FONDATION POUR LES ÉTUDES ET RECHERCHES SUR LE DÉVELOPPEMENT INTERNATIONAL

Working Paper 123 March 2015 Development Policies

Can mirror data help to capture informal international trade?*

Céline Carrère
Christopher Grigoriou

Céline Carrère is Professor at the European Institute, University of Geneva. Her research interests include international trade, regional integration and Preferential Market Access. She is Senior Fellow at Ferdi.

Christopher Grigoriou (Corresponding author) is external Expert for the Fiscal Affair Department at the International Monetary Fund, Associate Professor of Economics at CERDI-University of Auvergne, and invited Professor of Econometrics at the University of Lausanne. He si also Senior Economist- Risk Management Expert at the SICPA and Fellow at the Ferdi. **E-mail : chgrigoriou@gmail.com**

Abstract

Empirical studies on international trade extensively rely on the use of mirror trade statistics, i.e data reported by trading partners. However, while extensive reviews have been done on how to use mirror data to compensate poor quality data or to proxy transportation costs, very few has been done to see if and how the gap between the declared and mirrored disaggregated bilateral data could be used to capture informal cross border trade. Indeed, beyond the valid logistic reasons to explain why reported bilateral export flows from one country do not match the respective reported imports of its partner country, deliberate misreporting could significantly contribute to explain those discrepancies, either through misevaluation or misclassification of the imported goods, notably to evade tariffs and taxes. This paper proposes a review of the reasons for the gap between matched partner data, before investigating stylized facts from UN-COMTRADE data. .../...

Keywords: Keywords: informal trade, mirror trade statistics, panel data, gravity equation.
JEL Classification: F14, C33

Acknowledgements
This paper has benefited from valuable comments from the participants of the Research Conference on Informality, International Trade and Customs, held at Brussels the 3rd of June 2013 (WCO-WB). Special thanks are due to Marco Fugazza (UNCTAD) and Olivier Cadot (HEC Lausanne) for fruitful discussions and comments.
Any mistakes or errors remain the authors' own.

…/ … Empirical analysis relying on econometrical panel data over a worldwide set of data at the 6 digits level evidences that discrepancies from the mirror data are not erratically driven. A statistically significant relationship between the gap and macroeconomic variables such bilateral distance, gdp per capita, average tariffs, foreign direct investments (FDI), implementation of regional trade agreements (RTA) have been evidenced. Based on these preliminary correlations, a probit has been run on orphan imports (imports reported by importing country without equivalent by exporting country) and predicts accurately up to 68% of these misclassification cases. Thus, part of the gap can be predicted by macroeconomic variables, some of them suggesting a relationship between cross-border trade flows misreporting and fraud opportunities to evade tariffs and taxes.

Contents

1 Introduction 6

2 Ratio CIF/FOB: what do we know? 7

2.1 Theoretical definition of the CIF/FOB ratio 7

2.2 Potential sources of discrepancies in CIF/FOB ratios 8

3 Ratio CIF/FOB: some stylized facts 12

3.1 Trade data 12

3.2 Fact #1: Product misclassification is substantial 13

3.3 Fact #2: Country pair characteristics as drivers of the gap 15

3.4 Fact #3: Importer country characteristics as driver of the gap 21

3.5 Fact #4: Product characteristics as driver of the gap 29

4 Empirical strategy: one step further 32

5 Concluding comments 36

References 38

List of figures

Figure 1. Mirror data at the country pair level, different product aggregation levels, 2008…… 14

Figure 2. Average percentage of lines with orphan imports at the county-pair level related to the contiguity and RTA dummy variables, 2008…….. 17

Figure 3. Median of the cif/fob at the country-pair level related to the contiguity and RTA dummy variables, 2008…….. 17

Figure 4. Correlation between the percentage of lines with orphan imports at the country-pair level and the bilateral distance (in log), 2008…… 19

Figure 5. Correlation between the cif/fob ratio median (in log) at the country-pair level and the bilateral distance (in log), 2008……... 19

Figure 6. Correlation between the percentage of lines with orphan imports and the percentage of reexports at the country-pair level (in log), 2008……. 20

Figure 7. Correlation between the cif/fob ratio median (in log) and the percentage of reexports (in log) at the country-pair level, 2008……. 20

Figure 8. Average percentage of lines with orphan imports at the country-pair level related to landlockedness of the importer country, 2008……... 22

Figure 9. Median of the cif/fob ratio at the country-pair level related to landlockedness of the importer country, 2008……... 23

Figure 10. Correlation between the percentage of lines with orphan imports at the country-pair level and the GDP per capita (in log) of the importer country, 2008…… 24

Figure 11. Correlation between the cif/fob ratio median (in log) at the country-pair level and the GDP per capita (in log) of the importer country, 2008…… 25

Figure 12. Correlation between the percentage of lines with orphan imports at the country-pair level and the unweighted average tariff of the importer country, 2008…… 25

Figure 13. Correlation between the cif/fob ratio median (in log) at the country-pair level and the unweighted average tariff of the importer country, 2008…… 26

Figure 14. Correlation between the percentage of lines with orphan imports at the country-pair level and the collected import duties in percentage of total imports of the importer country, 2008……. 26

Figure 15. Correlation between the cif/fob ratio median (in log) at the country-pair level and the collected import duties in percentage of total imports of the importer country, 2008…….. 27

Figure 16. Correlation between the percentage of lines with orphan imports at the country-pair level and the stock of inflows FDI (in log) in the importer country, 2008……. 27

Figure 17. Correlation between the cif/fob ratio median (in log) at the country-pair level and the stock of inflows FDI (in log) in the importer country, 2008…… 28

Figure 18. Correlation between the percentage of lines with orphan imports at the country-pair level and CPIA corruption index of the importer country, 2008… 28

Figure 19. Correlation between the cif/fob ratio median (in log) at the country-pair level and the CPIA corruption index of the importer country, 2008…… 29

Figure 20. Distribution of CIF/FOB ratio at the country level, 2008…… 42

Figure 21. Distribution of the share of orphan imports at the country level, 2008…… 42

Figure 22. Distribution of the share of lost exports at the country level, 2008…… 43

Figure 23. Distribution of the CIF/FOB ratio at the product level, 2008……... 43

Figure 24. Distribution of the share of the orphan imports at the product level, 2008…... 44

Figure 25. Distribution of the share of the lost exports at the product level, 2008…… 44

List of tables

Table 1.	Sum of the bilateral trade, imports and exports, 2008…….……	13
Table 2.	Distribution of the cif/fob ratio at the county-pair level, different product aggregation levels, 2008……..……..……..……	15
Table 3.	Some descriptive statistics of our two measures of the gap in mirror data at the country-pair level, 2008……..……..……..…	16
Table 4.	Countries in the top 10, 2008……..……..…….	21
Table 5.	10 products with the highest share of orphan imports corresponding to the most exported products over the world, 2008……..……..……..	30
Table 6.	10 products with the highest share of lost exports corresponding to the most imported products over the world, 2008……..……..……...	30
Table 7.	10 products with the highest CIF/FOB ratios, 2008……..……..…....	31
Table 8.	10 products with the lowest CIF/FOB ratios, 2008……..……..……	31
Table 9.	Probit estimates of orphan imports, 2008……..……..……	34
Table 10.	Regression results, within estimator, 2008……..……..……	35
Annex A1.	List of countries including in the sample……..……..……..	40
Annex A2.	Some descriptive statistics at the country level……..……..……	42
Annex A3	Some descriptive statistics at the product (HS6) level……..……...	43

1. Introduction

Empirical studies on international trade extensively rely on the use of mirror trade statistics, i.e data reported by trading partners. Most of the time, this is done to compensate missing values or poor quality data issues: instead of using export flows as declared by a country, the mirror imports flows - as reported by its partners - are preferred, imports being assumed to be better reported due to taxation issues (see the large literature on gravity model estimates including Anderson and Van Wincoop 2003, Carrere 2006, Anson, Cadot and Olarreaga 2006 among others). Some studies also use the ratio between a trade flow and its mirror to proxy transportation costs (e.g. Yeats 1978, Rose 1991, Baier and Bergstrand 2001, Hummels and Lugovskyy 2006).

On one hand, there are valid structural or logistic reasons other than transportation costs to explain why reported bilateral export flows from one country do not match the respective reported imports of its partner country. This includes notably passing through third countries or having countries with different trading or clearing systems; see among others Hummels and Lugovskyy (2006) or Gaulier and Zignago (2010) which have respectively reviewed and discussed the issues associated to such a use of mirror data and how reported and mirror data could be reconciliated.

But on the other hand, very few has been done to see if and how the gap between the declared and mirrored disaggregated bilateral data could be used to proxy informal cross border trade. Indeed, deliberate misreporting could significantly contribute to explain a part of the discrepancies, either through misevaluation or misclassification of the imported goods, notably to evade tariffs and taxes.

This paper proposes such a review of the reasons for the gap between matched partner data, before investigating stylized facts from UN-COMTRADE data. Empirical analysis relying on panel data over a worldwide set of data at the 6 digits level evidences that discrepancies from the mirror data are not erratically driven. A statistically significant relationship between the gap and macroeconomic variables such bilateral distance, gdp per capita, average tariffs, foreign direct investments (FDI), implementation of regional trade agreements (RTA) have been evidenced. Based on these preliminary correlations, a probit has been run on orphan imports (imports reported by importing country without equivalent by exporting country) and predicts accurately up to 68% of these misclassification cases. In the same way these variables turn to have a significant impact on the cif/fob ratio in a within panel data at the 6 digits level. Thus, part of the gap can be predicted by macroeconomic variables, some of them suggesting a relationship between cross-border trade flows misreporting and fraud opportunities to evade tariffs and taxes. The paper is organized as follows. Section 2 defines the cif/fob ratios, and reviews the literature on the reasons for discrepancies. Section 3 presents the data from COMTRADE and proposes some stylized facts and correlation analysis between two measures of

potential misreporting and country-pair or country specific variables. Section 4 tests econometrically the previously detailed stylized facts from panel data. Section 5 concludes.

2. Ratio CIF/FOB : what do we know?

The "matched partner" CIF/FOB ratio technique consists in comparing the valuation of the same flow reported by both the importer and the exporter. This technique has notably been used in the literature to overcome the lack of data on the transportation cost. As the imports are reported including the cost insurance and freight (CIF) while the exports are net of these charges, the difference between the two trade flows should yield a difference proxying the transport costs. This section details this definition and reviews the literature on the reasons which could explain the gap, beyond the only transportation costs.

2.1. Theoretical definition of the CIF/FOB ratio

The so-called cif/fob ratio of a trade flow for a product k imported by a country i from a country j can be defined as following:

$$R_{ijkt}^{cif/fob} = \frac{p_{ijk}^{M} Q_{ijk}^{M}}{p_{ijk}^{X} Q_{ijk}^{X}} \qquad (1)$$

With:

p and *Q* being respectively the price and quantity of the trade flows,

$p_{ijk}^{M} Q_{ijk}^{M}$ being the value (in current US dollars) of the import flows of product k from j to i *as reported by the importing country i*,

$p_{ijk}^{X} Q_{ijk}^{X}$ being the value (in current US dollars) of the import flows of product k from j to i *as reported by the exporting country j*.

p_{ijk}^{M} being the price of the imported unit value of product k as declared by the importer country i which includes the freight charges as well as the insurances on these freight charges compared to the price declared by the exporter country j. We assume that such freight charges are ad valorem and (denoted τ_{ijk}), i.e.:

$$p_{ijk}^{M} = p_{ijk}^{X} \left(1 + \tau_{ijk}\right) \qquad (2)$$

Hence in a "perfect World" with no discrepancy in reported volumes by trade partners, we should have $Q_{ijk}^{M} = Q_{ijk}^{X}$ (we will come back on this point below). Then, given equations (1) and (2) we can write:

$$R_{ijkt}^{cif/fob} = \left(1+\tau_{ijk}\right) \qquad (3)$$

So in such a "perfect World" using mirror trade date and computing the corresponding cif/fob ratio would provide good estimates of transport costs and associated insurance. This ratio is used as proxy for transport costs in numerous studies (see for instance Harrigan 2013, Baier and Berstrand 2001).

However, as mentioned by Hummels and Lugovskyy (2006) or Gaulier and Zignago (2010) and as we will see in the next section, this ratio is far from only capturing transport costs. For instance, Hummels and Lugovskyy (2006) report from a sample of 17'790 country pairs in 1997 that hardly 50% of the total bilateral pairings are in range of data variation considered as "reasonable"[1] and a significant part of the cif/fob ratios computed from their sample are lower than 1 implying negative transport costs.

What can explain such discrepancies?

2.2. Potential sources of discrepancies in cif/fob ratios

Our purpose is here to review the potential origins of the discrepancies in the matched partner trade flows to isolate the deliberate errors from the structural or "logistic errors". While the latter are more or less unavoidable or due to structural considerations[2], the former are motivated by illicit operations, from tariff evasion to capital flight, resulting in deliberate mis-valuation (either over or under) or misclassification (either on the origin/destination or on the product).

Measurement errors of trade value: $Q_{ijk}^{M} \neq Q_{ijk}^{X}$ **,** $p_{ijk}^{M} \neq p_{ijk}^{X}\left(1+\tau_{ijk}\right)$

Definition of the recorded flows and timing issues

Customs administrations traditionally record more carefully the imported goods than the exports as tariffs and non-tariff barriers for sensitive goods are mostly related to the imports (see for instance Yeats (1995) or Gaullier and Zignago 2010)[3]. Moreover, the definition of the minimum thresholds for trade flows to be officially reported may differ from a country to another. This could be all the more important as both of the considered countries are part of a

[1] Hummels and Lugovskyy (2006) defines reasonable range of values as ratios in the interval (1, 2), implying an *ad valorem* transportation cost between 0% and 100%: a ratio below 1 would reflect negative transport costs, above 2 would imply transportation costs exceeding the values of the shipped goods. The authors find that 46.55% of the total bilateral pairs of countries are in the so-colled "reasonable interval", 63% if weighted by the volumes of trade.

[2] Federico and Tena (1991) classify the cases as (1) unavoidable factors such as transportation costs and other service charges, (2) stuructural differences in compilation and statistical criteria, (3) unintentiontal errors and deliberate misreporting.

[3] Note that Comtrade does not report flows before 1'000 dollars.

same regional trade agreement with a lack of internal customs controls and thus no official documentation to rely on.

