UNITED NATIONS UNIVERSITY
UNU-WIDER

WIDER Working Paper 2022/24

Measuring illicit financial flows

A gravity model approach to estimate international trade misinvoicing

Lourenço S. Paz*

March 2022

Abstract: Illicit financial flows have recently attracted the attention of academia, practitioners, and multilateral organizations who consider them harmful to economic development. Some observers suggest that many of these flows occur via the misinvoicing of international trade transactions. This study develops a novel methodology based on the gravity model of international trade to estimate illicit financial flows using publicly available product-level international trade data. It contributes to the literature by estimating product-level transportation and insurance costs and providing, for the first time, an estimation for all countries of an upper bound of the values of export and of import over- and under-misinvoicing separately, without resorting to assumptions about trade statistics reliability. Application of the methodology is illustrated using six-digit product-level UN Comtrade data for 2013–16. The results indicate that misinvoicing is not confined to a few products or countries but is a widespread phenomenon that deserves future research.

Key words: gravity model, illicit financial flows, international trade, misinvoicing, transportation and insurance cost

JEL classification: C23, F14, F24, O17

Note: figures and tables at the end of the paper

---

* Baylor University, Waco, Texas, United States of America, email lourenco_paz@baylor.edu

This study has been prepared within the UNU-WIDER project Detecting and countering illicit financial flows that is implemented in collaboration with the University of Copenhagen. The project is part of the Domestic Revenue Mobilization programme, which is financed through specific contributions by the Norwegian Agency for Development Cooperation (Norad).

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9267-155-6

https://doi.org/10.35188/UNU-WIDER/2022/155-6

Typescript prepared by Lesley Ellen.

United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland, Sweden, and the United Kingdom as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

1 Introduction

Increasing global integration has placed illicit financial flows (IFFs) centre stage in the international economic development policy debate. Such flows reduce the tax base and lower tax revenues which could be used to supply public goods such as education, healthcare, and infrastructure (Addison et al. 2018). Additionally, the cross-border nature of IFFs can drain a country's foreign currency reserves and compromise its balance of payments, especially in the least developed countries. Furthermore, IFFs can be used by the richest households to evade taxes and to hide wealth, which magnifies inequality in society and raises concerns about state legitimacy. This perception that IFFs are an important barrier to improving the living standards of developing nations made it target 16.4 of the United Nations Sustainable Development Goals (UNSDGs).

Unfortunately, the study of IFFs is hindered by difficulties in defining and measuring them. As Forstater (2018) and Cobham and Janský (2020) discussed, there is no single accepted definition of IFFs. Some well-known definitions, such as those in OECD (2014), United Nations (2016), and World Bank (2016), all view IFFs as cross-border transfers of financial capital in contravention of national or international laws, regardless of the legality of the origin of these financial resources.[1] This study adopts this common characterization as the definition of IFFs. In addition, Cobham and Janský (2020) pointed out that IFFs can only be measured indirectly because of the illegal nature of both the remittance and some of the remitted capital. These hurdles may be the reason why the UNSDGs have so far defined only one indicator for IFFs: '(16.4.1) the total value of inward and outward illicit financial flows (in current USD)'. Furthermore, the target levels for this value have not yet been established.

Carbonnier and de Cadena (2015) identified three ways in which money is moved across borders with limited scrutiny by government authorities. The first is via the financial system, using disguised remittances. The second is through the physical movement of banknotes, currency, gold, and gemstones. The third is through the misinvoicing of international trade transactions. Global Financial Integrity (Kar and Spanjers 2014: 2) claimed that this third channel accounts for the vast majority of IFFs from the countries analysed in its report. Most importantly, developing countries' increasing openness to trade in recent years, coupled with their weak governance, has created a favourable environment for IFFs through trade misinvoicing.

This study develops a novel empirical methodology to indirectly estimate the IFFs that take place through misinvoicing of international trade transactions. It employs the gravity model of international trade flows and publicly available product-level international trade statistics to estimate the transportation and insurance costs as well as the actual value of trade flows. The misinvoiced value is then estimated as the difference between the reported value and the estimated actual value of the transaction.

The proposed methodology makes three important contributions to the literature. First, it estimates product-level transportation and insurance costs and uses them to estimate the misinvoiced values in lieu of the usual rule of thumb margin of 10 per cent (Spanjers and Salomon 2017). Second, the actual value of trade flows is estimated empirically, while the extant literature makes the arbitrary assumption that developed countries always report the actual value of trade flows. Third, this enables the identification of misinvoicing for all countries and for import and export flows separately, which is an invaluable feature for policy makers.

[1] See Table 1.1 of Cobham and Janský (2020) for a typology of IFFs.

To illustrate this methodology this study employs publicly available data from the UN Comtrade system. These data cover the years 2013-16 and comprise bilateral costs for insurance and freight (CIF) imports and free on board (FOB) exports among 173 countries and 5,268 six-digit HS2012 commodities. The estimated export and import misinvoicing values suggest that the trade misinvoicing phenomenon is present across all types of products and across both developed and developing countries. Most importantly, the direction of the misinvoicing seems to depend upon the economic level of development of the trade partners.

A major limitation of the methodology proposed in this study is that a non-zero estimated misinvoicing value cannot be considered evidence of an actual and deliberate behaviour of misinvoicing a trade transaction. Indeed, trade data discrepancies may arise for reasons other than misinvoicing (Cobham and Janský 2020). Nevertheless, the estimated misinvoicing values are useful to researchers because these figures can shed some light on which countries, products, and trade flow direction seem to be more affected by this phenomenon. For policy makers and government officials, an unreasonably large estimated misinvoicing value may be a good indicator of where to concentrate resources for a profound investigation of the unusual trade values.

The remainder of this study is organized as follows. Section 2 describes the publicly available trade data used to illustrate the methodology proposed here. Section 3 develops the novel methodology designed to estimate the export and import misinvoicing values. Section 4 presents and discusses the results obtained. Finally, Section 5 contains the conclusions.

2 Data description

The data used to illustrate the methodology developed in this study are from Comtrade (2021).[2] Countries and some territories report their imports and exports to the UN Statistical Division using the Harmonized System (HS) classification. The imports are reported as CIF values, which means that the reported import figures contain the value of the goods plus the respective insurance and freight costs. The exports are generally reported as FOB values and therefore do not include insurance and freight costs. The UN Statistics Division then compiles the reported data and make them available to the public for download via Comtrade (the Commodities Trade Statistics database).

The analysis covers the years 2013–16 and the international trade flow data are extracted at the six-digit (also known as sub-headings) of the HS 2012 classification. A product is defined as a six-digit code.[3] It was decided not to use the most recent data for three reasons. First, the least developed countries typically take longer to report trade data and using older data therefore ensures coverage of the trade flows of almost all countries. Second, the data coverage begins one year after the HS 2012 classification came into effect and ends one year before a new classification (HS 2017) took place. Hence the issues arising from the creation and merging of product codes due to the use of different HS classifications are significantly attenuated in this timeframe.[4] Third, this period does not feature the creation, merging, and extinction of countries.

---

[2] According to Gaulier and Zignago (2010), the Comtrade database accounts for at least 95 per cent of world trade.

[3] Note that countries can adopt an eight-, ten-, or even twelve-digit product classification. Nonetheless, only the first six digits are harmonized across countries.

[4] Although the majority of trade flows in this period were reported using HS 2012 classification, a few countries still used older HS classifications, particularly in 2013 and 2014.

To be able to use the econometric estimation, the extracted data needed to undergo the following cleaning procedure. The records in which the reporter was '490 Other Asia, nes' were removed.[5] The total number of remaining reporters was 173. The records using the partners reported in Table 1 were also removed. The number of remaining partners is 232. These observations were removed because there is no observed distance available between them and their trade partners. Panel a. of Table 2 presents the descriptive statistics of the extracted raw data as made available by Comtrade. The trade volume shows considerable dispersion with a coefficient of variation always above 50. Panel b. shows the descriptive statistics after the cleaning process outlined above. The number of observations was reduced by approximately 10 per cent. Also, the removed observations had a large magnitude, as the maximum, standard deviation, and mean statistics declined considerably after their removal.

Another relevant characteristic of Comtrade data is that there are only a few reported zero-trade flows. Panel c. exhibits the number of reported zero-trade flows available in the raw data according to type of trade flow and year. We can see that only very few zero-trade observations are actually reported; indeed, less than 0.10 per cent of observations are of this type. Interestingly, zero-trade flows are more commonly reported for exports than for imports. The absence of observed trade data may be explained by either legitimate zero-trade flows or missing data. The following calculation provides the maximum number of observations per year and the flow direction if all trade flows (including all zero flows) were reported. This figure is 173 x (232-1) x 5,268 (six-digit HS12 product codes) = 210,525,084 observations per year. This means that only approximately 4 per cent of the theoretical maximum number of observations per year are actually available in the Comtrade database. This study assumes that the non-reported trade flows consist of zero trade.

The second dataset used in this study contains geographical variables for each of the county-pairs and comes from CEPII (2021). The dyadic variables of interest are the distance between capitals and indicators for common colonizer, common official language, and for contiguity. Table 3 displays the descriptive statistics for these variables. The average distance between countries' capital cities is approximately 8,400 km. Almost 1.2 per cent of country-pairs are made of contiguous countries and 17.3 per cent of country-pairs have a common official language, while 12 per cent have a common colonizer. This study now turns to description of the methodology.

3 Methodology

According to Kellenberg and Levinson (2019) and Nitsch (2017), there are myriad reasons why agents perpetrate international trade misinvoicing. For instance, export under-invoicing can be used to bypass export restrictions or to illegally remit funds abroad as importers can deposit the price difference in the exporters' bank account located in a tax haven country. Such trade misinvoicing schemes have been investigated since the 1950s, for example in the seminal works of Morgenstern (1950) and Bhagwati (1964).

Bhagwati (1964) developed an empirical methodology to assess trade-misinvoicing-based IFFs, in which the trade statistics of some countries (typically developed countries) are deemed to be free of misinvoicing. They are therefore reliable in the sense that they reflect the true value of the shipment. The misinvoicing experienced by the unreliable country is then estimated by contrasting the declared export and import values between the reliable and the unreliable country. As the declared exports are FOB values and the declared imports are CIF values, the insurance and freight

[5] 'Nes' means 'not elsewhere specified'.

cost adjustments are usually assumed to be 10 per cent of the FOB value for all products (Cobham and Janský 2020). The remaining discrepancy after the deduction of insurance and freight costs is considered to be the misinvoiced value of the trade flow.

