Insuring one ton of diamonds does not bear the same cost as a ton of wood). The database provides us with the quantity (in terms of kilograms) of the shipment (q_{ik}), as well as the price, so that we have the value of the shipment. However, we have no dedicated variable to measure the insurance cost in the database. Accordingly, we capture this dimension by including product fixed effects. Insurance grids retain a degree of product categorization much broader than the 3-digit level we retain as benchmark sectorial degree (our k index for the product dimension). In most pieces of information we have collected, insurance grids adopt a ranking in **20** (??) types of products. Accordingly, we include product fixed effects of dimension **20**, that we

¹From 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.

²See Appendix ??? for details.

refer to with the subscript k' , i.e. $FE_{k'}$. If we denote p_{ik}^{fob} and q_{ik} the (fas) price and the quantity of the shipment, total insurance costs (in USD) are accordingly measured by $FE_{k'} \times p_{ik}^{fob} \times q_{ik}$. As insurance cost depend on the value of the shipment, we expect it to be proportional to the price, ie rather play on the iceberg part of the transport costs.

- Costs (for handling costs). The amount paid to load the bulk on transport mode is likely to depend on the quantity of product charged, as well as the bulk implied in the shipping (A ton of cars does not represent the same volume on board as a ton of apparel); further, handling costs also depend on the country of origin (labor and various other costs). If the database provides us with the quantity that is imported (of product k from country i), we do not have information about the volume it takes, which affects the handling costs. We capture this dimension by including fixed effects by product category. In doing so, we adopt the same product categorization as for insurance costs, for similar reasons (it is very unlikely that the handling costs varies within the textile category). Last, handling costs are likely to be country-dependent, which we capture by including country fixed effects (FE_i). Denoting the fixed effects associated to handling costs by $\widehat{FE}_{k'}$, we accordingly measure total handling costs (in USD) by $FE_i \times \widehat{FE}_{k'} \times q_{ik}$. As such costs depend on the quantity that is shipped, but not on the price, we expect this dimension to play on the additive part of transport costs.
- Freight. Quite intuitively, freight costs should depend on the distance between the origin country and the U.S. (denoted $dist_i$). Further, they likely depend on the quantity (q_{ik}) and the bulk of the shipment. Again, we capture this dimension through product fixed effects, based on the same product categorization (based on k') as for handling and insurance costs. Accordingly, we capture freight costs through $\widetilde{FE}_{k'} \times dist_i \times q_{ik}$. Because this cost is most likely dependent on the quantity of goods exported rather than their value, we expect this term to have more impact on the additive part of total transport costs.

As a result, total transport costs by TC_{ik} (expressed in USD) can be decomposed in the three components of insurance, handling costs and freight, according to:

$$TC_{ik} = \underbrace{FE_i \times \widehat{FE}_{k'} \times q_{ik}}_{\text{Costs}} + \underbrace{FE_{k'} \times p_{ik}^{fob} \times q_{ik}}_{\text{Insurance}} + \underbrace{\widetilde{FE}_{k'} \times dist_i \times q_{ik}}_{\text{Freight}}$$

By definition, such transport costs can be written as $TC_{ik} = (p_{ik}^{cif} - p_{ik}^{fas})q_{ik}$. Accordingly, we can write:

$$\begin{aligned} (p_{ik}^{cif} - p_{ik}^{fas})q_{ik} &= FE_i \times \widehat{FE}_{k'} \times q_{ik} + FE_{k'} \times p_{ik}^{fob} \times q_{ik} + \widetilde{FE}_{k'} \times dist_i \times q_{ik} \\ \Leftrightarrow \left(\frac{p_{ik}^{cif}}{p_{ik}^{fas}} - 1 \right) p_{ik}^{fas} q_{ik} &= FE_i \times \widehat{FE}_{k'} \times q_{ik} + FE_{k'} \times p_{ik}^{fob} \times q_{ik} + \widetilde{FE}_{k'} \times dist_i \times q_{ik} \end{aligned}$$

From the above equation, it comes that $\frac{TC_{ik}}{p_{ik}^{fas} q_{ik}} = \left(\frac{p_{ik}^{cif}}{p_{ik}^{fas}} - 1 \right)$, which is exactly the measure of

trade costs obtained in the first stage.³

Accordingly, we can decompose each of our three measures of trade costs (τ_{ik}^{nII} , τ_{ik} and $\frac{t_{ik}}{p_{ik}^{fas}}$, expressed in the three cases as a percentage of the fob price, and denoted tc_{ik} to gain in generality, according to:

$$tc_{ik} = \underbrace{\frac{FE_i \times \widehat{FE}_{k'}}{p_{ik}^{fas}}}_C + \underbrace{FE_{k'}}_I + \underbrace{\frac{\widetilde{FE}_{k'} \times dist_i}{p_{ik}^{fas}}}_F \quad (3)$$

Based on Equation (3), we specify the estimated equation as follows:

$$tc_{ik} = \left[\frac{FE_i \times \widehat{FE}_{k'}}{p_{ik}^{fas}} + FE_{k'} + \frac{\widetilde{FE}_{k'} \times dist_i}{p_{ik}^{fas}} + FE_i \right] e^{\varepsilon_{ik}} \quad (4)$$

with ε_{ik} the error term, centered on 0. This calls for two comments. First, we add the country fixed effects separately in the regression, in order to correctly identify the associated interaction term. Second, we specify the error term multiplicatively. 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). In this case, it is very likely 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. Adopting this functional form would then be subject to a potential bias in point estimates due to the heteroscedasticity of error terms. This drives us to specify the error term in a multiplicative manner (such that the error term is in percentage of each explicative variable and centered on 0), and to take the log of the equation.

Accordingly, the functional form that we estimate, for each of the three transport costs obtained at the first stage ($tc_{ik} = \tau_{ik}^{nII}$, τ_{ik} and $\frac{t_{ik}}{p_{ik}^{fas}}$, conditional on the year and the transport mode), is expressed as:

$$\log tc_{ik} = \log \left[\frac{FE_i \times \widehat{FE}_{k'}}{p_{ik}^{fas}} + FE_{k'} + \frac{\widetilde{FE}_{k'} \times dist_i}{p_{ik}^{fas}} + FE_i \right] + \varepsilon_{ik} \quad (5)$$

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.” ...

Questions raised

- 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

³Recall that our first stage starts from the equation: $\left(\frac{p_{ik}^{cif}}{p_{ik}^{fas}} - 1 \right) = \tau_{ik}^{nII}$ in the “raw” equation and $\left(\frac{p_{ik}^{cif}}{p_{ik}^{fas}} - 1 \right) = \tau_{ik} + \frac{t_{ik}}{p_{ik}^{fas}}$ when decomposing between iceberg and additive transport costs.

by the gap between the fas price and the caf price. Our measure of transport cost is only part of the larger size of trade costs, and there is no a priori reason why the other gravity variables except distance might intervene in the difference between the price at the export departure point and the import arrival point.

- 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.

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 Product classification for C,I, F fixed effects

to be done