Transit time $Q_{ijk}^{M} \neq Q_{ijk}^{X}$

Yeats (1995) lists the transit period as a potential cause for misreporting. Exports and matching imports could indeed be reported in two different periods of time depending on the transit time. Hamanaka (2011) however considers this problem as marginal when dealing with yearly data.

Exchange rate conversions issues $p_{ijk}^{M} \neq p_{ijk}^{X}\left(1+\tau_{ijk}\right)$

Exchange rate conversion issues could also contribute to the gap, in a volatile environment, depending on the timing of the transaction and possible changes in the exchange rate between the beginning and end the of the transaction.

Measurement errors of trade volumes: the case of reexports $Q_{ijk}^{M} \neq Q_{ijk}^{X}$, $p_{ijk}^{M} \neq p_{ijk}^{X}\left(1+\tau_{ijk}\right)$

Reexport and transhipment account for a substantial part of the gap in the CIF/FOB partner matched data. The United Nations (1998) defines reexports as trade flows with goods entering the customs territory of a country and then being shipped to another, without having been transformed. The United Nations recommends reporting the country of origin, and not of provenance, for the imports, but the last know destination for the exports. There will henceforth be a mechanic gap when the exporter does not know the "true" destination (see Yeats 1995 or Guo 2009). Moreover, despite the UN rule, the country of shipment will still be considered as the import partner where the country of origin cannot be identified. Box 1 briefly describes the incidence of reexport flows on the CIF/FOB ratios.

Mellens et al. (2007) estimate that reexports represent 15% of Germany's total exports, over 50% for Singapore and approximately 95% for Hong-Kong. Ferrantino and Wang (2008) also find that more than 90% of Hong Kong's total exports represent re-exports either from China or from a third economy (see also Hamanaka 2011 for Singapore or Hong Kong and China, or Rotterdam for Europe). Easterly and Reshef (2010) explain the case of landlocked countries which do not have their own port and hence do not ship directly their exports by sea. They quote the case of Rwanda's exports (notably coffee) which could be erroneously documented as being shipped to Kenya as running through Mombasa port.

Box 1- Incidence of the reexports flows on the CIF/FOB ratio

Following UN recommendations, A exporting to C through B will in most of the cases report an export to B, the last know destination, while C will report an import from A, the country of origin. This implies distorted CIF/FOB ratios between the three pair of countries. (1) The country pair AC CIF/FOB will have an upward bias as exports (FOB) declared by A do not include goods going through B (the first destination) contrary to the import declared by C (CIF). (2) Country pair BC will have a downward bias in its bilateral CIF/FOB ratio as the ratio will miss the imports C has reported from the country of origin (A) instead of from the country of reexport (B), while the former has counted the transaction in its exports to C.

Trade flow	Reported Exports by the "exporting side"	Reported Imports	Impact on the CIF/FOB ratio
Import from the export flow (A-C)	- "direct" exports from A to C only	- "direct" imports from A - "reexported" by B imports from A	Upward bias: exports through B are "missing"
Import from the reexport flow (B-C)	- "direct" exports from B to C - "reexported" goods from B to C	- "direct" imports from B only (i.e for which B is the country of origin)	Downward bias: imports through B are missing

Note:

(1) The trade discrepancy associated to reexport or transhipment is not only on quantities but also on prices if additional mark-ups is applied to the initial price during the transhipment phase via branding or repackaging.

(2) Reexports could not only result in bias in the cif/fob ratio but also in an increased number of non-matching lines. If A does not report ANY exports to C while C declares the imports from A, the cif/fob ratio will be as if they were no export at all from A to C.

Beyond these "logistic" errors there are errors resulting from deliberate[4] misreporting, mostly for tax evasion purposes, either through misevaluation or through misclassification.

Deliberate misreporting for fraudulent purposes: $Q_{ijk}^{M} \neq Q_{ijk}^{X}$, $p_{ijk}^{M} \neq p_{ijk}^{X}\left(1+\tau_{ijk}\right)$

Tariff evasion (undervaluation p_{ijk}^{M} or Q_{ijk}^{M} *)*

A classic reason for misreporting trade flows is the undervaluation of imports for tariff evasion purposes, notably for products that have relatively high tariffs. Several studies have empirically

[4] Note that misclassification or misvaluation could be unintended and just result from the potential complexity of the valuation and classification of some products. However we consider all systematic and substantial discrepancies in a single category as our purpose is to identify mismatch between the reality and the recorded flows, and they will all result in evaded taxes from the Government point of view.

evidenced a positive correlation between the missing trade and the tariff rate. Fisman and Wei (2004) on trade flows between China and Hong-Kong, Javorcik and Narciso (2008) in ten transition economies, Mishra, Topalova and Subramanian (2008) in India, Rotunno and Vézina (2011) from Chinese imports from multiple exporters. Note that misreporting can also result from misclassification as importer could deliberately declare the imports from a lower tariff line category or from a country with a preferential trade agreement.

Customs officials role (overvaluation of p_{ijk}^{M} or Q_{ijk}^{M})

Javorcik and Narciso (2012) question the role of the customs officials in the CIF/FOB ratio discrepancy: as most of the tariffs are *ad valorem*, customs officials could have incentives to overvalue the imports to increase the perceived tariffs. Moreover, there could also be misclassification to re-classify goods in a higher tariff line category. Assuming the article VII of the GATT on Customs Valuation Agreement limits the discretion of customs officials, they assess the impact of WTO accession on the relationship between the tariff rate and the unit value gap (including country-pair, product and time effects) and they expect the mis-reporting to be decreased by the WTO accession. Their study rely on data at the 6 digit HS level over the exports of US, Germany, Japan and France to 15 importing countries over 1992-2009 and from differentiated products according to Rauch (1999) classification. They conclude that fraud on the valuation is actually decreased after the accession to WTO but is replaced by fraud on the quantity[5].

Transfer pricing, profit shifting and evading capital (overvaluation of p_{ijk}^{M})

Overvaluation of imports might occur when importers attempt to shift profit from countries with a strong capital control or with high corporate taxes (cf Bagwati 1964 or 1967, Sheik 1974 or Yeats 1995). Profit shifting through transfer pricing leads to overvalue the price of the imported goods to evade local corporate taxes and shift profits away from the importing country (Ferrantino and Wang 2008 or Hamanaka 2011). Misclassification could also occur for profit shifting purposes, reporting imports from a lower tariff line.

Reasons to hide the truth on the exports (undervaluation of Q_{ijk}^{X})

Some countries do not classify products properly for political or economic reasons. This is why there is the code labelled "999'999" in the 6 digits HS classification for unspecified items. This could be for confidentiality issues if a small number of enterprises are doing most of the operations, or for products for which exports are submitted to quotas resulting from

[5] Note that Javorcik and Narciso only consider misevaluation related to customs officials in such a way that they would lift-up the social benefit through increased collected taxes. However, in a corrupted environment, the incentives could also be to increase the private benefits, importer and customs officers sharing the amounts due by the importer. In that situation, the "discretion" of the customs officials that will be all the higher than the level of corruption is important could result in underreporting the value of the imported goods.

international commodity agreements (cf. Yeats 1995 on petroleum exports or on coffee or cocoa markets).

Carrousel mechanisms (overvaluation of Q_{ijk}^{X})

Deliberate overvaluation of goods and misclassification of the products and/or of the destination could also occur to take advantage of reductions or duty drawback (carousel schemes for instance) or to receive specific export subsidies. This could in turn explain a "too large" CIF/FOB gap (Yeats 1995, Hamanaka 2012) once controlled for structural elements as transportation costs etc.

3. Ratio CIF/FOB: some stylized facts

3.1. Trade data

The trade database used is the raw data as reported by countries to the UN, i.e. the COMTRADE database. The reported flows are bilateral at the product level according to the Harmonized System revision 1 at the 6-digit levels, HS6 thereafter. Given that around 200 countries and 5106 products are reported in the COMTRADE database the potential number of observations is 203 million per year.

In the COMTRADE database, countries report only strictly positive trade flows. Hence, there is no distinction between zero trade flows and missing values in raw data. We then follow the usual assumption to discriminate between zero and missing values: for a given year, when a country reports at least one strictly positive import flow we assume that no information on other country-pair / product lines means zero for this country and this year. We retreat export flows in the same way. We then exclude from the sample the HS6 lines with a zero trade flow reported by both partners.

Moreover, we exclude the "999'999" classification at the HS6 digit, as this code correspond to "unspecified" goods (cf. section 2).

3.2. Fact #1: product misclassification is substantial

World level

When looking at the most aggregate level, i.e. the world total trade in 2008[6], the cif/fob ratio is around 1.06 (see Table 1).

Table 1 : Sum of the bilateral trade, imports and exports, 2008

At the aggregated level (in $)	
Total imported goods ($)	1.52E+13
Total exported goods ($)	1.43E+13
CIF/FOB ratio	1.06

Source: Author's computation based on UN COMTRADE database

This ratio corresponds to a cost of transport on average equal to 6% of the fob value of the shipped goods, which is consistent with the usual 10% imputational rule assumed by the IMF.

Country pair level, different level of product aggregation

One of the main issues of the cif/fob ratio used in empirical studies is that the computation is often based on the ratio of aggregate trade flows over all products or by main categories (agriculture, manufactures, etc.). However, to understand how mirror data can help to detect some informal trade we have to go beyond traditional analysis and look at the most disaggregated data available on a world basis (i.e. harmonised over all countries in the World). Indeed, using the importing country as a benchmark would imply a double aggregation: on product and on trade partners which would confuse the interpretation. To disentangle the effect of product aggregation on mirror matching estimates, we use as benchmark the country-pair relationship.

Hence, for each country-pair we look at different levels of aggregation: total trade (one flow per country pair), the HS-2 digits level (99 chapters), the HS-4 digits levels (1240 headings) and the HS-6 digits level (5106 sub-headings). Once excluded countries that do not report data we use a balanced database of 155 countries (see the list in Annex A1).

When looking at a trade flow as reported by both partners, i.e. the import flows reported by the importer and its corresponding mirror export flow as reported by the exporter country, 3 cases can occur:

- "*both"* trade partners report a strictly positive flows, a cif/fob ratio can be computed;
- "*orphan imports*": the importer country declares a strictly positive flow with no corresponding flow reported by the exporter,

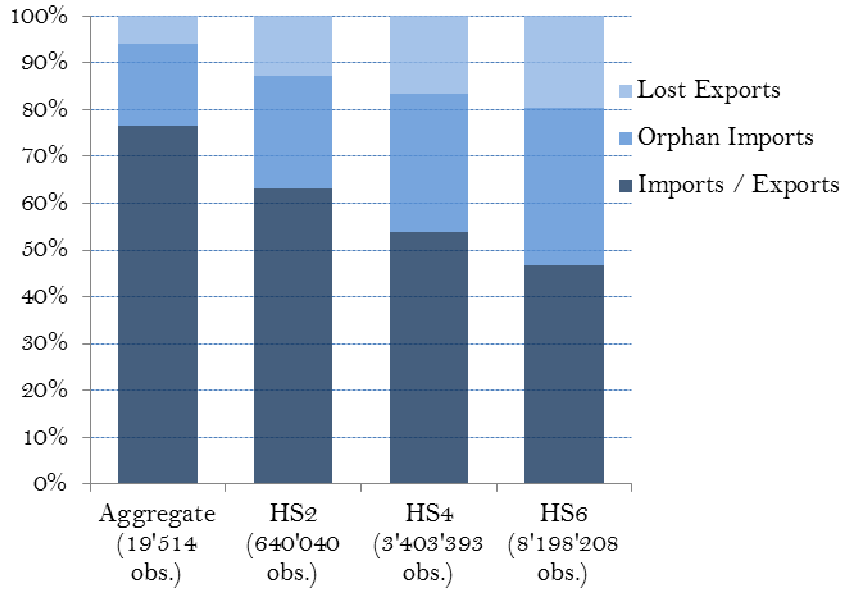
[6] The year 2008 was chosen preferably to a more recent one to avoid mixing our results with those from the economic crisis.

- "*lost export*": the exporter country declares a strictly positive flow with no corresponding flow reported by the importer.

Note that we exclude from the sample the case where both trade partners report a zero trade flows.

As reported in Figure 1, when looking at the aggregate level (one flow per country-pair), strictly positive trade flows are reported by both trade partners for 77% of the lines while 17% correspond to "orphan imports" and the remaining 6% to "lost exports".

Figure 1 : mirror data at the country pair level, different product aggregation levels, 2008



Note: we exclude from the sample the HS6 lines with a zero trade flow reported by both partners
Source: Author's computation based on UN COMTRADE database

While global matching is observed at the total level, with the 1.06 ratio previously defined in table 1, substantial discrepancies prevail. The latter are all the larger as the level of disaggregation is important, as evidenced by the reduction of the dark blue area of the figure 1 along with the successive disaggregation from total bilateral trade to the 6 digits level. The dark blue area represents the share of the lines for which each line is reported by both the importing side and exporting side.

While 77% of the lines are reported by the both partners at the aggregate level, only 47% of the lines are (still) matching at the 6 digits level.

The remaining 53% lines are either "lost exports" (20%), i.e trade flows declared by the "exporting side" (e.g from country A to country B) while not reported at all from the "importing side" (country B), or to "orphan imports" (33%), i.e imports declared by the "importing side" (e.g country B from country A) but not reported by the "exporting side".

Note that a line that is reported by the two partners is not necessarily properly recorded as the cif/fob ratio may still be under or over-valued or even misclassified. Indeed, even data at the 6 digits level are still aggregated data over all the transactions between the two countries for a specific product and could henceforth hide misreporting. So 53% represent the very minimum number of misclassified lines, either on destination/origin, or on the category of product.

When both trade flows are available (reported as "import/export" in Figure 1), we can compute the cif/fob ratio. The distribution of this ratio for the different levels of product aggregation is reported in Table 2. Note that the larger the level of disaggregation, the lower the median cif/fob ratio and the larger the dispersion of the observations (see the inter-quartile gap in Table 2).

Table 2 : distribution of the cif/fob ratio at the country-pair level, different product aggregation levels, 2008

Percentile	25%	50%	75%	obs.
Aggregate	0.73	1.16	2.16	14'942
HS2	0.58	1.07	2.03	403'406
HS4	0.47	1.04	2.18	1'834'998
HS6	0.38	1.01	2.36	3'819'876

Source: Author's computation based on UN COMTRADE database

Either the gaps and misreporting are purely erratic notably due to the weak quality of the COMTRADE data, which would be rather problematic given the extensive use of these data in the international trade literature, or there may have trends and drivers to be identified in the way these gaps occur. This is what we study in the next section.