The evidence uncovered by de Boyrie et al. (2005) suggests misinvoicing in the trade flows between the USA and Switzerland. This casts serious doubt on developed countries' status as reliable partner countries. Also, bilateral collusion—likely to happen in transfer pricing—is difficult to detect using this method (Yalta and Demir 2010). Moreover, the ad hoc 10 per cent margin for transport and insurance costs is at odds with the finding of Ndikumana (2016) of substantial product-level heterogeneity in such costs. One possibility is the use of actual freight and insurance costs together with the cost of holding inventory during shipment. Unfortunately, data on transport costs are generally rarely found (Limão and Venables 2001) and are even more scarce at highly disaggregated levels, such as at the six-digit HS level. One notable exception is Hummels and Skiba (2004) whose data cover only the USA and the Southern Cone countries. Nonetheless, little is known for other countries. This paucity of data is the central argument for using estimated trade costs.

The methodology presented in this study addresses these issues and comprises three steps. In the first step it uses the gravity model of international trade to estimate product-level transportation and insurance costs. In the second step it estimates the actual value of trade flows based upon fitted values of the gravity model and the estimated transportation and insurance cost. This replaces the problematic and highly controversial reliable-country assumption used in the literature. In the third step the actual value of the shipment estimates is then used to separately estimate export and import misinvoiced values for all countries. This is important because it addresses the bilateral collusion issue, and these two types of misinvoicing when taken together can cancel each other out and severely bias misinvoicing estimates (Hong et al. 2017).

Let the reported FOB value of an export flow of product *p* from country *i* to country *j* at time *t* (FOB_export_{ijpt}) consist of the actual value of the export flow ($exportvalue_{ijpt}$) plus any misinvoicing performed by the exporters ($export\_misinvoicing_{ijpt}$). Similarly, the reported CIF value of an import flow of product *p* of country *j* from country *i* at time *t* (CIF_import_{ijpt}) consists of the actual value of the import flow ($importvalue_{ijpt}$) plus transportation and insurance costs ($transportation_{ijpt}$) plus any misinvoicing performed by the importers ($import\_misinvoicing_{ijpt}$). Note that only the reported FOB and CIF values are observed by the researcher.

The first identification assumption is that both export and import misinvoicing are determined by the quadruple product, importer, exporter, and time. Product matters for the extent of misinvoicing because the larger its price dispersion, the less detectable is the mispricing, for instance. The source and destination countries matter due to their specific natural and artificial trade barriers, tax systems and rates, legal systems, currencies, trade agreements, and tastes, etc. This implies that for a given product–source–destination–time quadruple, any profit- (or welfare) maximizing agent will exhibit the same optimal behaviour of over-, under-, or no misinvoicing. This reasonable assumption allows for the decomposition of misinvoicing between export and import operations.

The first step of the methodology is to estimate a gravity equation for each product and trade flow (FOB exports and CIF imports) to recover product-level freight and insurance costs, as specified in equation (1).

$$Reportedflow_{ijpt} = exp\left[\alpha_{it} + \alpha_{jt} + \delta_p \log\left(Distance_{ij}\right) + \boldsymbol{x}_{ijt}\boldsymbol{\beta}\right]\varepsilon_{ijpt} \qquad (1)$$

where the dependent variable ($Reportedflow_{ijpt}$) is either FOB_export_{ijpt} or CIF_import_{ijpt} and includes zero-trade flows; α_{it} is the exporter-year fixed effects; α_{jt} is the importer-year fixed effects; $Distance_{ij}$ is the distance between importer and exporter; $\boldsymbol{x}_{ijt}$ is a vector of other dyadic variables like common language, contiguity, common colonizer, regional trade agreement membership, etc; and ε_{ijpt} is the error term, which is assumed to be independent of the explanatory variables. The distance coefficient (δ_p) is expected to have a negative sign, whereas the coefficents for common language, contiguity, and common colonizer are expected to be positive.

The exporter-year (α_{it}) and importer-year (α_{jt}) fixed effects capture country-level variables which affect trade flows, such as economic size, area, population, and remoteness, etc. The second identification assumption is that the gravity equation, equation (1), is properly specified and its explanatory variables are able to capture all the factors that lead to legitimate price dispersion across trade flows due to differences in quality and mode of transportation, etc.

As discussed in Head and Mayer (2014) and in Gervais (2019), a dyadic variable like distance between the countries captures not only the effects of transportation and insurance costs on trade flows but also the impacts of import tariffs, non-tariff barriers, and differences in tastes and in production costs. Note that the distance coefficient estimated using FOB data captures all factors affecting trade flows other than transportation and insurance costs, whereas the distance coefficient estimated using CIF data captures these other factors plus transportation and insurance costs. Thus, we can estimate product-level transportation and insurance costs through the difference between the estimated coefficients of the distance of the CIF and the FOB specifications ($\widehat{\delta_p^{CIF}} - \widehat{\delta_p^{FOB}}$).

Although gravity models are usually estimated using a double-log specification, Santos Silva and Tenreyro (2006, 2010) stressed two serious econometric issues about the log-linear specification of the gravity model of equation (1). First, log-linearized models cannot deal with zero-trade observations as the natural logarithm of zero is not defined. These observations are usually dropped and not used in the estimation. The zero-trade observations are not random because likelihood depends on the explanatory variables. For instance, larger economies are less likely to exhibit zero-trade flows. Therefore, the omission of these zero-trade observations leads to biased estimates.

The second issue is the presence of heteroskedasticity as the expected value of the logarithm of the error ($\log \varepsilon_{ijpt}$) depends on the variance and other higher moments of ε_{ijpt}. As Santos Silva and Tenreyro (2006) pointed out, it is very likely that a bilateral trade flow model exhibits a heteroskedastic error term. This means that in a log-linear specification, the regressors will be correlated with the error term, and this makes the estimates of both coefficients and standard errors biased and inconsistent.[6]

To address these two issues Santos Silva and Tenreyro (2006, 2010) proposed estimating the gravity model using the Poisson Pseudo-Maximum Likelihood (PPML) estimator. This estimator can deal with zero values of the dependent variables and provides consistent parameter estimates even with heteroskedastic errors, as long as the gravitational model is correctly specified. An additional advantage of the PPML estimator is that it does not suffer from the incidental parameter problem (see Fernández-Val and Weidner 2016), which is an important concern whenever fixed

[6] As coefficient estimates are also biased, this type of heteroskedasticity cannot be addressed simply by using a robust covariance matrix estimator.

effects are used. In other words, the PPML allows for 'differencing out' of the fixed effects, and this improves the computational speed and accuracy. A shortcoming of the PPML estimator is that its numerical implementation sometimes does not converge. This problem also affects several non-linear estimators and can be attenuated by rescaling and centering variables. As it is not clear whether the assumption about the conditional variance actually holds, to account fully for heteroskedasticity, it is recommended to make inferences based on the Eicker-White robust covariance matrix estimator. In view of all these arguments, equation (1) is estimated using the PPML estimator following the recommendations of Santos Silva and Tenreyro (2006) in order to obtain consistent estimates of the parameters.

One additional criticism of the econometric model based on equation (1) comes from Egger (2005), Baier and Bergstrand (2007), and Magee (2008). Their argument is that the country-pair characteristics—like similar legal systems—are likely to affect bilateral trade flows and to bias estimates if left unaccounted for. This means that estimates based on equation (1) will be biased due to omitted variables. One way to address this problem is to add country-pair fixed effects to equation (1) to obtain more robust estimates. The country-pair fixed effects control for all observed and unobserved country-pair, time-invariant characteristics that influence trade flows, such as distance between countries, adjacency, colonial ties, and common language indicators. They also account for country-pair unobservable effects such as natural proclivity to be trade partners, cultural and institutional aspects, and all other unmeasurable characteristics which affect trade flows. Nevertheless, the coefficient of the log of the distance between countries will not be identified in this case. One solution is to first estimate equation (1) using the country-pair fixed effects. This provides consistent estimates of the country-year fixed effects. These country-year effects are then subtracted from the dependent variable. This modified dependent variable is now regressed on distance between countries, adjacency, colonial ties, and common language indicators as before.

Although the PPML estimator is able to handle zero-trade observations, not all zero-trade observations will be used in the computation of the estimates. Indeed, the identification source for the estimated coefficients is the countries that have at least one non-zero trade. The significance of this is that it is only possible to obtain predicted values of trade flows for country-pairs in which both countries have at least one non-zero-trade flow for the product of interest. This can be problematic for product-level analysis because the more disaggregated the unit of analysis, the larger is the share of zero-trade observations. As a result, only observations with predicted trade values will be used in the remainder of the analysis.

In the second step of the methodology, the $exportvalue_{ijpt}$ is estimated as the predicted value of the shipment paid by the importer less the transportation and insurance costs, as shown in equation (2). Similarly, the $importvalue_{ijpt}$ is estimated as the predicted value charged by exporters plus the estimated transportation and insurance costs, as depicted in equation (3).[7] These estimates will be free from misinvoicing as long as the *export_misinvoicing*$_{ijpt}$ and *import_misinvoicing*$_{ijpt}$ are independent. This is the third identification assumption.

$$\widehat{exportvalue}_{ijpt} = CIF\widehat{import}_{ijpt} exp[(\widehat{\delta_p}^{FOB} - \widehat{\delta_p}^{CIF})\log(Distance_{ij})] \quad (2)$$

$$\widehat{importvalue}_{ijpt} = FOB\widehat{export}_{ijpt} exp[(\widehat{\delta_p}^{CIF} - \widehat{\delta_p}^{FOB})\log(Distance_{ij})] \quad (3)$$

[7] Note that this framework is robust to special cases like the USA which calculate import duties on FOB values while other countries do so on CIF values.