3.3. Fact #2: country pair characteristics as drivers of the gap

For each country-pair level we compute 2 measures of the gap in mirror data:

- The percentage of HS6 lines with orphan imports (capturing pure misclassification), the total number of HS6 lines being computed from lines with at least one of the trade partner reporting a strictly positive trade flow;

- The median of cif/fob ratio computed for HS6 lines having strictly positive trade flows reported by both trade partner (capturing both misclassification and misevaluation), i.e. computed from the 3'819'876 lines with both imports and exports as reported in Figure 1. Note that using the weighted average of cif/fob ratio instead of the median does not change the qualitative results reported below.

We report in Table 3 some basic descriptive statistics of our two measures of the gap in mirror data. Note that the average level of the percentage of orphan imports is here reported as 55% at the country-pair level while Figure 1 had showed an average level of 33%. This highlights that orphan imports occur relatively more frequently in "small" country-pairs, i.e. in country-pairs

that trade a relatively small number of HS6 lines. The same explanation holds for the gap between the median of the cif/fob ratio reported in Table 2 and Table 3.

Table 3 : some descriptive statistics of our two measures of the gap in mirror data at the country-pair level, 2008

Variable	Obs	Median	Min	Max
% of orphan imports	19514	0.533	0	1
Median of the cif/fob ratio	12871	1.057	0.0000296	2544340

Source: Author's computation based on UN COMTRADE database

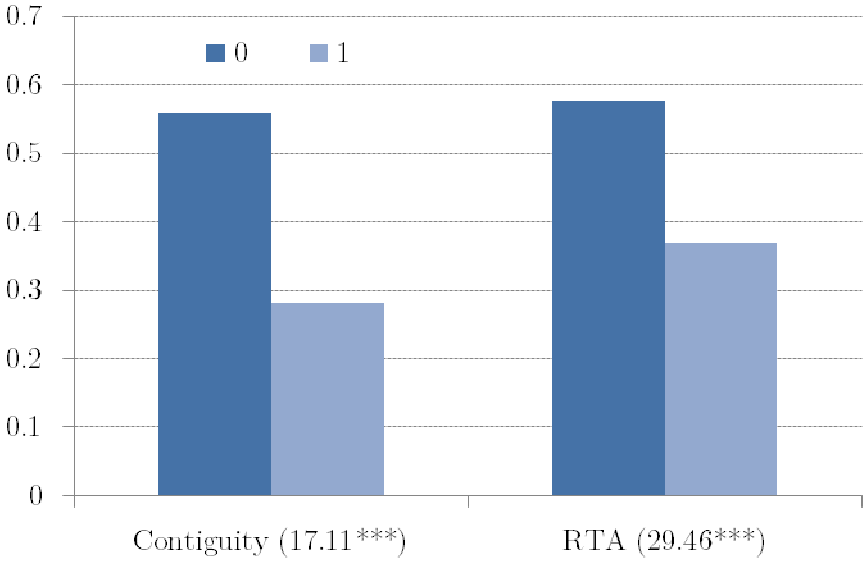
We then correlate these 2 measures with the following country-pair characteristics resulting from the discussion on the literature review section 2 on what we called the "logistic" reasons for discrepancies (notably transportations costs and re-exports);

- Contiguity between the two trading partners;
- The implementation of a regional trade agreement (RTA);
- The bilateral distance between most populated cities;
- The % of reexport at the country pair level (value of total reexport from *j* to *i* in % of total export value from *j* to *i*).

The 3 first variables are extracted from the CEPII's database, the latter comes from UN-COMTRADE database.

We propose two types of descriptive statistics depending on the nature of the qualitative or quantitative characteristics we test. First, when the country-pair characteristics is a dummy variable, we propose a two-group mean-comparison test to test that the percentages of orphan imports have the same mean within the two groups defined by the dummy variables (t-test). This is done in Figure 2 for the contiguity and RTA dummy variables. In the case of the cif/fob ratio we will test differences in median rather than means due to outliers in the cif/fob ratios. We then compute the Pearson's chi-squared test, a nonparametric test that tests the null hypothesis that the medians between two groups are not significantly different. Results for the contiguity and RTA dummy variables are reported in Figure 3.

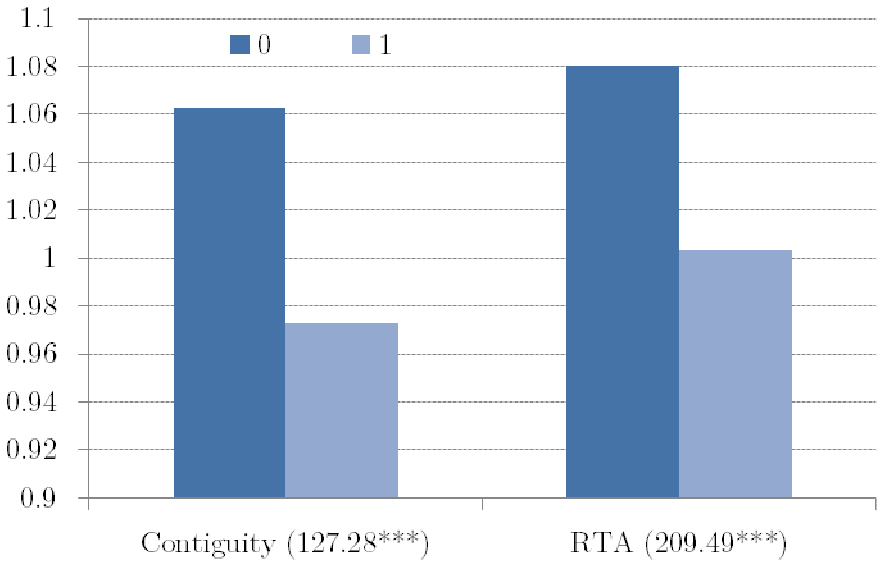
Figure 2: Average percentage of lines with orphan imports at the country-pair level related to the contiguity and RTA dummy variables, 2008



Note: 19'232 observations, t-test in parenthesis with ***, ** and *: significant at the 1%, 5% and 10% level respectively, Contiguity and RTA from CEPII database.
Source: Author's computation based on UN COMTRADE database.

Both contiguous country-pair and "countries part of a same RTA" have significantly lower average percentages of lines with orphan imports, which can illustrate facilitated administrative burden and fewer cases of misclassifications in the both situations due to a stronger proximity and collaboration.

Figure 3: Median of the cif/fob ratio at the country-pair level related to the contiguity and RTA dummy variables, 2008



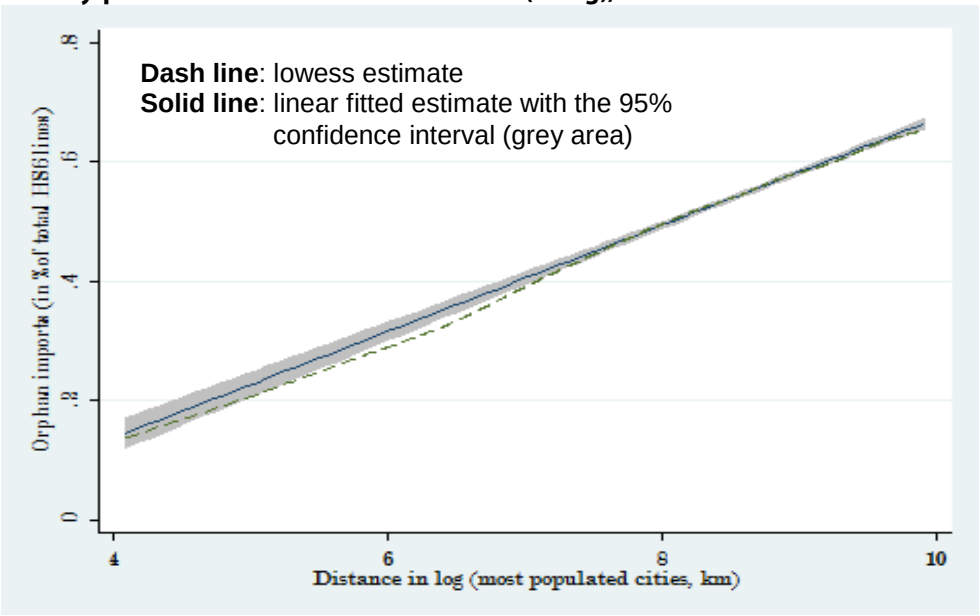
Note: 12'768 observations, Pearson chi2(1) test in parenthesis with ***, ** and *: significant at the 1%, 5% and 10% level respectively, Contiguity and RTA from CEPII database.
Source: Author's computation based on UN COMTRADE database.

The cif/fob ratio is also significantly lower for contiguous country pairs or countries that are both part of the same RTA, which is consistent with reduced trade costs by comparison with either non-contiguous countries or with countries with no trade preferential agreements.

Second, when the country-pair characteristics are continuous variables, we perform on the same graph both (i) a nonparametric analysis with a locally weighted regression (Lowess) and (ii) a parametric analysis with an OLS linear fitted estimate with the 95% confidence interval. These parametric and nonparametric regressions are reported in Figure 4 and Figure 5 for the bilateral distance (in log) and confirm that we can approximate the relation between both the percentage of orphan imports and the log of bilateral distance on one hand and the log of cif/fob ratio and the log of bilateral distance on the other hand by a linear form. The positive correlation between the (log of the) bilateral distance and the cif/fob ratio is as expected: the longer the distance, the higher the transportation costs and the larger the cif/fob ratio. The positive correlation between the share of orphan imports and the bilateral distance might reflect the fact that the likelihood of misreporting will be all the higher as there is a large distance, as a longer distance might increase the probability of reexport or transhipment (see box 1) and decrease the likelihood of the two countries being part of a same RTA and of collaboration between the two countries.

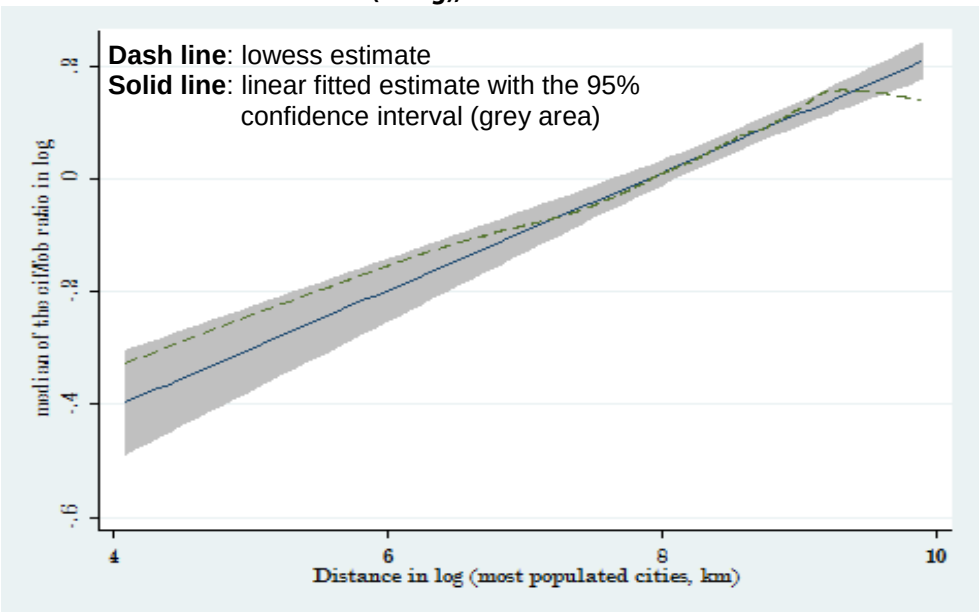
The parametric and nonparametric regressions are reported in Figure 6 and Figure 7 for % of reexport at the country pair level. The observed correlations between the percentage of reexports of the partner country at the country pair level with both the share of orphan imports and the cif/fob ratio are as expected. As stated in section 2, given the UN recommendations to record transhipment trade flows, orphan imports are assumed to be all the higher as there is reexport, as the importing side will report the country of origin and not the country of provenance. This will lead either to an increased likelihood to have a mis-match (figure 6) or to under-estimate the cif/fob ratio (figure 7).

Figure 4: Correlation between the percentage of lines with orphan imports at the country-pair level and the bilateral distance (in log), 2008



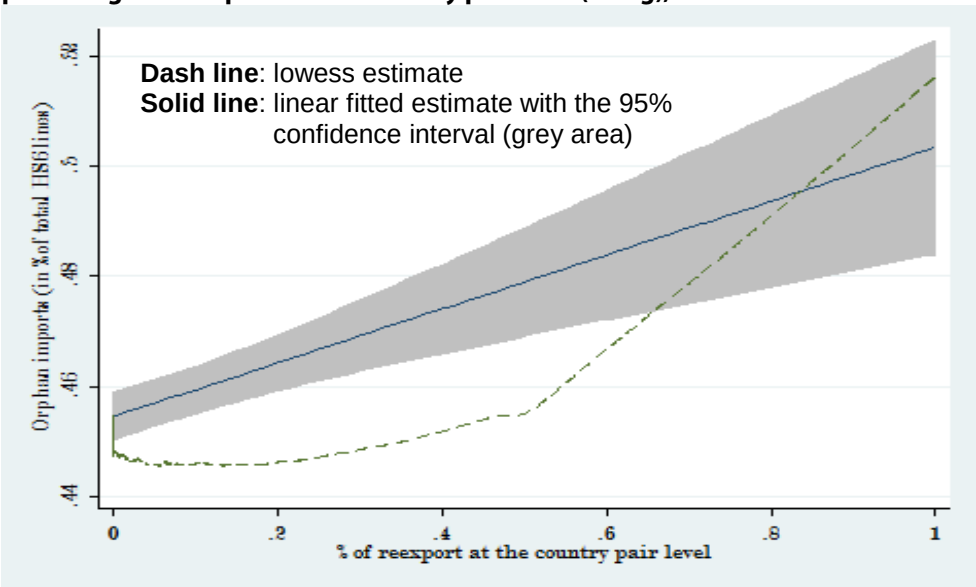
Note: 19'232 observations, distance from CEPII database.
Source: Author's computation based on UN COMTRADE database.

Figure 5: Correlation between the cif/fob ratio median (in log) at the country-pair level and the bilateral distance (in log), 2008



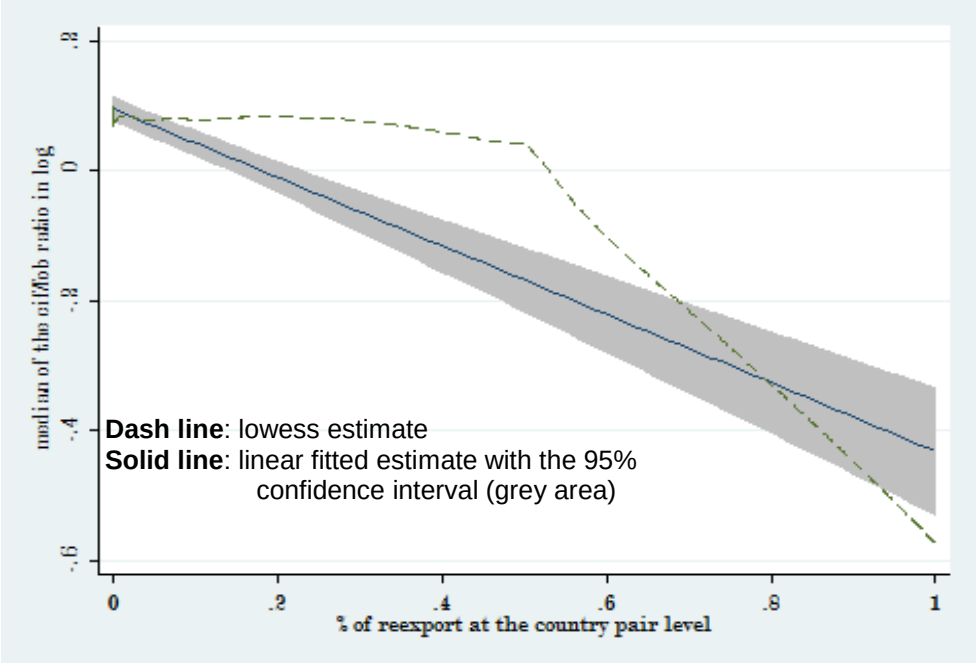
Note: 12'768 observations, distance from CEPII database.
Source: Author's computation based on UN COMTRADE database.

Figure 6: Correlation between the percentage of lines with orphan imports and the percentage of reexports at the country pair level (in log), 2008



Note: 16'103 observations.
Source: Author's computation based on UN COMTRADE database.

Figure 7: Correlation between the cif/fob ratio median (in log) and the percentage of reexports (in log) at the country pair level, 2008



Note: 12'871 observations.
Source: Author's computation based on UN COMTRADE database.

3.4. Fact #3: importer country characteristics as driver of the gap

Going further to try to identify patterns and drivers of the discrepancies in the cif/fob ratio, we rank the countries depending on the value of their cif/fob ratio as well as on the share of orphan imports or lost exports they have (see Appendix 2 for additional descriptive statistics over the cif/fob ratio distribution at the country level). We will then present correlations at the country level.

Table 4 : Countries in the Top 10, 2008

10 countries with the lowest CIF/FOB ratios		10 countries with the highest CIF/FOB ratios		10 countries with the highest shares of orphan imports		10 countries with the largest share of lost exports	
KGZ	0.20	BTN	2.36	LSO	0.99	LSO	0.987
BEN	0.21	KIR	1.73	BWA	0.84	AFG	0.901
LSO	0.22	UGA	1.50	NAM	0.77	LCA	0.875
GMB	0.33	ETH	1.50	BLR	0.66	BHS	0.649
LBY	0.44	BWA	1.50	AFG	0.65	CAF	0.632
KHM	0.54	ARM	1.46	COG	0.55	PAN	0.514
COG	0.57	MWI	1.43	TON	0.53	SLB	0.502
NGA	0.57	FJI	1.37	GMB	0.50	TON	0.420
LVA	0.61	MLI	1.30	SDN	0.47	MLT	0.417
PAN	0.61	LTU	1.27	SRB	0.45	MRT	0.414

Note: correspondence between country codes and names are reported in annex A1
Source: Author's computation based on UN COMTRADE database.

Table 4 shows that most of the countries with the poorest ranking are low income countries. Moreover, landlocked countries are over-represented amongst these countries: respectively 6 and 5 amongst 10 for the highest cif/fob ratios and for the the highest shares of orphan imports. This is however not a surprise as landlockedness represents a significant increase in transportation costs (see Grigoriou 2007) and often results in transhipment.

Pairwise correlations from country characteristics

We propose the same exercise as in the previous section, i.e. we correlate the percentage of HS6 lines with orphan imports and the median of cif/fob ratio with importer country characteristics at the country-pair level such as;

- Landlockness from CEPII.

- GDP per capita, 2008, data from WDI

- Unweighted average tariff, 2008. We use the unweighted measure from Nicita and Olarreaga (2013), as using weighted tariff would bias the result due to imports with low tariff. We expect first a positive correlation between the unweighted average tariff and the share of orphan imports as the bigger the tariff the higher the misclassification likelihood.

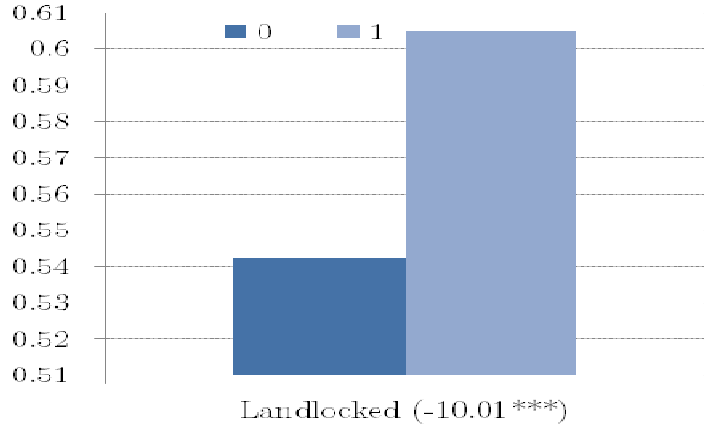
On the other hand, a negative correlation is expected with the cif/fob ratio as the higher the tax burden, the bigger the incentive to either misclassify or under-value the imported goods.

- Collected import duties in % of total imports from WDI in 2008.

- CPIA transparency, accountability, and corruption in the public sector rating (1=low to 6=high, CPIA: The Bank's Country Policy and Institutional Assessment Indicators) from the WDI, 2008. Transparency, accountability, and corruption in the public sector assess the extent to which the executive can be held accountable for its use of funds and for the results of its actions by the electorate and by the legislature and judiciary, and the extent to which public employees within the executive are required to account for administrative decisions, use of resources, and results obtained. The three main dimensions assessed here are (1) the accountability of the executive to oversight institutions and of public employees for their performance, (2) access of civil society to information on public affairs, and (3) state captured by narrow vested interests.

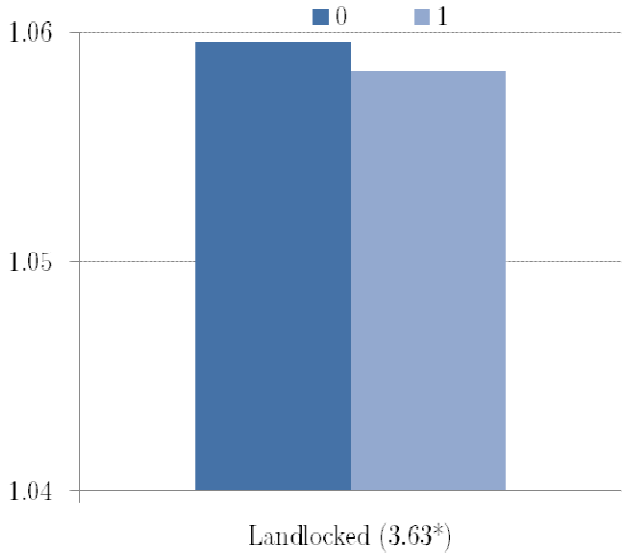
- Stock of inflows FDI (in log) from the UNCTAD database. FDI stocks are presented at book value or historical cost, reflecting prices at the time when the investment was made. For a large number of economies, FDI stocks are estimated by either cumulating FDI flows over a period of time or adding flows to an FDI stock that has been obtained for a particular year from national official sources or the IMF data series on assets and liabilities of direct investment. While very cautious on the interpretation from a basic correlation, we would expect a positive relationship between the stock of inflows FDI and the cif/fob ratios as FDI could be positively correlated with transfer pricing policies from the foreign companies at the origin of the FDI, which might imply overvaluation of the imported goods (see section 2).

Figure 8: Average percentage of lines with orphan imports at the country-pair level related to landlockedness of the importer country, 2008



Note: 19'232 observations, t-test in parenthesis with ***, ** and *: significant at the 1%, 5% and 10% level respectively, landlockedness from CEPII database.
Source: Author's computation based on UN COMTRADE database.

Figure 9: Median of the cif/fob ratio at the country-pair level related to landlockedness of the importer country, 2008



Note: 11'064 observations, Pearson chi2(1) test in parenthesis with ***, ** and *: significant at the 1%, 5% and 10% level respectively, landlockedness from CEPII database.
Source: Author's computation based on UN COMTRADE database.

As expected, landlocked countries have on average higher shares of orphan imports (figure 8). On the other hand, landlocked countries have statistically significantly smaller cif/fob ratios than non-landlocked countries (figure 9), contrary to what was expected from the higher transportation costs landlocked countries have to cope with. This may illustrate a selection bias in the calculation of their cif/fob ratio. Indeed, the landlocked countries' international trade statistics may by definition suffer significantly more than the others from misreporting due to transhipment and re-export (through the countries of transit), which is fully consistent with the substantially higher shares of orphan imports landlocked countries have. This misreporting leads to over-weight the trade of the landlocked countries with their contiguous countries (i.e with the countries of transit), the transportation costs of which being significantly lower (see figure 3).

Figure 10 and 11 evidence the positive correlation between the per capita GDP, capturing the level of economic development of the country, and the registration of its trade flow; the richer the country, the better the quality of the administration and the lower the share of lines with orphan imports.

Correlations of figure 12 and 13 on the relationship between the unweighted average tariff and the percentage of lines with orphan imports or the cif/fob ratios are as expected; the higher the average tariffs, the higher the orphan imports (misclassification for tariff evasion) and the lower the cif/fob ratios (under-valuation for tariff evasion).

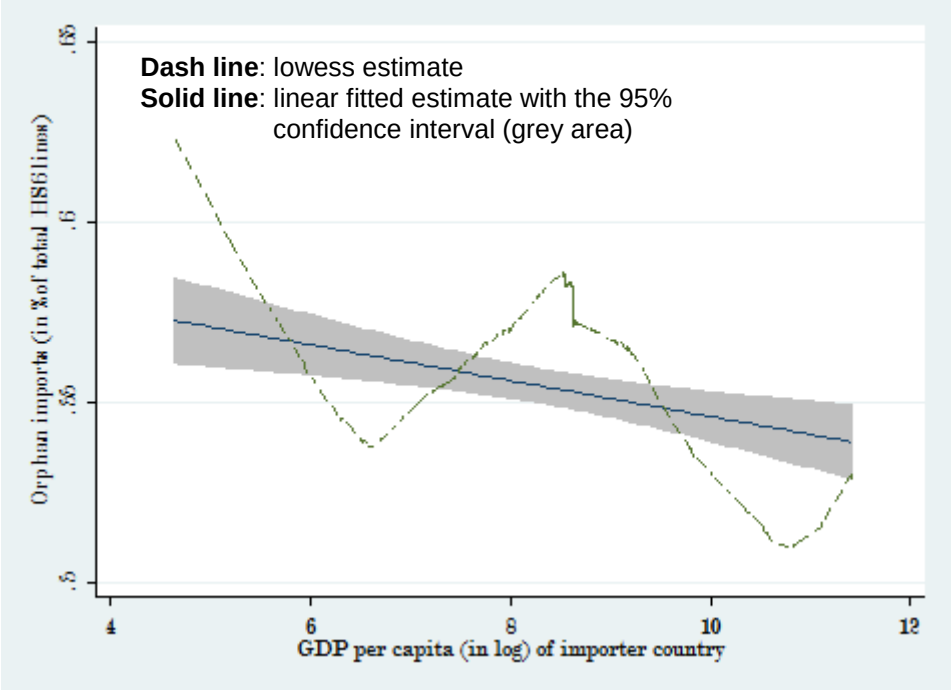
Interpreting the ratio relating collected import duties to total imports as a measure of the efficiency of the Customs administration to collect duties and taxes, figure 14 and 15 may reveal

that the fraud will shift from misclassification to misevaluation as this ratio increases. Indeed, the percentage of lines with orphan imports decreases with the increase in the collected import duties while the cif/fob ratio diminishes.

Figures 16 and 17 highlight a negative correlation between the stock of inflows FDI and the share of orphan imports, a positive correlation with the cif/fob ratio. We observe positive relationship between the stock of inflows FDI and the cif/fob ratios. This could result from transfer pricing policies from foreign companies: the higher the FDI the higher the overvaluation of the associated imported goods. However, higher inflows FDI could be simply correlated with per capita GDP which is also positively correlated with the cif/fob ratios. One should also control for the level of tariff as it represents the cost of over-valuating the imports for these companies. We will run econometric estimates to control for such mechanisms.