In the third and final step of the methodology, the estimated export misinvoicing ($export\widehat{misinvoicing}_{it}$) and the estimated import misinvoiced ($import\widehat{misinvoicing}_{ijpt}$) consist of the difference between reported trade flow values and estimated trade flow values, as shown in equations (4) and (5). A nice feature of the methodology proposed here is that under- and over-invoicing for both imports and exports can be separately identified. Indeed, a negative value of $export\widehat{misinvoicing}_{ijpt}$ means an under-invoicing of exports. Conversely, a positive value suggests over-invoicing of exports.

$$export\widehat{misinvoicing}_{ijpt} = FOBexport_{ijpt} - export\widehat{value}_{jipt} \tag{4}$$

$$import\widehat{misinvoicing}_{jipt} = CIFimport_{ijpt} - import\widehat{value}_{jipt} \tag{5}$$

The residual nature of the estimated misinvoicing means that this measure may not necessarily capture misinvoicing only. In addition to idiosyncratic shocks, this residual can contain legitimate price heterogeneity—due to quality or transportation heterogeneity, for instance—which is not captured by the gravity equation explanatory variables, product misclassification, or even due to trade data discrepancies arising for reasons other than misinvoicing (see Cobham and Janský (2020) for several examples). If the third identification assumption holds, these concerns will be attenuated to some extent. Nonetheless, to be on the safer side, the estimated misinvoicing should be interpreted as an upper bound on the actual misinvoicing. If it is believed that this legitimate price heterogeneity is determined at the country-pair level, one remedy would be to use the predictions generated through the gravity model with country-pair fixed effects in equations (2) and (3). Another possible remedy would be to subtract the average residual at the country-pair level from the estimated misinvoicing. This is ultimately an empirical matter which also depends on the time period covered and should be further explored in future work.

A few words of caution are in order regarding the methodology proposed here. These misinvoicing estimates are based on an average of transactions at the product–exporter–importer–year quadruple, and therefore some transactions may be over-invoiced while others may be under-invoiced, that is, the first identification assumption fails to hold.[8] Nonetheless, this is a very unlikely scenario because large variations in unit value will easily attract the attention of customs officials. Second, note that the gravity model is estimated using potentially misinvoiced transactions. Suppose that a certain product is commonly used for trade misinvoicing operations. In this case firms will generally report artificially high CIF prices to remit money away from countries which import this type of good. As a result, the difference between CIF prices and FOB prices will be very large, and the transportation and insurance costs will be over-estimated. Moreover, the precision of the estimates of transportation and insurance costs will be smaller for products that are not traded by many partners. Therefore, the fewer the non-zero-trade observations, the larger will be the bias caused by misinvoiced transactions on the estimated transportation and insurance costs. One way to assess the sensitivity of our estimates to these two issues is to conduct the estimates at the four-digit product level, as will be done in the robustness subsection.

The estimated country-level total export ($export\widehat{misinvoicing}_{it}$) and import misinvoiced ($import\widehat{misinvoicing}_{ijpt}$) are the sum across trade partners and products of the absolute value

[8] On the one hand, the use of higher frequency data, such as quarterly data, would attenuate this problem. On the other hand, it would introduce systematic measurement error due to different transit times, in which some shipments take more than 20 days.

of the difference between reported values and estimated trade flow values, as shown in equations (6) and (7).

$$\widehat{exportmisinvoicing}_{it} = \sum_{j \neq i} \sum_{p} \left(\left| FOBexport_{ijpt} - \widehat{exportvalue}_{jipt} \right| \right) \quad (6)$$

$$\widehat{importmisinvoicing}_{jt} = \sum_{i \neq j} \sum_{p} \left(\left| CIFimport_{ijpt} - \widehat{importvalue}_{jipt} \right| \right) \quad (7)$$

The summation of the absolute value of the misinvoicing at the product level avoids the problem of over-invoiced trade flows cancelling out under-invoiced flows. This would lead to a misrepresentation of the actual volume of misinvoicing. The results obtained using this methodology and the 2013–16 sample are presented in the next section.

4 Results

The econometric model based on equation (1) is estimated for each product using six-digit HS product-level export and import data. This is a computationally intensive procedure. The descriptive statistics of the estimated coefficients are reported in Table 4. Panel a. displays the descriptive statistics for the coefficients obtained using imports as the dependent variable, while those in Panel b. were obtained using exports as the dependent variable. The means of the estimated coefficients have the expected signs—negative for the distance and positive for the others. In terms of magnitude the means of the coefficients are in line with gravity equation estimates in the extant literature (Gervais 2019). Nevertheless, there were several cases in which the estimated coefficients held the opposite signs and their magnitudes varied widely. With the exception of common colonizer, the other variables' coefficients exhibit a larger magnitude in Panel b. The number of observations is different across panels. This is so because there were 5,266 imported products for which the estimation of equation (1) was feasible but only 5,251 exported products for which the estimation of equation (1) was feasible. Of these products only 5,228 had estimates for both imports and exports and these are the products that will be utilized in the remainder of the analysis.

Table 5 gives more-detailed descriptive statistics of these 5,228 estimated coefficients for the distance ($\widehat{\delta^{CIF}}$ and $\widehat{\delta^{FOB}}$). Both coefficients exhibit negative skewness, although $\widehat{\delta^{CIF}}$ has a more pronounced skewness. Approximately 90 per cent of the coefficients have the expected negative sign. There are also several outliers which are both positive and negative. Figure 1 shows the kernel density of both coefficients. The difference between the CIF and the FOB coefficients has a positive mean. Actually, it is positive in almost 75 per cent of the cases. This is confirmed in Figure 2, which shows the kernel density of the difference between the two estimated coefficients. The finding that approximately a quarter of the differences had a negative sign is an unexpected result. Unfortunately, there is no benchmark in the literature with which to compare our results. Indeed, there is a paucity of studies which estimate the gravity equation at the product level.

The next step is to use the estimated coefficients of equation (1) to predict the trade flows and then use the distance and its FOB and CIF estimated coefficients to calculate the estimated actual export and import values according to equations (2) and (3), respectively. Then the estimated import and the estimated export misinvoicings are calculated according to equations (4) and (5), respectively. This exercise is conducted using only the observations for which a predicted trade flow can be calculated.

The descriptive statistics of the actual export flows, the estimated export values, and the export misinvoicing estimates are reported at the year level in Table 6. The figures reported present several patterns that are present in every year of the sample and merit discussion. The first pattern which emerges from this table is that the means of the estimated export values are greater than the means of the actual exports. A second pattern is that the maximum values of the estimated export values are substantially larger than the maximum of the actual exports. This suggests that some estimated export values may not be good estimates. This is corroborated by the fact that the estimated export value has a standard deviation which is approximately 75 per cent larger than that of the actual exports. Perhaps this larger variability is because of extreme values of the estimated coefficients for the distance. Third, the means of the export misinvoicing are positive—which means over-invoicing—and approximately 10 per cent of the means of the actual exports. Fourth, the estimated export misinvoicing can be either positive or negative and presents large values relative to the actual exports as a result of being either zero or a very large magnitude of estimated exported values.

Table 7 displays the descriptive statistics at the year level of the actual import flows, the estimated import values, and the estimated import misinvoicing. The figures in Table 7 exhibit similar patterns relative to those in Table 6, albeit with some differences in terms of magnitude. The means of the estimated import values are smaller than the means of the actual imports. Although the maximum values of the estimated import values are larger than the maximum of the actual imports, they are so by a much smaller factor than that exhibited by the export data in Table 6. The standard deviations of the actual imports and estimated import values are comparable. The means of the estimated import misinvoicing are positive and their magnitudes are approximately 7 per cent of the mean of the actual imports, which is smaller than the 10 per cent found for the export misinvoicing. The import misinvoicing also shows substantial variability and displays both positive and negative values. The concern that aggregation leads to over-invoicing offsetting under-invoicing is corroborated by the results found in this study.

Tables 8 reports the descriptive statistics of the actual exports and the estimated export values broken down between high-income countries (OECD countries except for Chile, Mexico, and Turkey) and developing countries (the rest of the world). Panel a. shows the statistics for high-income countries. We can see that the means of the estimated export values are smaller than the means of actual exports for OECD countries. The standard deviation of estimated export values is approximately 50 per cent larger than the standard deviation of the actual exports. The means of the estimated export misinvoicing are positive and are approximately 20 per cent of the actual exports. The figures for developing countries, which are displayed in Panel b., are at variance with those from Panel a. The means of the estimated export values are larger than those of actual exports. The standard deviation of the estimated export value is about two times that of the actual exports. The mean of the estimated export misinvoicing is negative and the magnitude is between half and 5 per cent of the actual exports.

Figure 3 shows the kernel density of the estimated export misinvoicing as a share of actual exports for exports originating in high-income (black line) and developing countries (grey line). The two densities are very different in shape and in skewness. The significance of this is that both the intensity and type of misinvoicing (under- or over-misinvoicing) vary according to the source of the export flow. The kernel density displayed in Figure 4 is a similar exercise but now according to the economic development level of the destination (partner) country. The densities are now more similar, albeit that export over-invoicing seems more frequent when the destination country is a developing country. These figures and the statistics from Table 8 reveal substantial heterogeneity of export misinvoicing depending upon the economic development level of countries. Legitimate price heterogeneity due to product quality could be one of the explanations for this pattern if richer countries tend to export higher-quality products.

The analysis now turns to import misinvoicing. Table 9 presents the descriptive statistics of the actual imports and the estimated import values broken down between high-income countries and developing countries. Panel a. shows the statistics for high-income countries. The means of the actual imports are greater than the means of the estimated import values. The standard deviations of these two variables are comparable. The means of the estimated import misinvoicing have a small magnitude (below 1 per cent of the mean of the actual imports) and show mixed signs. This does not mean that there is no misinvoicing. Indeed, import misinvoicing shows a standard deviation of approximately 70 per cent of the standard deviation of the actual imports. Moving to the statistics for the developing countries in Panel b., we can see that the means of the actual imports are larger than the means of the estimated import values. Their standard deviations are of similar magnitude. The means of the estimated import misinvoicing are all positive and between 15 and 20 per cent of the mean of the actual imports. As in Panel a., estimated import misinvoicing exhibits a large standard deviation and a large range encompassing both positive and negative values. Interestingly, this pattern is at odds with the hypothesis that taste for quality is increasing in the income per capita, for instance.