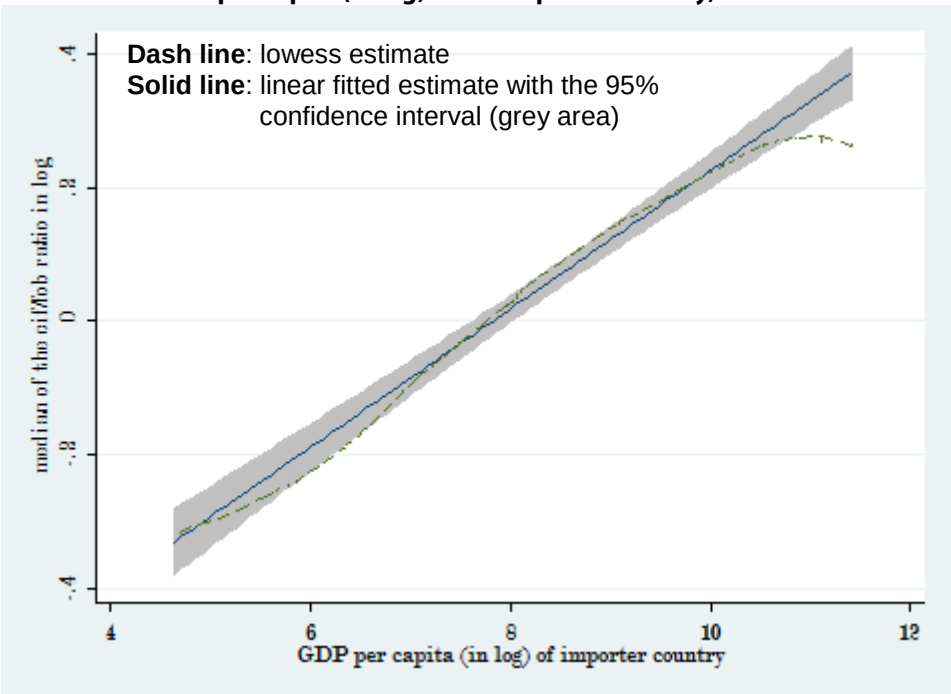
Last, figures 18 and 19 evidence a positive relationship between the level of corruption and both the share of lines with orphan imports and the cif/fob ratio, which is consistent with Yeats (1995) and Javorcik and Narciso (2012) who explain that customs officials might have incentives to overvalue or shift the classification towards high tariff lines in corrupted environment. This is also consistent with our comments resulting from figure 14 and 15 as the efficiency of the administration would be expected to be negatively correlated with the corruption.

Figure 10: Correlation between the percentage of lines with orphan imports at the country-pair level and the GDP per capita (in log) of the importer country, 2008



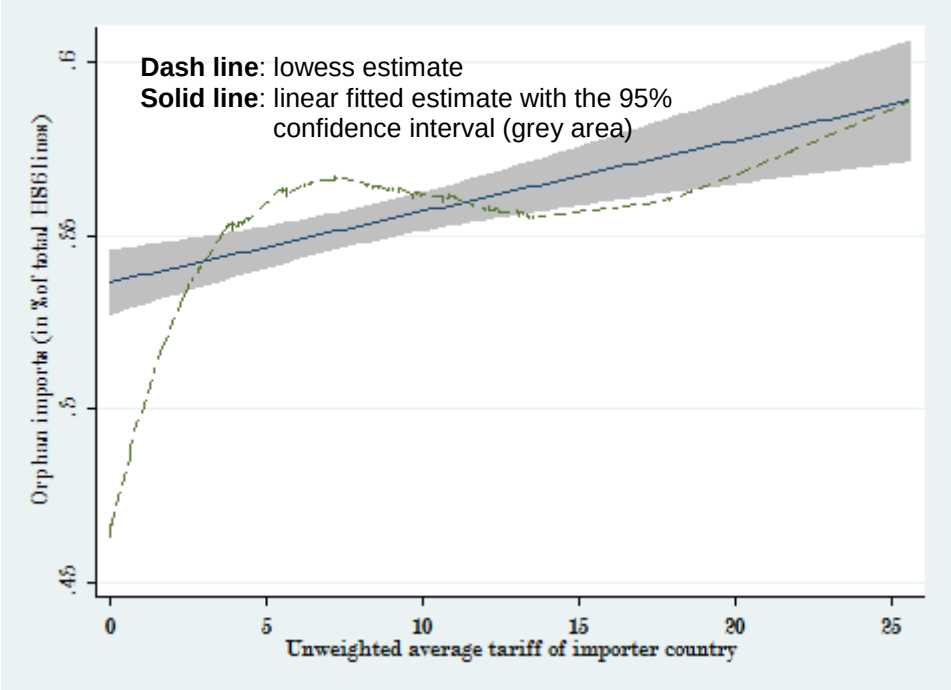
Note: 17'267 observations.
Source: Author's computation based on UN COMTRADE database.

Figure 11: Correlation between the cif/fob ratio median (in log) at the country-pair level and the GDP per capita (in log) of the importer country, 2008



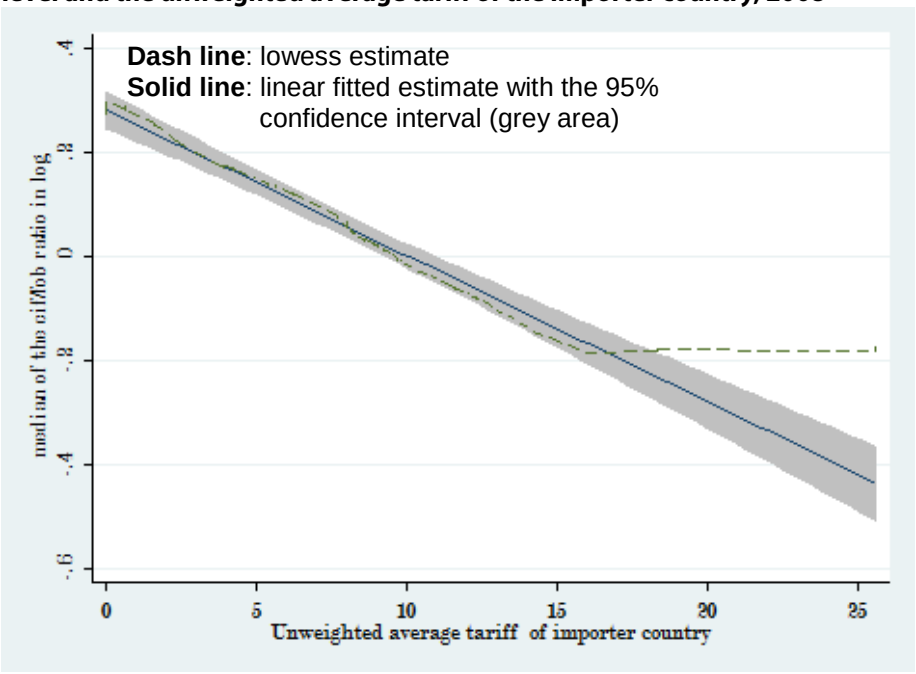
Note: 11'713 observations.
Source: Author's computation based on UN COMTRADE database.

Figure 12: Correlation between the %age of lines with orphan imports at the country-pair level and the unweighted average tariff of the importer country, 2008



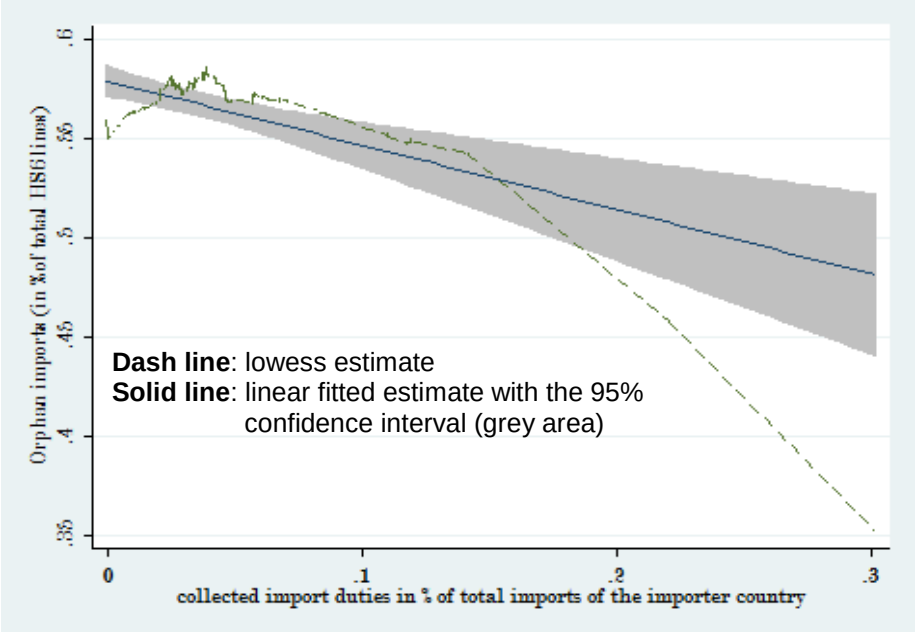
Note: 14'824 observations.
Source: Author's computation based on UN COMTRADE database.

Figure 13: Correlation between the cif/fob ratio median (in log) at the country-pair level and the unweighted average tariff of the importer country, 2008



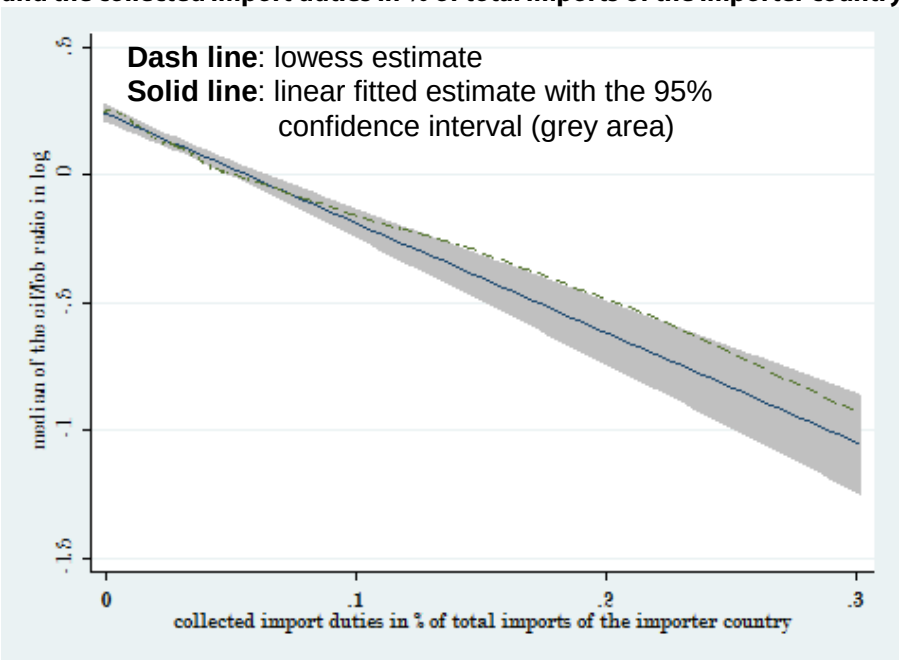
Note: 9'941 observations.
Source: Author's computation based on UN COMTRADE database.

Figure 14: Correlation between the %age of lines with orphan imports at the country-pair level and the collected import duties in % of total imports of the importer country, 2008



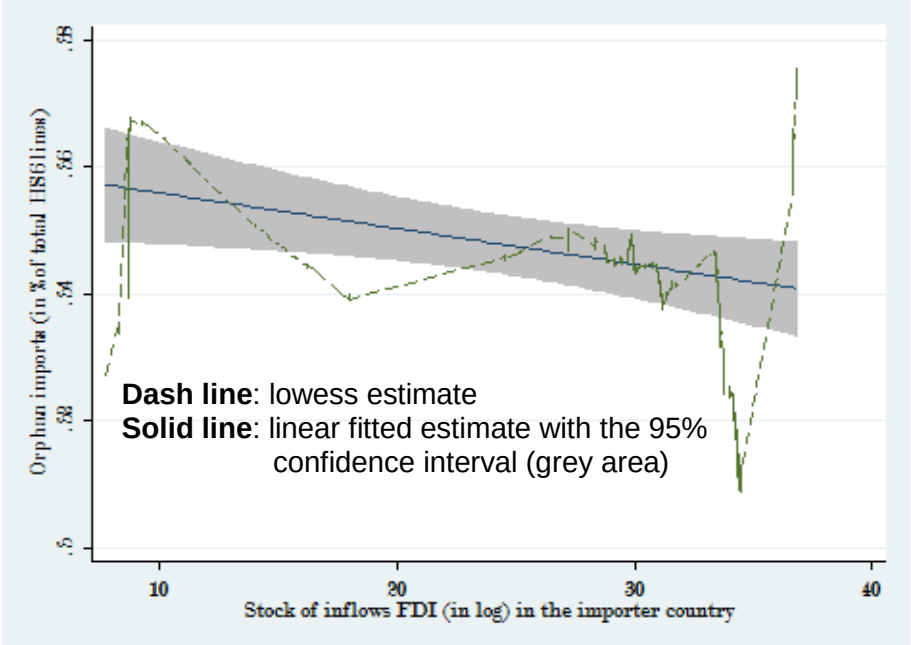
Note: 11'429 observations.
Source: Author's computation based on UN COMTRADE database.

Figure 15: Correlation between the cif/fob ratio median (in log) at the country-pair level and the collected import duties in % of total imports of the importer country, 2008



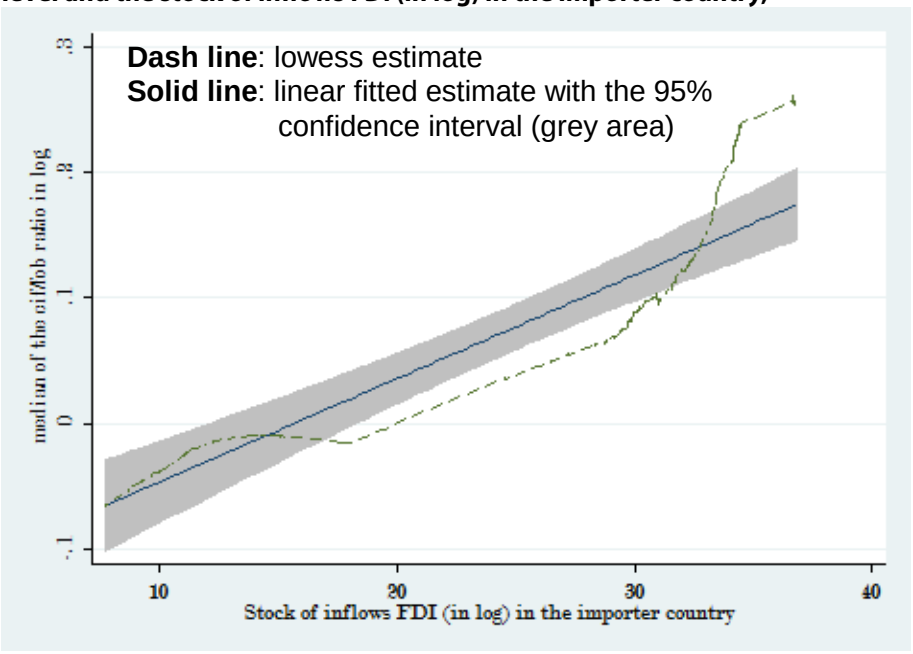
Note: 7'782 observations.
Source: Author's computation based on UN COMTRADE database.