Figure 5 displays the kernel density of the estimated import misinvoicing as a share of actual imports for imports heading to high-income (black line) and developing countries (grey line). The two densities seem similar except for the extreme right-hand side of the diagram, where import over-invoicing is more frequent in developing countries. Figure 6 shows the density of the estimated import misinvoicing as a share of actual imports according to the country of origin of the import flows. The density for the high-income countries has a very different shape and skewness relative to that of developing countries. More precisely, the former has a substantial mass in the negative values of import misinvoicing, while most of the mass of the latter is located in the range of positive values of misinvoicing. The significance of this is that imports originating in high-income countries are more likely to be under-invoiced, while those from developing countries tend to be over-invoiced. In summary, the statistics in Table 9 and Figures 5 and 6 reveal considerable heterogeneity in import misinvoicing depending on the countries' economic development level.

Table 10 gives the top five products in terms of estimated export over-invoicing (Panel a.) and under-invoicing (Panel b.). This ranking is populated by agricultural products such as bambara beans, minimally processed mineral products (unwrought tungsten), and manufactured products such as computer memories. As these products have a different scope for quality differentiation, this means that the estimated misinvoicing is not just capturing legitimate price heterogeneity due to product differentiation. Table 11 shows the top products according to estimated import over-invoicing (Panel a.) and under-invoicing (Panel b.). As in Table 10 this ranking is populated by products of different natures. Figure 7 presents the kernel density of import misinvoicing aggregated at the product level as a share of actual imports. Interestingly, under- and over-invoicing of imports seems equally likely. This result highlights the concern that under- and over-invoicing may cancel each other out depending on how the data are aggregated. Figure 8 shows the kernel density of export misinvoicing aggregated at the product level as a share of actual exports. This density suggests that export over-invoicing is more frequent than under-invoicing. Taken together these results point to a substantial heterogeneity of misinvoicing across products and across trade flows.

One important assumption of the methodology developed here relates to the direction of the misinvoicing being the same for the quadruple year–exporter–importer–product. Although it is not possible to assess the validity of this assumption, it would be advisable to see whether some countries present more misinvoicing in one direction or in the other. Table 12 shows the share of transactions that were estimated to be over-misinvoiced by trade flow according to economic development. The share of under-invoiced transactions is just the difference between 100 per cent and the share of over-invoiced transactions. We can see that the most frequent misinvoicing for developing countries' imports is over-invoicing, while for developed countries it is under-invoicing. When it comes to exports, developing countries feature a larger share of under-invoiced transactions, whereas developed countries exhibit the opposite pattern.

4.1 Robustness check

The robustness check conducted here consists in assessing the sensitivity of the estimates obtained using the proposed methodology with four-digit HS-level data instead of the six-digit data used before. At this aggregation level there are 1,228 products instead of the 5,268 available at the six-digit level. Using this aggregated data would address the misinvoicing that could take place via product misclassification (Carrère and Grigoriou 2014)—for instance ores with different concentration levels being misclassified or special steel alloys classified as standard steel. Second, this also alleviates the zero-trade observation issue. And third, this addresses the two concerns raised in the methodology section regarding the estimation bias introduced on the distance coefficient estimation due to misinvoicing and the lack of precision resulting from smaller sample sizes. These new estimates are summarized in Tables 13 and 14, which are equivalent to Tables 4 and 5 of the estimates using six-digit data, respectively. These figures exhibit less extreme maximum and minimum estimates of the transportation costs. We can also see a reduction in the dispersion of the estimated coefficients. Indeed, the figures in Table 14 suggest that the density of the estimated coefficients became more concentrated on a range of reasonable values. Figure 9 exhibits the kernel density of difference between the estimated CIF and FOB distance coefficients using four-digit HS-level data. In contrast to Figure 2, which used six-digit HS-level data, in addition to a decrease in dispersion, we can also see an increase in the mass of positive values.

5 Conclusions

The unprecedented surge in international trade in recent decades may have inadvertently facilitated IFFs. These illicit flows harm the tax base and foreign currency reserves of countries and may ultimately reduce their economic growth. Such negative effects have a disproportionate impact on developing countries, and this is why a reduction in IFFs became target 16.4 of the UNSDGs.

No consensus on the definition of IFFs has yet been achieved. Based on the extant literature IFFs are defined here as cross-border transfers of financial capital in contravention of national or international laws, regardless of the legality of the origin of these financial resources. Amongst the different ways these cross-border flows take place, misinvoicing of international trade flows has attracted the attention of researchers and policy makers for several decades. Given the illegal nature of IFFs, there is no direct way of estimating their magnitudes.

This study develops a new methodology to indirectly estimate the IFFs that take place via misinvoicing of international trade in goods. This methodology features the use of publicly available data on international trade statistics and addresses many of the flaws of the existing indirect methods. In contrast to the previous methods, the methodology developed here does not rely on the assumption that misinvoicing only takes place in a certain group of countries (e.g. developing countries), whereas other countries' trade statistics are reliable. Also, it does not employ ad hoc measures of shipping costs. To avoid such flaws this study employs a gravity equation to estimate shipping costs at the product level and uses these estimates to obtain an estimate of misinvoiced values.

The publicly available UN Comtrade international trade data for 2013–16 were used to illustrate the methodology. The estimated export and import misinvoicing values indicate that the trade misinvoicing phenomenon is present across both developed and developing countries. Also, an exporter (or importer) may over-invoice for some destinations (or source countries) and under-invoice for others. At the product level the results indicate that under-invoicing of imports is as frequent as over-invoicing of imports, while over-invoicing is much more frequent for exports. These results justify the concerns raised by many observers about the extant methodologies for estimating misinvoicing using mirror trade data.

Despite the advancements introduced by the proposed methodology, it is important to understand its limitations. Most importantly, part of the estimated misinvoicing could be due to factors other than IFFs such as product misclassification, or legitimate price heterogeneity due to quality differentiation. Also, the time lag between the recording of exports and imports due to transit time can lead to the same transaction being declared in different calendar years. This means that the results obtained via the methodology proposed here should be cautiously interpreted as an upper bound of misinvoicing. Still, they are useful for pointing to trade flows that merit careful scrutiny by authorities.

References

Addison, T., M. Niño-Zarazúa, and J. Pirttilä (2018). 'Fiscal Policy, State Building and Economic Development'. *Journal of International Development*, 30(2): 161–72. https://doi.org/10.1002/jid.3355

Baier, S. L., and J.H. Bergstrand (2007). 'Do Free Trade Agreements Actually Increase Members' International Trade?'. *Journal of International Economics*, 71(1): 72–95. https://doi.org/10.1016/j.jinteco.2006.02.005

Bhagwati, J.N. (1964). 'On the Underinvoicing of Imports'. *Bulletin of Oxford University Institute of Economics and Statistics*, 27(4): 389–97. https://doi.org/10.1111/j.1468-0084.1964.mp27004007.x

Carbonnier, G., and A.Z. de Cadena (2015). 'Commodity Trading and Illicit Financial Flows'. *International Development Policy*. Available at: http://journals.openedition.org/poldev/2054 (accessed 10 July 2021).

Carrère, C., and C. Grigoriou (2014). 'Can Mirror Data Help to Capture Informal International Trade?' Policy Issues in International Trade and Commodities Research Study Series 65. New York, NY: UNCTAD.

CEPII (2021). 'The CEPII Gravity Database'. Paris: CEPII. Available at: http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=8 (accessed: 7 February 2021).

Cobham, A., and P. Janský (2020). Estimating Illicit Financial Flows: A Critical Guide to the Data, Methodologies, and Findings. Oxford: Oxford University Press https://doi.org/10.1093/oso/9780198854418.001.0001

Comtrade (2021). 'The UN Comtrade Database'. New York, NY: United Nations. Available at: https://comtrade.un.org/ (accessed 21 February 2021).

de Boyrie, M.E., S.J. Pak, and J.S. Zdanowicz (2005). 'The Impact of Switzerland's Money Laundering Law on Capital Flows Through Abnormal Pricing in International Trade'. *Applied Financial Economics*, 15(4): 217–30. https://doi.org/10.1080/0960310042000313200

Egger, P. (2005). 'Alternative Techniques for Estimation of Cross-section Gravity Models'. *Review of International Economics*, 13(5): 881–91. https://doi.org/10.1111/j.1467-9396.2005.00542.x

Fernández-Val, I., and M. Weidner (2016). 'Individual and Time Effects in Nonlinear Panel Models with Large *N, T*'. *Journal of Econometrics*, 192(1): 291–312. https://doi.org/10.1016/j.jeconom.2015.12.014

Forstater, M. (2018). 'Illicit Financial Flows, Trade Misinvoicing, and Multinational Tax Avoidance: The Same or Different?'. CGD Policy Paper. Washington, DC: Center for Global Development.

Available at: https://www.cgdev.org/publication/illicit-financial-flows-trade-misinvoicing-and-multinational-taxavoidance (accessed 10 July 2021).

Gaulier, G., and S. Zignago (2010). 'BACI: International Trade Database at the Product-Level (The 1994-2007 Version)'. CEPII Working Paper, 2010-23, October. Paris: CEPII. https://doi.org/10.2139/ssrn.1994500

Gervais, A. (2019). 'Indirect Measures of Trade Costs: Limitations and Caveats'. *Economics Letters*, 176: 96–102. https://doi.org/10.1016/j.econlet.2018.12.027

Head, K., and T. Mayer (2014). 'Gravity Equations: Workhorse, Toolkit, and Cookbook'. *Handbook of International Economics*, 4: 131–95. https://doi.org/10.1016/B978-0-444-54314-1.00003-3

Hong, K.P., and S.J. Pak (2017). 'Estimating Trade Misinvoicing from Bilateral Trade Statistics: The Devil is in the Details'. *The International Trade Journal*, 31(1): 3–28. https://doi.org/10.1080/08853908.2016.1202160

Hummels, D., and A. Skiba (2004). 'Shipping the Good Apples Out? An Empirical Confirmation of the Alchian-Allen Conjecture' *Journal of Political Economy*, 112(6): 1384–402. https://doi.org/10.1086/422562

Kar, D., and J. Spanjers (2014). 'Illicit Financial Flows from Developing Countries: 2003–2012'. Global Financial Integrity [online]. Available at: https://gfintegrity.org/report/2014-global-report-illicit-financial-flows-from-developing-countries-2003-2012/ (accessed 10 July 2021).