Figure 16: Correlation between the percentage of lines with orphan imports at the country-pair level and the stock of inflows FDI (in log) in the importer country, 2008



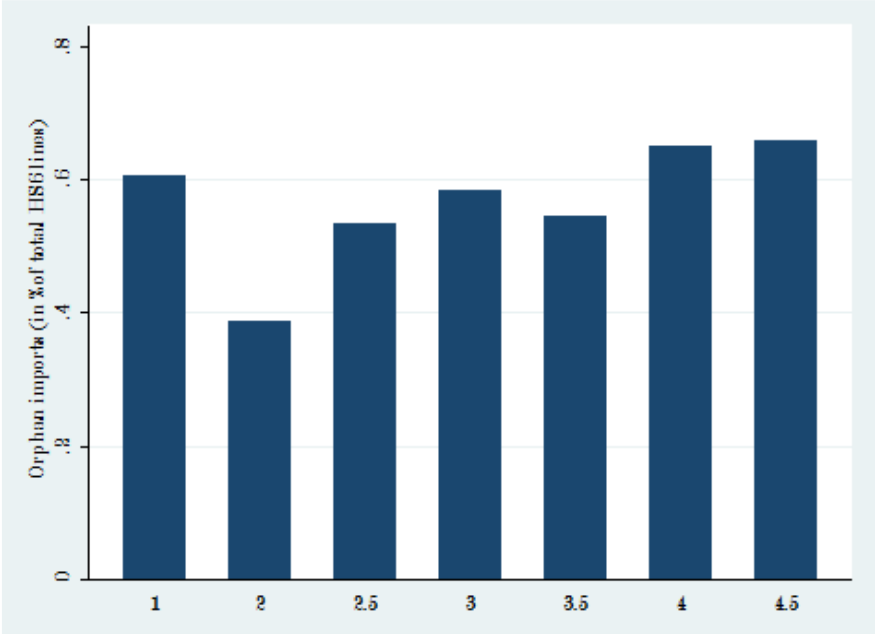
Note: 19'140 observations.
Source: Author's computation based on UN COMTRADE database.

Figure 17: Correlation between the cif/fob ratio median (in log) at the country-pair level and the stock of inflows FDI (in log) in the importer country,



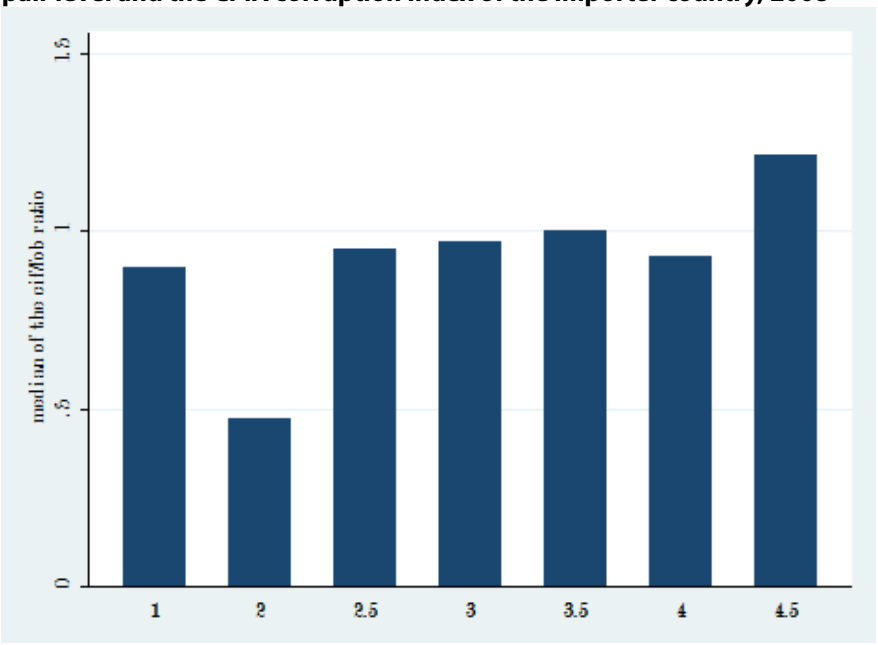
Note: 12'738 observations.
Source: Author's computation based on UN COMTRADE database.

Figure 18: Correlation between the percentage of lines with orphan imports at the country-pair level and CPIA corruption index of the importer country, 2008



Note: 6'140 observations.
Source: Author's computation based on UN COMTRADE database.

Figure 19: Correlation between the cif/fob ratio median (in log) at the country-pair level and the CPIA corruption index of the importer country, 2008



Note: 3'484 observations.
Source: Author's computation based on UN COMTRADE database.

3.5. Fact #4: product characteristics as driver of the gap

A part of the discrepancy observed in the cif/fob ratio is obviously not independent from the traded products, as some products face higher tariffs than others etc. We show hereafter in the following tables (tables 5 and 6) the ranking of the products with the highest share of orphan imports and of lost exports. 214 products amongst the 5'104 considered at the 6 digit level have 100% as share of orphan imports, meaning there is systematically a mismatch between what the importers have declared and the reported exports. We report those which correspond to the most exported products in 2008.

Note that mineral fuels and oils (chapter 27) is at the third position of this ranking, two products amongst the ten are connected to textile (chapter 61), and two to electrical machinery equipment (chapter 85).

Table 5 : 10 products with the highest share of orphan imports corresponding to the most exported products over the world, 2008

Corresponding Chapter	Orphan Imports
Electrical mchy equip parts thereof; sound recorder etc	854219
Ships, boats and floating structures.	890130
Mineral fuels, oils & product of their distillation; etc	270760
Raw hides and skins (other than furskins) and leather.	410410
Nuclear reactors, boilers, mchy & mech appliance; parts	846911
Art of apparel & clothing access, knitted or crocheted.	610110
Paper & paperboard; art of paper pulp, paper/paperboard	480521
Art of apparel & clothing access, knitted or crocheted.	610311
Optical, photo, cine, meas, checking, precision, etc	900921
Oil seed, oleagi fruits; miscell grain, seed, fruit etc	121210
Electrical mchy equip parts thereof; sound recorder etc	851992

Source: Author's computation based on UN COMTRADE database.

The same ranking is processed for the lost exports, computed from the importing countries point of view. 186 products amongst the 5'104 considered at the 6 digits level have 100% as share of lost exports, meaning that there are exports that the partners report but that never match what the importing countries declare. We report those which correspond to the ten most imported products in 2008.

Table 6 : 10 products with the highest share of lost exports corresponding to the most imported products over the world, 2008

Corresponding chapter	lost exports
Electrical mchy equip parts thereof; sound recorder etc	854219
Fertilisers.	310320
Raw hides and skins (other than furskins) and leather.	410410
Ships, boats and floating structures.	890130
Optical, photo, cine, meas, checking, precision, etc	900921
Man-made filaments.	540320
Optical, photo, cine, meas, checking, precision, etc	900912
Fertilisers.	310410
Organic chemicals.	293010
Electrical mchy equip parts thereof; sound recorder etc	852822

Source: Author's computation based on UN COMTRADE database.

Note there are again two products related to the electrical machinery equipment (chapter 85). Moreover, the first product for the lost exports is the same as the first product for the orphan imports (chapter 85, subheading 854219) which could be linked to re-export processes or to deliberate misclassification on the origin or destination.

Tables 7 and 8 report the products with the highest and lowest cif/fob ratios to describe the extreme discrepancies there are at the product level. See appendix 3 for further descriptive statistics at the 6 digit level. Note the large range of values of the cif/fob ratio; the largest (chemical products) ratio is beyond 400 while the smallest (iron and steel) is only 0.08.

Table 7 : 10 products with the highest CIF/FOB ratios, 2008

Products (chapters and subheadings)	hs6code	CIF/FOB ratio
Chemical products (ch38; naphtenic acids)	382420	495.76
Tanning/dyeing extracts (ch32; pigments and preparations based on chromium compounds)	320630	419.39
Optical, photo (ch90; profile projectors)	903130	182.87
Wood and articles of wood (ch44; fribreboard)	441111	85.306
Wood and articles of wood (ch44; other)	441191	68.73
Eletrical machinery equipment (ch85; floor poolishers)	850920	47.13
Eletrical machinery equipment (ch85; Monolithic digital integrated circuits)	854213	39.83
Oil seed, oleagi fruits and grains (ch12; safflower seeds)	120760	39.03
Cotton (ch.52; of yarns of different colours)	521042	36.78
Organic chemicals (ch29. derivates containing only sulpho groups)	290820	28.01

Source: Author's computation based on UN COMTRADE database.

Table 8 : 10 products with the lowest CIF/FOB ratios, 2008

Products (chapter and subheadings)	hs6code	CIF/FOB ratio
Iron and steel (ch72; high speed steel)	722520	0.08
Natural/cultured pearls, prec stones & metals, coin etc (ch71; precious metal whether or not plated or clad with precious metal)	711419	0.09
Electrical mchy equip parts thereof; sound recorder etc (ch.85; cards incorporating a magnetic stripe)	852460	0.10
Wood and articles of wood; wood charcoal (ch44; Other, with at least one outer ply of non-coniferous wood)	441222	0.10
Pharmaceutical products (ch. 30; Containing hormones or other products of heading No. 29.37 but not containing antibiotics :-- Containing insulin)	300331	0.11
Headgear (ch65; Felt hats and other felt headgear)	650300	0.11
Cotton (ch52; printed, 3 thread or 4 thread twill)	520853	0.11
Man-made staple fibres (ch55; other)	550490	0.13
Wool, fine/coarse animal hair, horsehair yarn & fabric (ch51; coarse animal hair)	510540	0.13
Art of apparel and clothing (ch61; suits of wool or finae animal hair)	610411	0.13

Source: Author's computation based on UN COMTRADE database.

4. Empirical strategy: one step further

We now look for multivariate correlations using econometric panel data at the country-pair product level. We first study the probability that a given HS6 product line at the country pair level correspond to an orphan import. We then focus on the cif-fob ratio for lines with both imports and its mirror reported data.

Orphan import: A probit analysis

Our dependent variable, denoted M_{ijk}^{orphan}, is a binary variable that takes the value of 1 if the HS6 lines k exported from country i to country j corresponds to an orphan import, otherwise 0. The comparison group for orphan imports consists of the lines at the country-pair level with no orphan imports, i.e. with the trade flows either reported by both countries or only by the exporting countries (lost exports). We estimate the following non-linear probit model where the binary dependant variable (M_{ijk}^{orphan}) is regressed on the several determinants already presented in the previous section, over 7'105'937 observations for the year 2008.

$$\Pr\left(M_{ijk}^{orphan}=1\right) = \phi\left(\begin{array}{l}\alpha_0+\alpha_1\ln(Dist_{ij})+\alpha_2\ln(GDPpc_i)+\alpha_3\ln(GDPpc_j)+\alpha_4 Landlocked_i+\alpha_5 Re\text{-}export_{ij} \\ \quad +\alpha_6 RTA_{ij}+\alpha_7 Contiguity_{ij}+\alpha_8 ln(FDI_i)+\alpha_9 Tariff_i+\alpha_{10} CPI_i+\sum_{k\in HS2}\beta_k D_k\end{array}\right) \quad (4)$$

where:

ϕ is the cumulative normal distribution;

$\ln(Dist_{ij})$ is the (log of the) bilateral distance between most populated cities of countries i et j;

$\ln(GDPpc_{i(j)})$ is the (log of the) per capita GDP of country i(j) in 2008;

$Landlocked_i$ is a dummy variable that takes the value of 1 if country i is landlocked (no direct access to the sea), otherwise 0;

$Re\text{-}export_{ij}$ is the percentage of total re-export at the country pair level in 2008 (value of total re-export from *j* to *i* in % of total export value from *j* to *i*);

RTA_{ij} is a dummy variable taking the value of 1 if a regional trade agreement is implemented between i and j in 2008, otherwise 0;

$Contiguity_{ij}$ is a dummy taking the value of 1 if countries i and j share a common border, otherwise 0;

$\ln(FDI_i)$ is the (log of the) stock of inflows FDI in country i in 2008;

$Tariff_i$ is the unweighted average tariff of country i in 2008;

CPI_i is the Corruption Perceptions Index of country i in 2008;

$\sum_{k \in HS2} D_k$ is a full set of chapter (i.e. HS-2 digit level) effects to capture product specific effect that is country-pair invariant.

Table 9 reports the recomputed marginal impacts corresponding to the estimation of (4). Hence, the reported coefficients give directly the change in the probability that an orphan import occurs for a change of 1 unity in the corresponding continuous variable or a discrete change of the corresponding dummy variable from 0 to 1.

We alternatively estimate equation (4) with and without the average tariff and the CPI score index variables as these variables significantly reduce the size of the sample.

The coefficients are all significantly different from zero and with the expected signs.

The proxies for transportation costs are all emphasizing a positive correlation between the transport cost and the probability to have orphan imports: having a shorter bilateral distance, a RTA with the partner, being contiguous to the partner and having an access to the sea all contribute to reduce the transportation costs and to decrease the probability to have orphan imports. Notably, the coefficient associated with RTA_i is significantly negative (at the 1% level), even once controlled for contiguity and bilateral distance: on average, implementing a RTA between two countries decreases *all other things being equal* the probability of experiencing an orphan import by 5.4 percentage points. This may reflect an increased cooperation between customs officials of countries sharing a same RTA plus decreased trade costs, both resulting in fewer misclassifications. Note that the coefficient remains significantly negative once controlled for the tariffs (col.2).

The GDP per capita of the two countries are both negatively correlated with the probability to have orphan imports, consistently with the assumption than higher level of economic development is connected with a better registration system for the imports.

As expected, we observe a conditional correlation between the share of reexport between i and j and the probability of orphan import: an additional percentage point in the share of reexport from *j* to *i* (in % of total export value from *j* to *i*) increases the probability to have a reported orphan import by 3.7 percentage points (significant at a 5% level).

Tariff is positively correlated with the probability to have orphan imports, possibly due to higher incentives to misclassify the product by shifting it to lower tariff lines.

Corruption is also positively correlated with the probability to have orphan imports which makes sense as corruption goes with lower efficiency in customs administration and higher discretion.