Kellenberg, D., and A. Levinson (2019) 'Misreporting Trade: Tariff Evasion, Corruption, and Auditing Standards'. *Review of International Economics*, 27: 106–29. https://doi.org/10.1111/roie.12363

Limão, N., and A.J. Venables (2001). 'Infrastructure, Geographical Disadvantage, Transport Costs, and Trade'. *The World Bank Economic Review*, 15(3): 451–79. https://doi.org/10.1093/wber/15.3.451

Magee, C. (2008). 'New Measures of Trade Creation and Trade Diversion' *Journal of International Economics*, 75(2): 349–62.

Morgenstern, O. (1950). *On the Accuracy of Economic Observations*. Princeton, NJ: Princeton University Press.

Ndikumana, L. (2016). *Trade Misinvoicing in Primary Commodities in Developing Countries: The Cases of Chile, Cote d'Ivoire, Nigeria, South Africa and Zambia*. Geneva: UNCTAD.

Nitsch, V. (2017). 'Trade Misinvoicing in Developing Countries'. CGD Policy Paper 103. Washington, DC: Center for Global Development.

OECD (2014). *Illicit Financial Flows from Developing Countries: Measuring OECD Responses*. Paris: Organization for Economic Cooperation and Development. https://doi.org/10.1787/9789264203501-en

Santos Silva, J.M.C., and S. Tenreyro (2006). 'The Log of Gravity'. *Review of Economics and Statistics*, 88(4): 641–58. https://doi.org/10.1162/rest.88.4.641

Santos Silva, J.M.C., and S. Tenreyro (2010). 'On the Existence of the Maximum Likelihood Estimates in Poisson Regression'. *Economics Letters*, 107(2): 310–12. https://doi.org/10.1016/j.econlet.2010.02.020

Spanjers, J., and M. Salomon (2017). *Illicit Financial Flows to and from Developing Countries: 2005–2014*. Washington DC, USA: Global Financial Integrity.

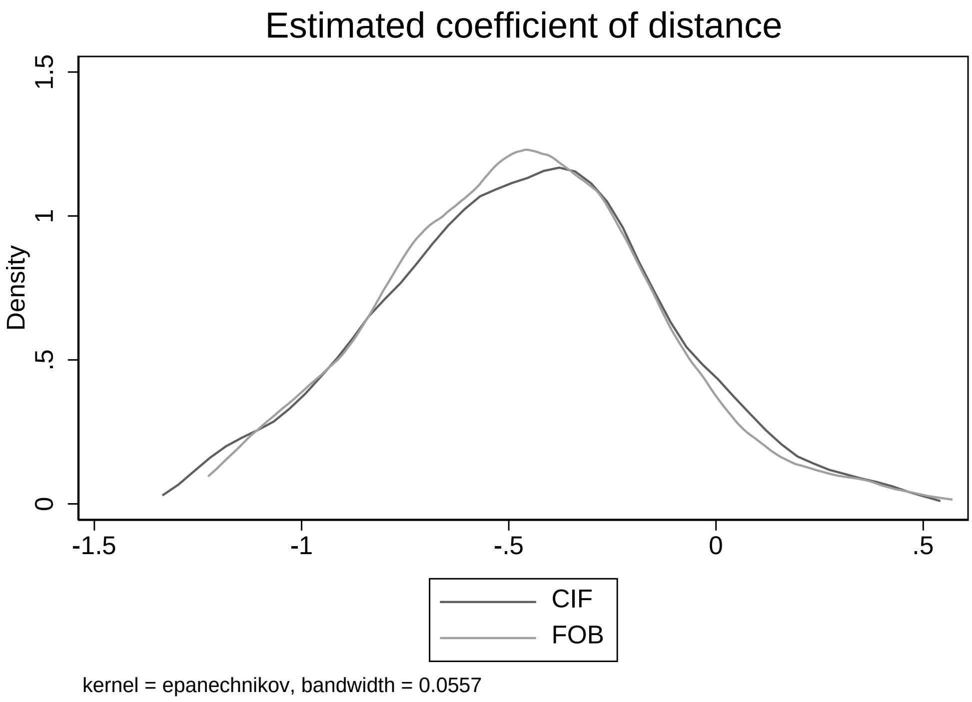
United Nations (2016). *Coherent Policies for Combating Illicit Financial Flows* (in association with the OECD). Inter-agency Task Force on Financing for Development Issue Brief. Available at: https://www.un.org/esa/ffd/wp-content/uploads/2016/01/Coherent-policies-for-combatting-Illicit-Financial-Flows_UNODC-OECD_IATF-Issue-Brief.pdf (accessed 10 July 2021).

World Bank (2016). 'The World Bank Group's Response to Illicit Financial Flows: A Stocktaking'. Board Report 104568. Washington, DC: The World Bank. Available at: http://documents.worldbank.org/curated/en/502341468179035132/pdf/104568-BR-SecM2016-0112-IDASecM2016-0071-IFC-SecM2016-00423-MIGA-SecM2016-0044-Box394878B-PUBLIC-disclosed-4-5-16.pdf (accessed 10 July 2021).

Yalta, A.Y., and I. Demir (2010). 'The Extent of Trade Mis-Invoicing in Turkey: Did Post-1990 Policies Matter?'. *Journal of Economic Cooperation and Development*, 31(3): 41–66.

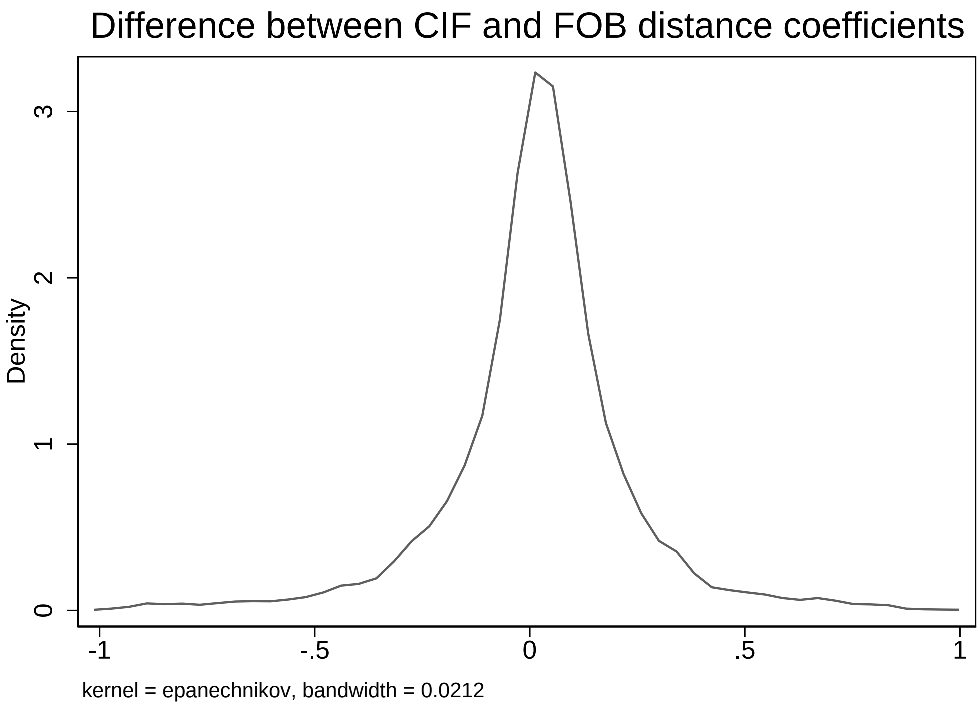
Figures and tables

Figure 1: Kernel density of the CIF and the FOB distance estimated coefficients using six-digit HS-level data



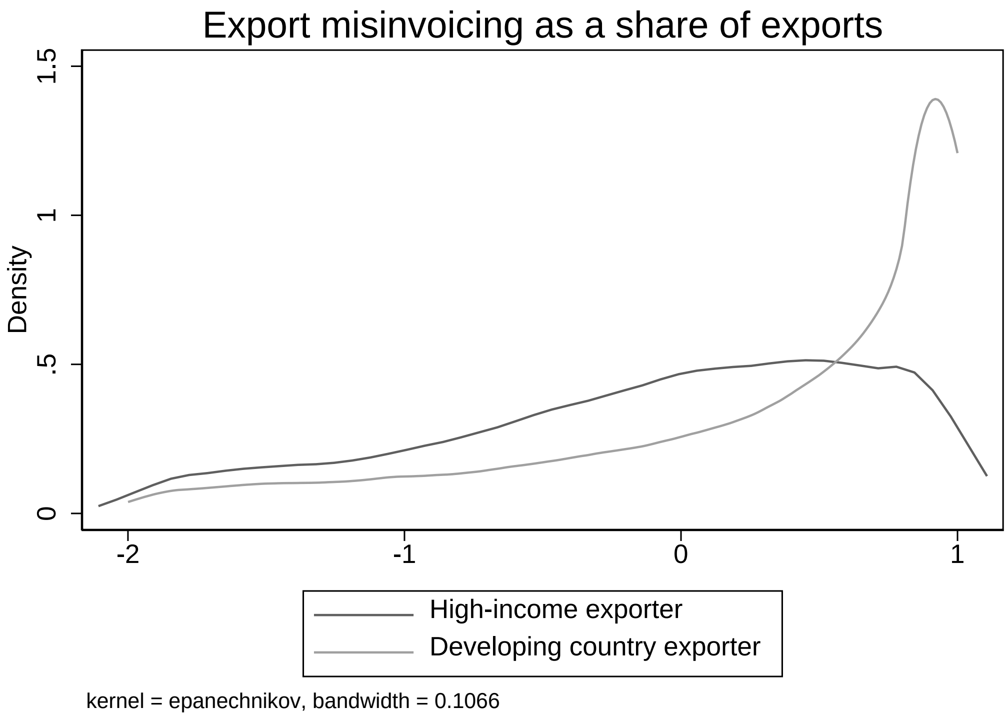
Source: author's calculations based on UN Comtrade data and CEPII gravity data.

Figure 2: Kernel density of the difference between the estimated CIF and FOB distance coefficients using using six-digit HS-level data



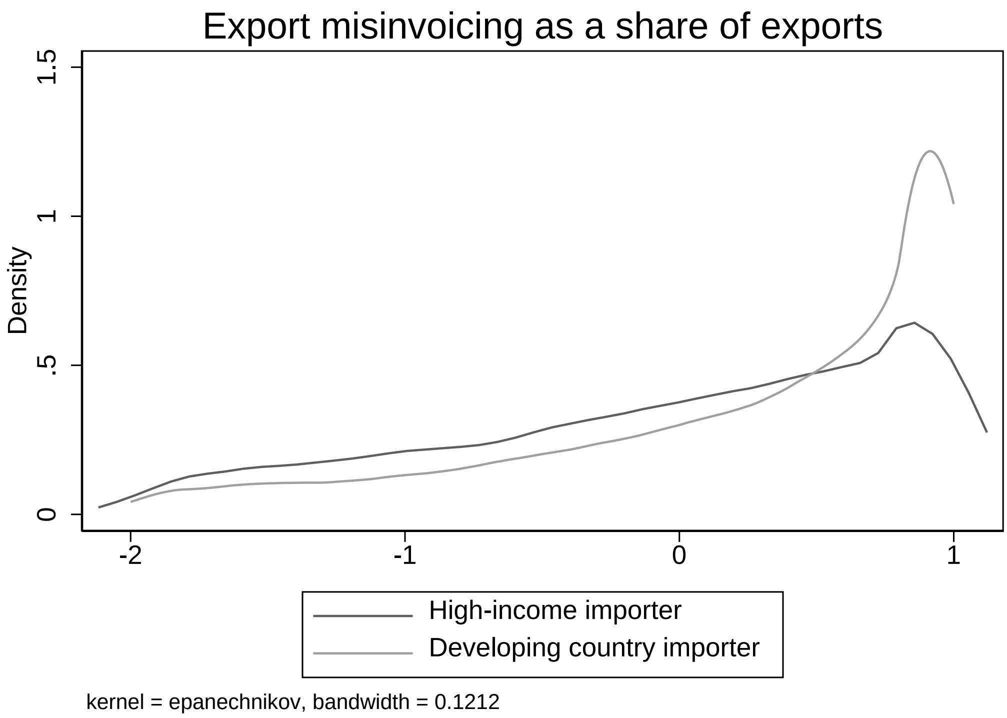
Source: author's calculations based on UN Comtrade data and CEPII gravity data.

Figure 3: Kernel density of export misinvoicing as a share of actual exports according to exporter (reporter) country economic development level using six-digit HS-level data



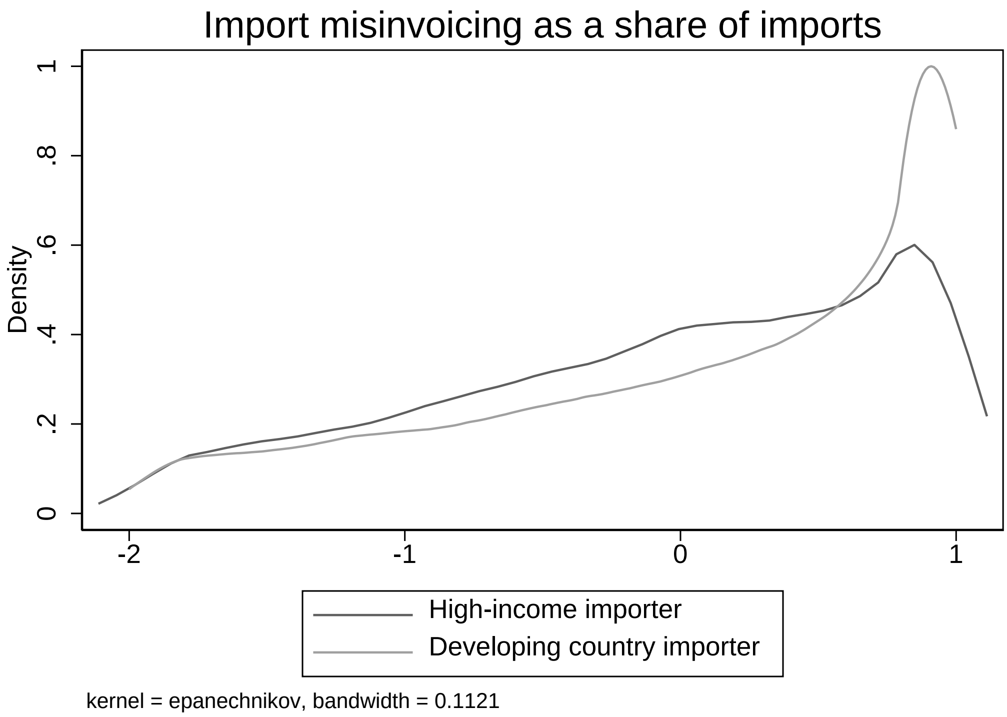
Source: author’s calculations based on UN Comtrade data and CEPII gravity data.

Figure 4: Kernel density of export misinvoicing as a share of actual exports according to importer (partner) country economic development level using six-digit HS-level data



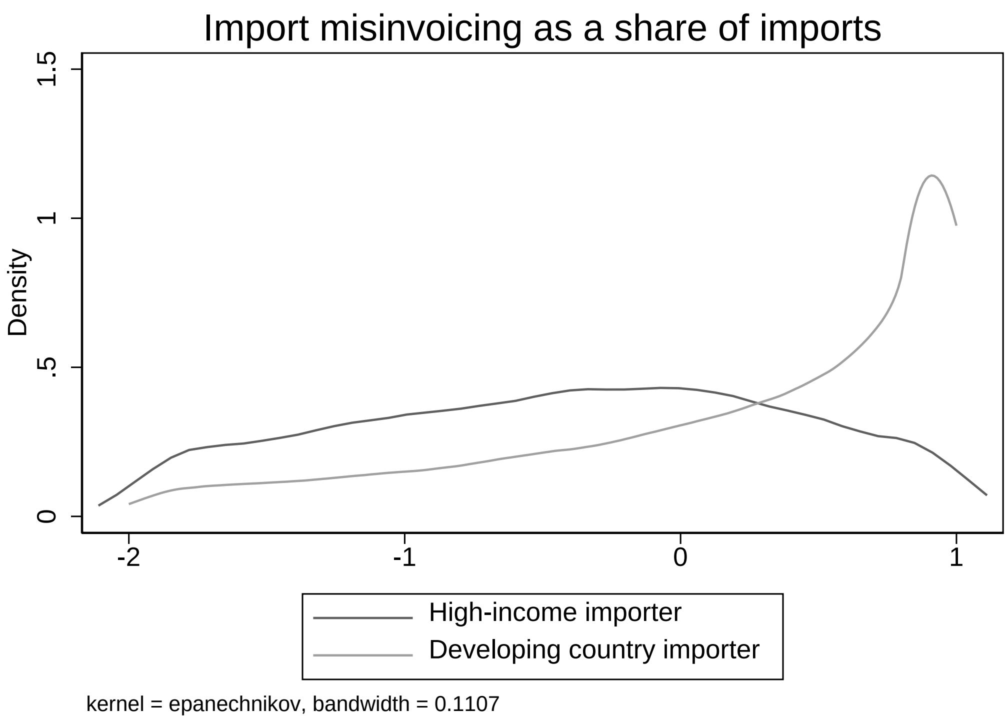
Source: author’s calculations based on UN Comtrade data and CEPII gravity data.

Figure 5: Kernel density of import misinvoicing as a share of actual exports according to importer (reporter) country economic development level using six-digit HS-level data



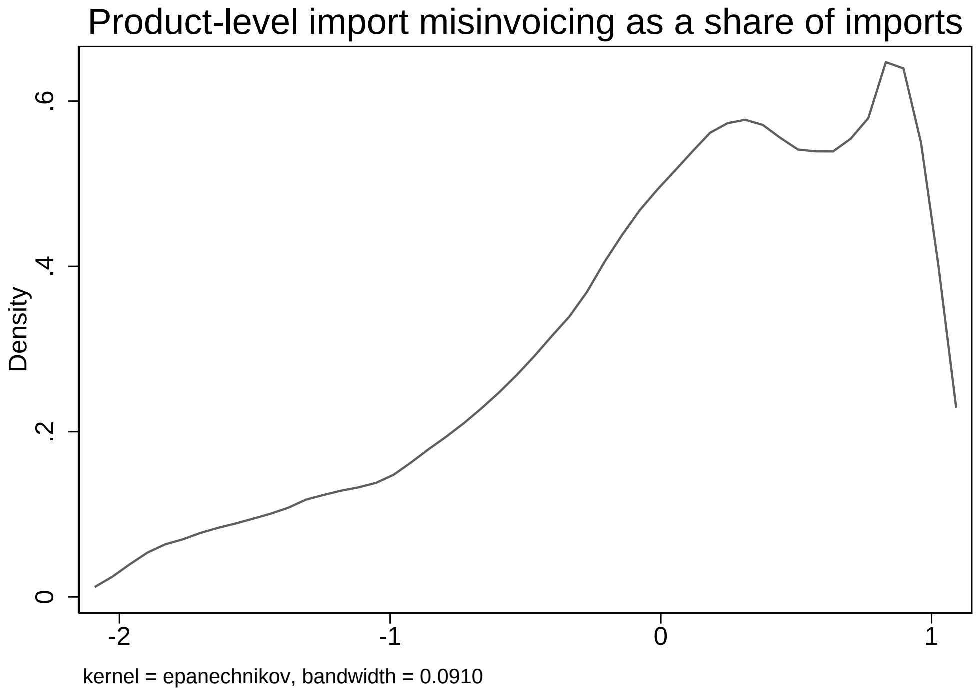
Source: author's calculations based on UN Comtrade data and CEPII gravity data.