Table 9 : Probit Estimates of Orphan imports, 2008

	(1)			(2)		
$\Pr\left(M_{ijk}^{orphan}=1\right)$	Coeff		std. Err.	Coeff		std. Err.
ln(Dist$_{ij}$)	0.049	***	*(0.0019)*	0.041	***	*(0.0017)*
ln(GDPpc$_i$)	-0.032	***	*(0.0015)*	-0.028	***	*(0.0017)*
ln(GDPpc$_j$)	-0.015	***	*(0.0037)*	-0.015	***	*(0.0037)*
Landlocked$_i$	0.091	***	*(0.0032)*	0.105	***	*(0.0038)*
Re-export$_{ij}$	0.037	**	*(0.0193)*	0.032	***	*(0.0195)*
RTA$_{ij}$	-0.054	***	*(0.0039)*	-0.047	***	*(0.0033)*
Contiguity$_{ij}$	-0.116	***	*(0.0033)*	-0.137	***	*(0.0025)*
ln(FDI$_i$)	0.001	***	*(0.0001)*	0.001	***	*(0.0001)*
Tariff$_i$				0.003	***	*(0.0002)*
CPI$_i$				0.002	*	*(0.0010)*
nber obs.	7'105'937			5'399'988		
product FE (HS2)	yes			yes		
Correctly classified						
cutoff = 0.34 (mean)	60.44%			59.26%		
cutoff = 0.5	68.05%			67.26%		

Note: Estimation by probit. Coefficients are marginal probabilities evaluated at the sample means. Standard errors below coefficients: heteroscedasticity consistent and adjusted for HS2 product clustering; *** : p=0.01, **: p=0.05, *: p=0.1 . See text for definition of variables.
Source: Author's computation based on UN COMTRADE database.

Finally, the 2 last lines of Table 9 reports statistics of the predictive ability of this Probit model. While there is no theoretical reason to chose this cutoff, it is customary to take a prediction rule with a threshold value like p* = 0.5, on the basis that we would predict a 1 if the model says a 1 is more likely than a zero :

$$\hat{M}_{ijk}^{orphan}=1 \text{ if the predicted probability } \phi > p^*$$

Taking this criterion, Table 9 suggests that the model successfully predicts 68% of total observations. Since this measure of goodness of fit depends on the cutoff selected to classify the predicted, one should also test robustness to another cutoff. Because of the unbalanced sample with many more 0s than 1s, we also set a p* equal to the proportion of 1's in the sample (which corresponds to the average predicted probability in the sample of 0.339).

Ratio cif-fob: Within estimator

We now focus on the cif-fob ratio for lines with both the imports and the reported mirror (exports) data. The baseline regression (1) counts 3'466'902 observations, and is reduced to 2'590'433 observations (2) once included the tariffs and the measure of corruption (CPI). The regression is processed with the WITHIN estimator, including products specific effects at the HS6 level.

$$\ln(R_{ijk}^{cif/fob}) = \alpha_0 + \alpha_1 \ln(Dist_{ij}) + \alpha_2 \ln(GDPpc_i) + \alpha_3 \ln(GDPpc_j) + \alpha_4 Landlocked_i + \alpha_5 re\exp ort_{ij} + \alpha_6 RTA_{ij} + \alpha_7 Contiguity_{ij} + \alpha_8 \ln(FDI_i) + \alpha_9 Tariff_i + \alpha_{10} CPI_i + u_k + \varepsilon_{ijk} \quad (5)$$

Table 10 : regression results, within estimator, 2008

	(1)			(2)		
$\ln\left(R_{ijk}^{cif/fob}\right)$	Coeff		std. Err.	Coeff		std. Err.
ln(Dist$_{ij}$)	0.075	***	*(0.0025)*	0.068	***	*(0.0027)*
ln(GDPpc$_i$)	0.092	***	*(0.0024)*	0.085	***	*(0.0030)*
ln(GDPpc$_j$)	-0.104	***	*(0.0036)*	-0.113	***	*(0.0037)*
Landlocked$_i$	0.066	***	*(0.0044)*	0.074	***	*(0.0056)*
Re-export$_{ij}$	-0.876	***	*(0.0190)*	-0.922	***	*(0.0192)*
RTA$_{ij}$	0.006	***	*(0.0041)*	0.019	***	*(0.0044)*
Contiguity$_{ij}$	-0.131	***	*(0.0043)*	-0.187	***	*(0.0049)*
ln(FDI$_i$)	0.005	***	*(0.0002)*	0.005	***	*(0.0002)*
Tariff$_i$				-0.013	***	*(0.0006)*
CPI$_i$				-0.014	***	*(0.0015)*
nber obs.	3'466'902			2'590'433		
product FE (HS6)	yes			yes		

Note: Estimation by within estimator. Coefficients are marginal probabilities evaluated at the sample means. Standard errors below coefficients: heteroscedasticity consistent and adjusted for HS6 product clustering; *** : p=0.01, **: p=0.05, *: p=0.1 . See text for definition of variables.
Source: Author's computation based on UN COMTRADE database.

Variables controlling for the transportation costs are all significantly different from zero and with the expected sign: cif/fob ratio increases with the distance and if the importing country is landlocked, while it is decreased if the trading countries are contiguous. The coefficient associated to the RTA is positive and significantly different from zero: once controlled for the bilateral distance, the contiguity and the tariffs which are all correlated with both the cif/fob ratio and the RTA, the remaining impact of setting a RTA on the cif/fob ratio is positive. This may

reflect the fact that undervaluation of the imported trade is significantly reduced by the RTA[7], which is consistent with the results of the probit estimate of a strengthened collaboration between countries part of a same RTA improving the efficiency of the administration.

The reexports are as expected negatively correlated with the cif/fob ratio, as the importing country should not report its imports from the "reexporting partner".

The stock of FDIs is still positively correlated with the cif/fob ratio, even once controlled for the per capita GDP and the tariffs.

Including the tariffs and the corruption decrease importantly the size of the sample but the coefficients remain very close to regression (1). Tariffs are negatively correlated with cif/fob ratio, reflecting a potential deliberate undervaluation or reclassification to lower tariff lines of the imported goods in the context of high tariffs. Corruption is negatively correlated with the cif/fob ratio, which may suggest collusion between customs officials and importers to evade tariffs by undervaluing the imported goods.

5. Concluding comments

This paper reviews and investigates the reasons why the reported bilateral trade from an importing country may not match the corresponding trade flow by the exporting country, apart from the only transportation costs. There are first valid structural or logistic reasons other than transportation costs to explain the gap between matched partner data, including notably re-export and transhipment issues, or countries having different trade reporting systems. Moreover, deliberate misreporting can, beyond these logistic reasons, significantly contribute to explain the discrepancies, either through misevaluation or misclassification of the imported goods, notably to evade tariffs and taxes.

An empirical analysis over the UN-Comtrade data evidences the margin of the gap from two measures of misreporting; the cif/fob ratio, which is a direct measure of the gap, and the share of the import lines that miss their counterpart from the export-side, defined as the “orphan imports”. Interestingly enough, these two measures are both significantly correlated with not only macroeconomic variables such as bilateral distance, landlockedness, GDP per capita and implementation of regional trade agreements, but also with variables capturing potential incentives for misreporting like; (1) average tariffs, measuring the incentive to use misevaluation or misclassification of the shipments to evade tariffs , (2) foreign direct investments, measuring the case for profit-shifting or capital flight from misevaluation of the transactions, and (3) corruption which leads to informal operations and tax evasion including both misevaluation and misclassification. Relying on a set of such variables and applied to more

[7] It could also result from the combination between the lack of mandatory documentation at customs in a RTA and the differences in trade flows thresholds to record transactions from a country to another inside a RTA, contributing to missing a part of the inside-RTA trade.

than 7 million of observations over products at the 6 digits level in 2008, our model accurately predicts up to 68% of the orphan imports against 32% chances for a pure random selection.

This suggests that there is a significant relationship between cross-border trade flow misreporting and fraud opportunities to evade tariffs and taxes. Far from being erratically driven, the gap can consequently be partly predicted by macroeconomic variables which should be used to better target the informal trade.

References

Anderson, J. and E.van Wincoop, 2003, Gravity with Gravitas: A Solution to the Border Puzzle, *American Economic Review*.

Anson J., Cadot O. and M. Olarreaga, 2006, Tariff Evasion and Customs Corruption: Does Pre-Shipment Inspection Help? *The B:E Journals in Economic Analysis and Policy*, Manuscript 1600.

Bernard A., Jensen J. and P. Schott, 2006, Transfer pricing by U.S.-based multinational firms, NBER 12493.

Bhagwati J., 1964, On the underinvoicing of imports, *Oxford Bulletin of Economic and Statistic* (November).

Bhagwati J., 1967, Fiscal policies, the faking of foreign trade declarations, and the balance of payments, *Oxford Bulletin of Economic and Statistics* (February).

Baier, Scott L. and Jeffrey H. Bergstrand, 2001, The Growth of World Trade: Tariffs, Transport Costs, and Income Similarity, *Journal of International Economics* 53(1):1–27

Carrere C., 2006, Revisiting the Effects of Regional Trading Agreements on trade flows with Proper Specification of the Gravity Model, *European Economic Review***,** February 2006, 50(2):223-247

Easterly W. and A. Reshef, 2010, African export successes: surprises, stylized facts and explanations, NBER working paper 16597.

Federico G. and A.Tena, 1991, On the accuracy of foreign trade statistics 1909-1935; Morgenstern revisied, Explorations in Economic History, 28, 259-2730

Ferrantino M. and Z. Wang, 2008, Accounting for discrepancies in bilateral trade: the case of China, Hong Kong and the United States, *China Economic Review*, 19, 502-520.

Fisman R. and S.-J. Wei, 2004, Tax rates and and tax evasion, evidence from missing imports. *Journal of Political Economy*, 112(2), 471-496.

Gaulier G. and S. Zignago, 2010, BACI: International Trade Database at the Product-level: the 1994-2007 Version, CEPII, document de travail n°2010-23.

Grigoriou C., 2007, Landlockedness, Infrastructure and Trade: New Estimates for Central Asian countries, Policy Research Working Paper 4335, The World Bank, Development Research Group, Trade Team.

Guo D., 2009, Mirror statistics of international trade in manufacturing goods: the case of China, UNIDO Research and statistics branch, working paper 19/2009.

Hamanaka S., 2011, Whose trade statistics are correct? Multiple mirror comparison techniques: a test of Cambodia, *Journal of Economic Policy Reform* 15(1), 33-56.

Hummels D. and V.Lugovskyy, 2006, Are Matched Partner Trade Statistics a Usable Measure of Transportation Costs?, *Review of International Economics*, 14(1):69-86.

Javorcik B. and G. Narciso (2008), Differentiated products and evasion of import tariffs, *Journal of International Economics*, 76 (2).

Javorcik B. and G. Narciso, 2012, WTO Accession and Tariff Evasion, mimeo.

Makhoul B. and S. Otterstrom, 2010, Exploring the accuracy of international trade statistics, *Applied Economics*, 1998 (30), 1603-1616

Mishra P., Subramanian A. and P. Topalova, 2007, Policies, enforcement and Customs evation: evidence from India, *IMF working papers* 07/60, International Monetary Fund.

Rose A., 1991, Whay has Trade Grown Faster than Income? *Canadian Journal of Economics* 24(2): 417-27.

L. Rotunno and P-L. Vézina, 2011, The rise and fall of (Chinese) African apparel exports," CSAE Working Paper Series 2012-12, Centre for the Study of African Economies, University of Oxford.

Sheik M., 1974, Underinvoicing of imports in Pakistan, *Oxford Bulletin of Economics and Statistics*, 36 (4).

Yeats A., 1978, On the Accuracy of Partner Country Trade Statistics, *Oxford Bulletin of Economics and Statistics*, 40(4):341-61.

Yeats A., 1995, Are partner-country statistics useful for estimating "missing" trade data? *Policy Research Working Paper 1501*. Washington DC: World Bank.

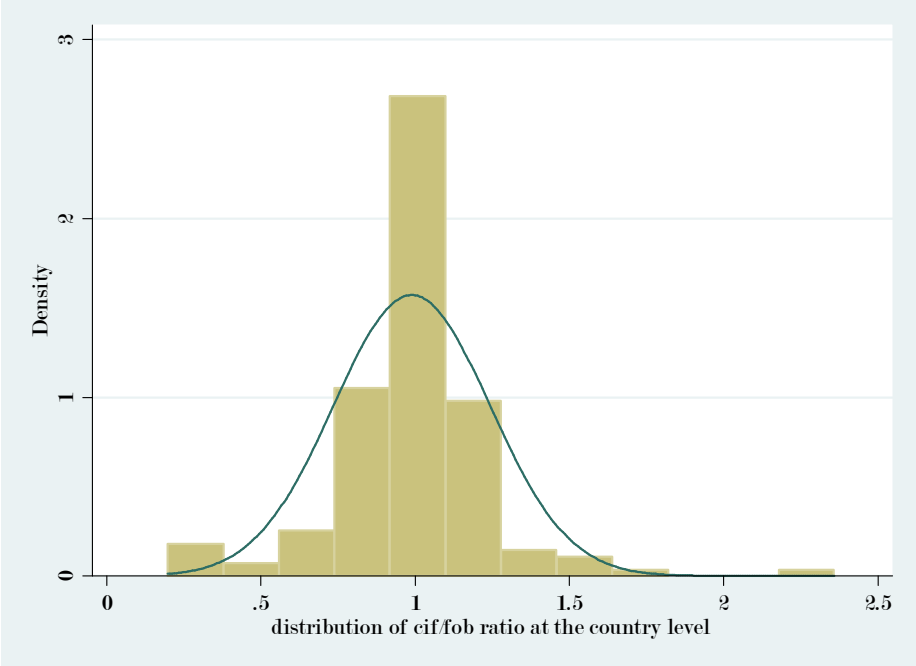
Annex A1: list of countries included in the sample