Figure 6: Kernel density of import misinvoicing as a share of actual exports according to exporter (partner) country economic development level using six-digit HS-level data



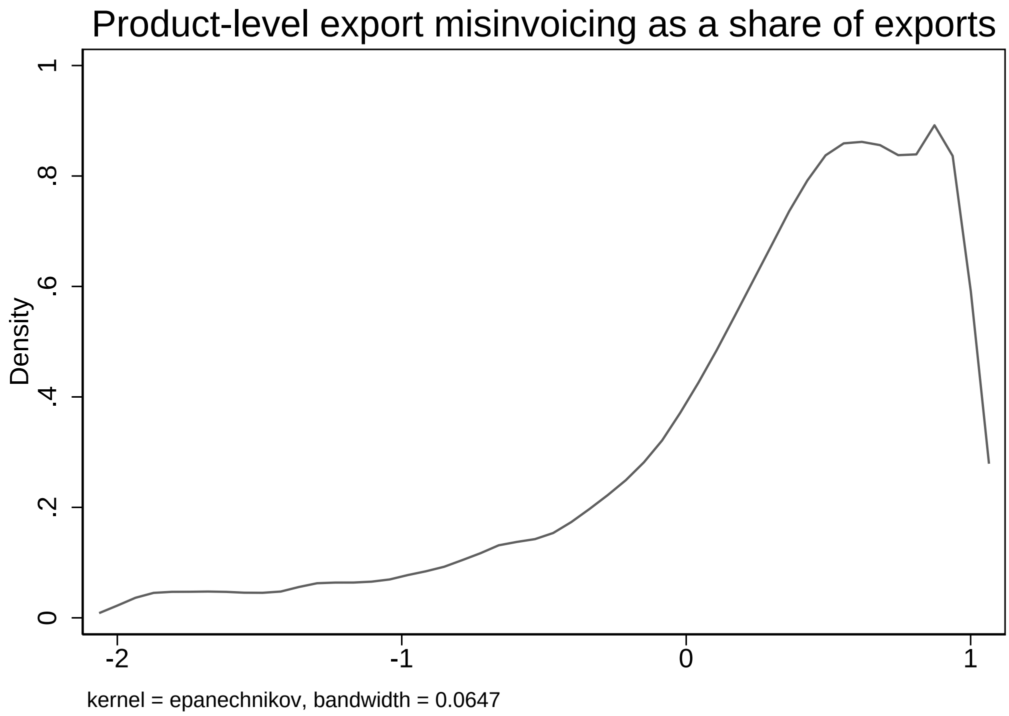
Source: author's calculations based on UN Comtrade data and CEPII gravity data.

Figure 7: Kernel density of import misinvoicing aggregated at the product level as a share of actual imports using six-digit HS-level data



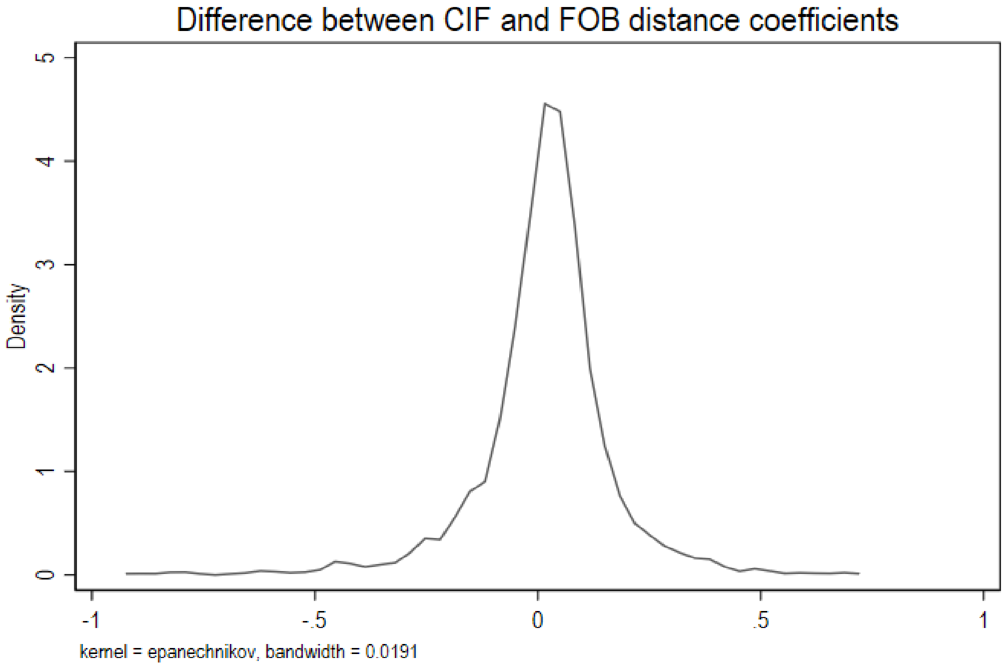
Source: author's calculations based on UN Comtrade data and CEPII gravity data.

Figure 8: Kernel density of export misinvoicing aggregated at the product level as a share of actual exports using six-digit HS-level data



Source: author's calculations based on UN Comtrade data and CEPII gravity data.

Figure 9: Kernel density of the difference between the estimated CIF and FOB distance coefficients using four-digit HS-level data



Source: author's calculations based on UN Comtrade data and CEPII gravity data.

Table 1: Partner countries removed from the raw Comtrade international trade data

Numeric ISO code	Partner	Three-letter ISO code
0	World	WLD
490	Other Asia, nes	
899	Areas, nes	
568	Other Europe, nes	
837	Bunkers	
839	Special Categories	
527	Oceania, nes	
838	Free Zones	
10	Antarctica	ATA
260	Fr. South Antarctic Terr.	ATF
577	Other Africa, nes	
473	LAIA, nes	
637	North America and Central America, nes	
80	Br. Antarctic Terr.	

Source: UN Comtrade data.

Table 2: Descriptive statistics of Comtrade international trade data

Panel a. Raw data

Trade flow	Year	Obs.	Mean	Std. Dev.	Minimum	Maximum
Imports	2013	7,630,988	4.693	245.815	0	279,485.1
Imports	2014	8,894,711	4.155	214.462	0	253,228.7
Imports	2015	8,955,316	3.598	160.603	0	134,342.9
Imports	2016	8,885,925	3.522	152.189	0	116,660.7
Exports	2013	7,387,952	4.854	264.921	0	293,994.6
Exports	2014	7,762,710	4.756	244.728	0	250,522.8
Exports	2015	7,944,171	4.028	185.265	0	160,892
Exports	2016	7,935,296	3.917	179.017	0	161,065.1

Panel b. Removed reporters and partners from Table 1

Trade flow	Year	Obs.	Mean	Std. Dev.	Minimum	Maximum
Imports	2013	6,823,869	2.488	99.031	0	78,481.79
Imports	2014	7,891,515	2.22	86.602	0	85,646.6
Imports	2015	7,977,846	1.912	66.176	0	49,884.65
Imports	2016	7,921,863	1.856	62.282	0	38,935.52
Exports	2013	6,784,514	2.412	88.248	0	78,396.09
Exports	2014	7,093,703	2.379	81.578	0	85,197.09
Exports	2015	7,270,324	2.047	65.859	0	49,685.57
Exports	2016	7,269,021	1.996	62.687	0	38,841.16

Panel c. Reported zero-trade flows in raw data

Trade flow	Year	Reported zero-trade obs.	Share of total obs. (%)
Imports	2013	1,456	0.02
Imports	2014	2,444	0.03
Imports	2015	3,756	0.04
Imports	2016	4,590	0.05
Exports	2013	2,667	0.04
Exports	2014	3,244	0.04
Exports	2015	6,785	0.09
Exports	2016	7,786	0.10

Note: values in USD million.

Source: author’s calculations based on UN Comtrade data.

Table 3: Descriptive statistics of the variables used in the gravity model estimates

Variable	Obs.	Mean	Std. Dev.	Minimum	Maximum
Contiguity	197,184	0.012	0.110	0.000	1.000
Distance between capitals (km)	202,668	8,432.22	4,704.4	0.995	19,951.16
Common official language	202,668	0.173	0.378	0.000	1.000
Common colonizer	202,668	0.119	0.323	0.000	1.000

Source: author's computation based on CEPII gravity data.

Table 4: Descriptive statistics of the estimated coefficients of equation (1)

Panel a. Imports

Variable	Obs.	Mean	Std. Dev.	Minimum	Maximum
Contiguity	5,266	1.297	1.478	-21.895	84.424
Common official language	5,266	0.593	0.929	-6.293	31.001
Common colonizer	5,266	0.809	1.363	-7.400	14.129
Distance (delta_cif)	5,266	-0.510	0.482	-17.088	3.131
Std. error contiguity	5,266	0.739	3.762	0.234	2.730
Std. error common official language	5,266	0.772	5.435	0.133	3.944
Std. error common colonizer	5,266	0.400	0.275	0.078	4.933
Std. error distance (delta_cif)	5,266	0.143	0.292	0.012	2.442

Panel b. Exports

Variable	Obs.	Mean	Std. Dev.	Minimum	Maximum
Contiguity	5,251	1.605	1.953	-5.171	97.823
Common official language	5,251	0.781	1.872	-4.218	108.311
Common colonizer	5,251	0.570	1.454	-6.861	40.414
Distance (delta_fob)	5,251	-0.532	0.562	-10.487	12.821
Std. error contiguity	5,251	1.664	1.205	0.031	8.750
Std. error common official language	5,251	0.378	1.436	0.085	10.42
Std. error common colonizer	5,251	0.568	1.064	0.108	2.980
Std. error distance (delta_fob)	5,251	0.241	0.675	0.062	12.225

Source: author's calculations based on UN Comtrade and CEPII gravity data.

Table 5: Descriptive statistics of the estimated Delta_cif and of the Delta_fob

	Delta CIF	Delta FOB	CIF - FOB
Mean	-0.518	-0.541	0.022
Std. Dev.	0.536	0.574	0.469
Variance	0.287	0.330	0.220
Skewness	-7.158	-0.606	-7.492
Kurtosis	193.641	93.858	293.434
Minimum	-17.088	-10.487	-14.274
Maximum	3.131	12.821	8.348
Percentiles			
1%	-2.012	-2.261	-1.037
5%	-1.239	-1.281	-0.369
10%	-1.020	-1.034	-0.221
25%	-0.742	-0.750	-0.064
50%	-0.477	-0.495	0.027
75%	-0.249	-0.267	0.117
90%	-0.026	-0.060	0.255
95%	0.122	0.096	0.395
99%	0.490	0.571	1.228
Observations	5,228	5,228	5,228

Source: author's calculations based on UN Comtrade and CEPII gravity data.