Country Name	ISO-Code	Country Name	ISO-Code	Country Name	ISO-Code	Country Name	ISO-Code	Country Name	ISO-Code	Country Name	ISO-Code
Afghanistan	AFG	China	CHN	Guatemala	GTM	Malaysia	MYS	Rwanda	RWA	Uruguay	URY
Albania	ALB	Colombia	COL	Guinea	GIN	Maldives	MDV	Samoa	WSM	Venezuela, Bolivarian Republic of	VEN
Algeria	DZA	Congo, the Democratic Republic of the	COD	Guyana	GUY	Mali	MLI	Sao Tome and Principe	STP	Viet Nam	VNM
Argentina	ARG	Costa Rica	CRI	Hong Kong, SAR	CHN, HKG SAR	Malta	MLT	Senegal	SEN	Yemen Rep.	YEM
Armenia	ARM	Côte d'Ivoire	CIV	Hungary	HUN	Mauritania	MRT	Serbia	SRB	Zambia	ZMB
Aruba	ABW	Croatia	HRV	Iceland	ISL	Mauritius	MUS	Seychelles	SYC	Zimbabwe	ZWE
Australia	AUS	Cyprus	CYP	India	IND	Mexico	MEX	Singapore	SGP		
Austria	AUT	Czech Republic	CZE	Indonesia	IDN	Moldova, Republic of	MDA	Slovak Rep.	SVK		
Azerbaijan	AZE	Denmark	DNK	Ireland	IRL	Morocco	MAR	Slovenia	SVN		
Bahamas	BHS	Dominica	DMA	Israel	ISR	Mozambique	MOZ	Solomon Islands	SLB		
Bahrain	BHR	Dominican Republic	DOM	Italy	ITA	Namibia	NAM	South Africa	ZAF		
Barbados	BRB	Ecuador	ECU	Jamaica	JAM	Netherlands	NLD	Spain	ESP		
Belarus	BLR	Egypt, Arab Rep.	EGY	Japan	JPN	New Caledonia	NCL	Sri Lanka	LKA		
Belgium	BEL	El Salvador	SLV	Jordan	JOR	New Zealand	NZL	Sweden	SWE		
Belize	BLZ	Estonia	EST	Kazakhstan	KAZ	Nicaragua	NIC	Switzerland	CHE		
Benin	BEN	Ethiopia	ETH	Kenya	KEN	Niger	NER	Syrian Arab Republic	SYR		
Bhutan	BTN	Fiji	FJI	Kiribati	KIR	Nigeria	NGA	Tanzania, United Republic of	TZA		

Bolivia, Plurinational State of	BOL	Finland	FIN	Korea, Republic of	KOR	Norway	NOR	Thailand	THA
Bosnia and Herzegovina	BIH	France	FRA	Kuwait	KWT	Oman	OMN	Togo	TGO
Botswana	BWA	French Guiana	GUF	Kyrgyz Rep.	KGZ	Pakistan	PAK	Tonga	TON
Brazil	BRA	French Polynesia	PYF	Latvia	LVA	Panama	PAN	Trinidad and Tobago	TTO
Bulgaria	BGR	Gabon	GAB	Lebanon	LBN	Paraguay	PRY	Tunisia	TUN
Burkina Faso	BFA	Gambia, The	GMB	Lesotho	LSO	Peru	PER	Turkey	TUR
Burundi	BDI	Georgia	GEO	Libya	LBY	Philippines	PHL	Turks and Caicos Islands	TCA
Cambodia	KHM	Germany	DEU	Lithuania	LTU	Poland	POL	Uganda	UGA
Cameroon	CMR	Ghana	GHA	Luxembourg	LUX	Portugal	PRT	Ukraine	UKR
Canada	CAN	Greece	GRC	CHN, Macao SAR	CHN, MAC SAR	Qatar	QAT	United Arab Emirates	ARE
Central African Republic	CAF	Greenland	GRL	Madagascar	MDG	Romania	ROU	United Kingdom	GBR
Chile	CHL	Grenada	GRD	Malawi	MWI	Russian Federation	RUS	United States	USA

Annex A2. Some descriptive statistics at the country level

Figure 20: distribution of cif/fob ratio at the country level, 2008



Note: Half of the countries have a CIF/FOB ratio < 1 (75 countries out of 153)
Interquartile = 0.9 to 1.087, Interdecile = 0.75 to 1.22, Mean and median not significantly different one from the other (respectively 0.99 and 1.0)

Figure 21: distribution of the share of orphan imports at the country level, 2008

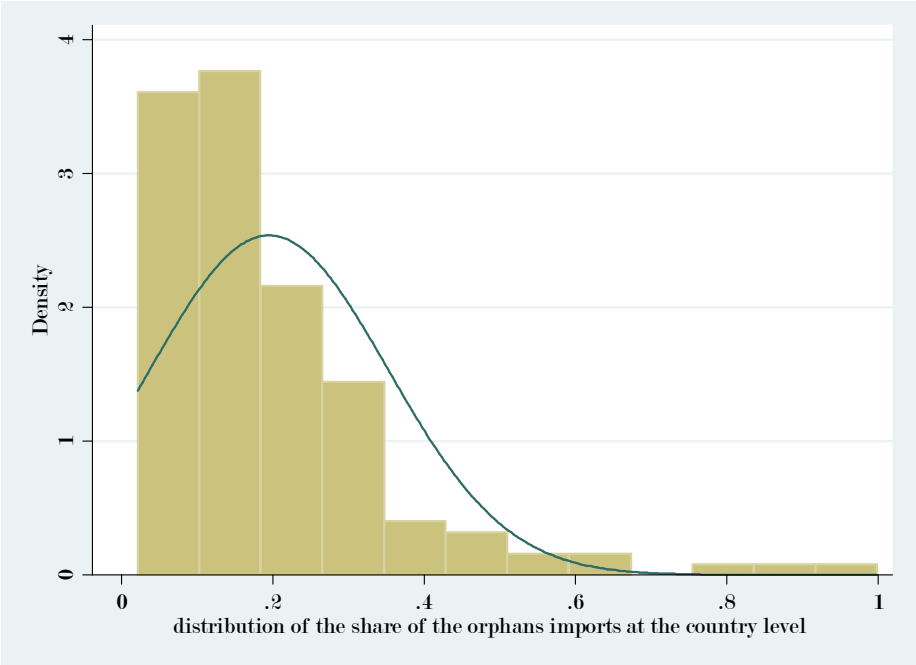
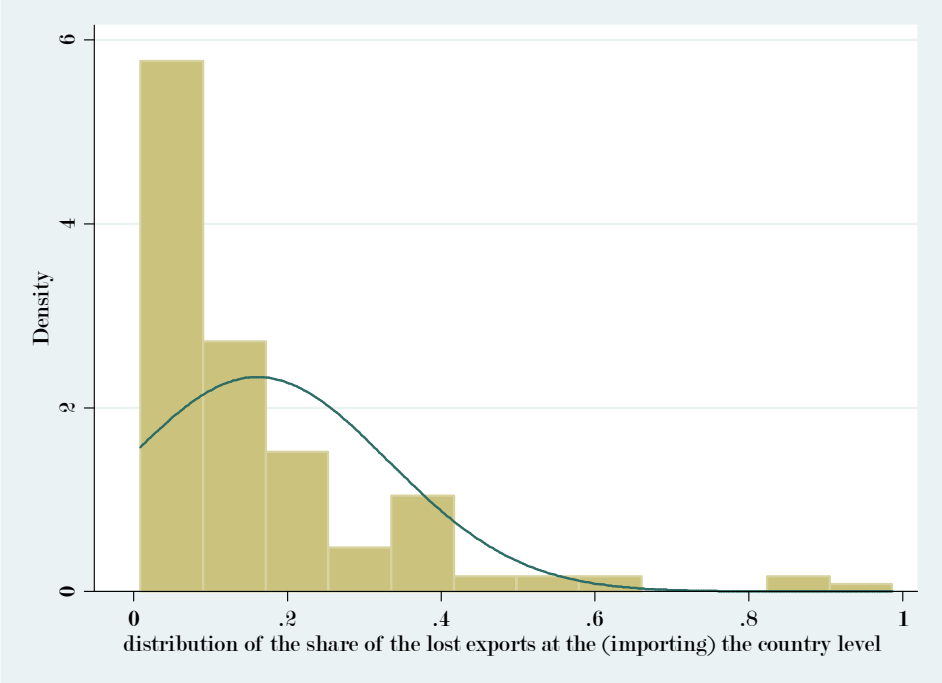


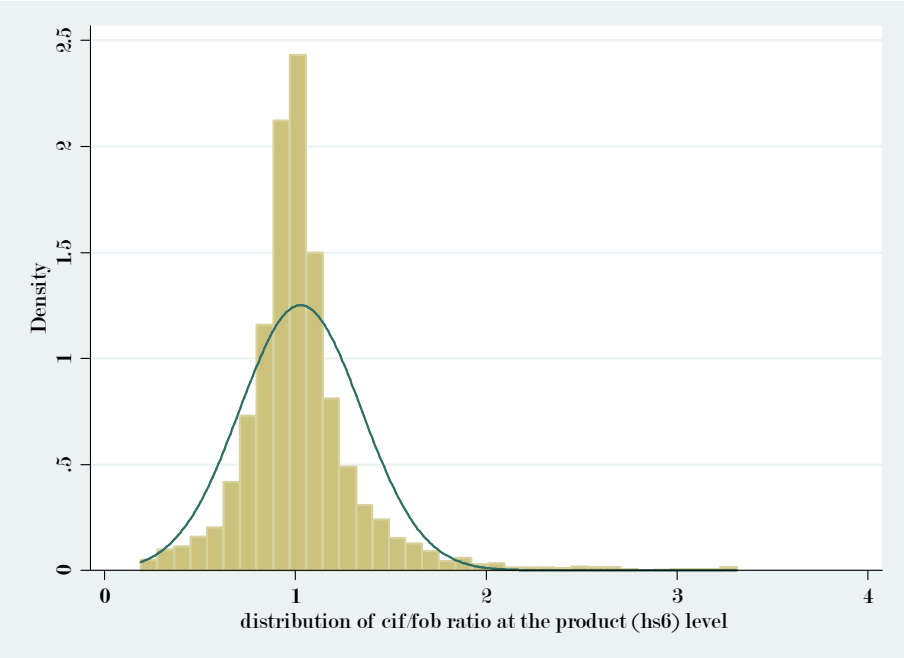
Figure 22: distribution of the share of lost exports at the country level, 2008



Note: the ranking are correlated at more than 75% (which makes sense: most of the lost exports correspond to orphan imports).

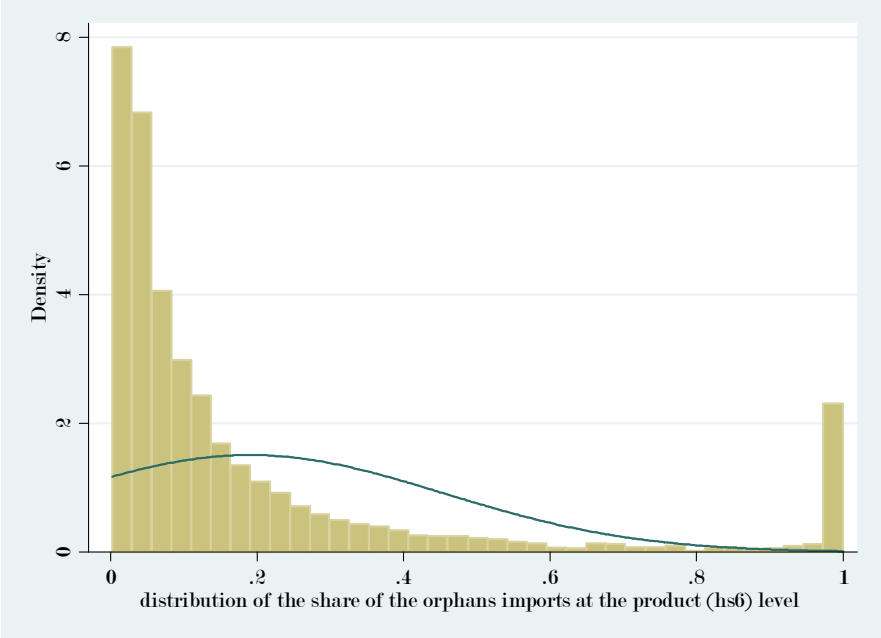
Annex A3. Some descriptive statistics at the product (HS6) level

Figure 23: distribution of the cif/fob ratio at the product level, 2008



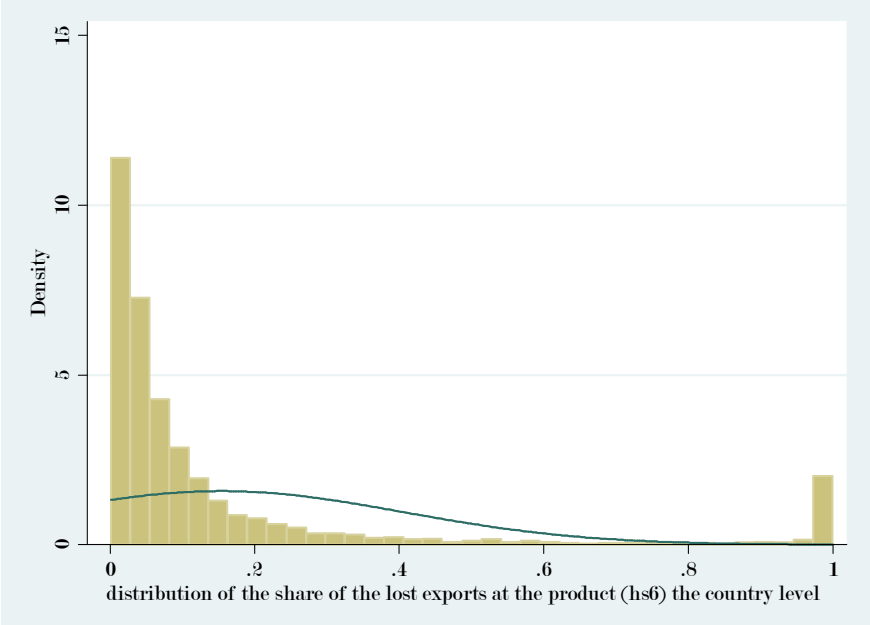
Note: More than half of the 4'882 products have a CIF/FOB ratio < 1 (2'533)
Interquartile = 0.88 to 1.125, Interdecile = 0.715 to 1.37, Mean 1.37, median 0.99

Figure 24: distribution of the share of the orphan imports at the product level, 2008



Note: 214 products are systematically orphan imports, i.e. never reported by the exporters while declared by the importers.

Figure 25: distribution of the share of the lost exports at the product level, 2008



Note: 186 products are systematically orphan imports, i.e. never reported by the exporters while declared by the importers.

"Sur quoi la fondera-t-il l'économie du monde qu'il veut gouverner? Sera-ce sur le caprice de chaque particulier? Quelle confusion! Sera-ce sur la justice? Il l'ignore."

Pascal

FERDI

Created in 2003 , the **Fondation pour les études et recherches sur le développement international** aims to promote a fuller understanding of international economic development and the factors that influence it.

Contact
www.ferdi.fr
contact@ferdi.fr
+33 (0)4 73 17 75 30