Table 6: Descriptive statistics of exports and estimated export value

Variable	Year	Obs.	Mean	Std. Dev.	Minimum	Maximum
Actual exports	2013	4,225,002	3.539	103.946	0.000	78,396.090
Estimated export value	2013	4,225,002	3.173	179.538	0.000	141,989.000
Estimated export misinvoicing	2013	4,225,002	0.366	131.721	-105,071.9	60,021.05
Actual exports	2014	5,011,972	3.157	92.807	0.000	85,197.090
Estimated export value	2014	5,011,972	2.794	168.346	0.000	138,674.100
Estimated export misinvoicing	2014	5,011,972	0.362	124.840	-102,417.4	41,203.840
Actual exports	2015	5,009,388	2.797	78.199	0.000	49,685.570
Estimated export value	2015	5,009,388	2.503	148.765	0.000	131,231.400
Estimated export misinvoicing	2015	5,009,388	0.294	113.155	-100,083.4	34,982.040
Actual exports	2016	5,037,713	2.716	74.658	0.000	38,841.160
Estimated export value	2016	5,037,713	2.442	144.443	0.000	118,875.800
Estimated export misinvoicing	2016	5,037,713	0.274	110.452	-91057.52	36,601.220

Note: values in USD million.

Source: author's calculations based on UN Comtrade and CEPII gravity data.

Table 7: Descriptive statistics of imports and estimated import value

Variable	Year	Obs.	Mean	Std. Dev.	Minimum	Maximum
Actual imports	2013	4,471,366	3.425	108.018	0.000	78,481.790
Estimated import values	2013	4,471,366	3.137	97.908	0.000	78,705.000
Estimated import misinvoicing	2013	4,471,366	0.289	100.117	-59,561.35	62,330.240
Actual imports	2014	5,319,044	3.031	94.560	0.000	85,646.600
Estimated import values	2014	5,319,044	2.798	94.257	0.000	77,917.130
Estimated import misinvoicing	2014	5,319,044	0.233	89.069	-59,192.99	37,459.220
Actual imports	2015	5,364,224	2.661	77.020	0.000	49,884.650
Estimated import values	2015	5,364,224	2.464	77.002	0.000	70,717.070
Estimated import misinvoicing	2015	5,364,224	0.196	73.423	-53,829.98	35,658.810
Actual imports	2016	5,311,964	2.591	73.108	0.000	38,935.520
Estimated import values	2016	5,311,964	2.418	75.094	0.000	70,517.430
Estimated import misinvoicing	2016	5,311,964	0.173	70.432	-56,753.67	32,156.320

Note: values in USD million.

Source: author's calculations based on UN Comtrade and CEPII gravity data.

Table 8: Descriptive statistics of exports and estimated export value according to economic development level

Panel a. High-income countries only

Variable	Year	Obs.	Mean	Std. Dev.	Minimum	Maximum
Actual exports	2013	2,673,437	3.369	100.368	0.000	78,396.090
Estimated export value	2013	2,673,437	2.749	143.332	0.000	116,674.500
Estimated export misinvoicing	2013	2,673,437	0.620	93.934	-65,925.65	60,021.050
Actual exports	2014	3,034,442	3.048	86.189	0.000	85,197.090
Estimated export value	2014	3,034,442	2.439	126.263	0.000	117,534.000
Estimated export misinvoicing	2014	3,034,442	0.609	78.211	-65,351.99	17,899.770
Actual exports	2015	3,038,951	2.694	69.665	0.000	49,685.570
Estimated export value	2015	3,038,951	2.162	107.670	0.000	72,029.430
Estimated export misinvoicing	2015	3,038,951	0.532	77.452	-58,312.54	23,804.260
Actual exports	2016	3,051,227	2.656	66.913	0.000	38,841.160
Estimated export value	2016	3,051,227	2.109	97.985	0.000	73,856.570
Estimated export misinvoicing	2016	3,051,227	0.547	69.356	-59,225.64	20,107.580

Panel b. Developing countries only

Variable	Year	Obs.	Mean	Std. Dev.	Minimum	Maximum
Actual exports	2013	1,551,565	3.832	109.836	0.000	36,917.090
Estimated export value	2013	1,551,565	3.903	228.856	0.000	141,989.000
Estimated export misinvoicing	2013	1,551,565	-0.071	179.004	-105071.9	23,952.910
Actual exports	2013	1,977,530	3.324	102.131	0.000	44,185.970
Estimated export value	2014	1,977,530	3.339	217.634	0.000	138,674.100
Estimated export misinvoicing	2014	1,977,530	-0.015	173.531	-102417.4	41,203.84
Actual exports	2014	1,970,437	2.956	89.786	0.000	39,428.580
Estimated export value	2015	1,970,437	3.029	195.918	0.000	131,231.400
Estimated export misinvoicing	2015	1,970,437	-0.073	152.642	-1,00083.4	34,982.040
Actual exports	2015	1,986,486	2.809	85.194	0.000	36,977.490
Estimated export value	2016	1,986,486	2.953	195.353	0.000	118,875.800
Estimated export misinvoicing	2016	1,986,486	-0.145	153.459	-91057.520	36,601.220

Note: values in USD million.

Source: author’s calculations based on UN Comtrade and CEPII gravity data.

Table 9: Descriptive statistics of imports and estimated import value according to economic development level

Panel a. High-income countries only

Variable	Year	Obs.	Mean	Std. Dev.	Minimum	Maximum
Actual imports	2013	1,904,359	4.868	128.484	0.000	78,481.790
Estimated import values	2013	1,904,359	4.751	111.115	0.000	51,596.590
Estimated import misinvoicing	2013	1,904,359	0.117	109.469	-40,770.17	62,330.240
Actual imports	2014	1,985,024	4.819	119.838	0.000	85,646.600
Estimated import values	2014	1,985,024	4.779	113.359	0.000	55,576.160
Estimated import misinvoicing	2014	1,985,024	0.040	100.615	-41,076.71	37,459.220
Actual imports	2015	2,080,156	4.106	93.565	0.000	49,884.650
Estimated import values	2015	2,080,156	4.108	92.829	0.000	37,215.860
Estimated import misinvoicing	2015	2,080,156	-0.001	82.325	-33,150.39	35,658.810
Actual imports	2016	2,060,741	4.096	89.614	0.000	38,935.520
Estimated import values	2016	2,060,741	4.100	88.260	0.000	37,909.980
Estimated import misinvoicing	2016	2,060,741	-0.004	78.081	-34,099.24	32,156.320

Panel b. Developing countries only

Variable	Year	Obs.	Mean	Std. Dev.	Minimum	Maximum
Actual imports	2013	2,567,007	2.355	89.858	0.000	54,925.030
Estimated import values	2013	2,567,007	1.939	86.803	0.000	78,705.000
Estimated import misinvoicing	2013	2,567,007	0.416	92.570	-59,561.35	46,256.770
Actual imports	2014	3,334,020	1.966	75.576	0.000	54,048.450
Estimated import values	2014	3,334,020	1.619	80.743	0.000	77,917.130
Estimated import misinvoicing	2014	3,334,020	0.347	81.421	-59,192.99	33,549.890
Actual imports	2015	3,284,068	1.745	64.360	0.000	35,605.290
Estimated import values	2015	3,284,068	1.424	64.991	0.000	70,717.070
Estimated import misinvoicing	2015	3,284,068	0.322	67.176	-53,829.98	30,031.530
Actual imports	2016	3,251,223	1.637	60.332	0.000	35,512.930
Estimated import values	2016	3,251,223	1.352	65.368	0.000	70,517.430
Estimated import misinvoicing	2016	3,251,223	0.284	65.120	-56,753.67	28,800.670

Note: values in USD million.

Source: author’s calculations based on UN Comtrade and CEPII gravity data.

Table 10: Products with the largest estimate of export misinvoicing

Panel a. Over-invoicing

Year	HS2012	Description	Actual exports	Estimated export values
2014	851420	Furnaces and ovens; electric, for industrial or laboratory use, functioning by induction or dielectric loss	492.729	49.350
2014	252329	Cement; portland, other than white, whether or not artificially coloured	6,421.582	644.961
2014	740729	Copper; bars, rods and profiles, of copper alloys (other than copper-zinc base alloys)	980.079	98.518
2015	600641	Fabrics; knitted or crocheted fabrics, other than those of headings 60.01 to 60.04, of artificial fibres, unbleached or bleached	40.318	4.055
2014	071334	Vegetables, leguminous; bambara beans (Vigna subterranea or Voandzeia subterranea), shelled, whether or not skinned or split, dried	1.219	0.123

Panel b. Under-invoicing

Year	HS2012	Description	Actual exports	Estimated export values
2013	282619	Fluorides; other than of ammonium or sodium or aluminium	237.105	711.143
2015	940381	Furniture (o/than seats) of bamboo or rattan	150.134	449.868
2013	810194	Tungsten, unwrought (including bars and rods obtained simply by sintering)	27.192	81.459
2016	854232	Memories	88,258.312	264,322.84
2013	551219	Woven fabrics containing 85% or more by weight of polyester staple fibres, other than unbleached or bleached	1,354.391	4,054.696

Note: values in USD million.

Source: author's calculations based on UN Comtrade and CEPII gravity data.

Table 11: Products with the largest estimated import misinvoicing

Panel a. Over-invoicing

Year	HS2012	Description	Actual imports	Estimated import values
2016	902219	Apparatus based on the use of x-rays, including radiography or radiotherapy apparatus; for other than medical, surgical, dental or veterinary uses	2,125.749	6,377.237
2013	252921	Fluorspar; containing by weight 97% or less of calcium fluoride	239.0214	717.0179
2014	490290	Newspapers, journals and periodicals; whether or not illustrated or containing advertising material, appearing less frequently than four times a week	4,469.699	13,406.21
2015	030771	Molluscs; clams, cockles and ark shells (families Arcidae, Arcticidae, Cardiidae, Donacidae, Hiatellidae, Mactridae, Mesodesmatidae, Myidae, Semelidae, Solecurtidae, Solenidae, Tridacnidae and Veneridae), whether in shell or not, live, fresh or chilled	333.7105	1,000.622
2016	640690	Footwear; parts, n.e.c. in heading 6406	2,365.429	7,087.989

Panel b. Under-invoicing

Year	HS2012	Description	Actual imports	Estimated import values
2014	030242	Fish; fresh or chilled, anchovies (Engraulis spp.), excluding fillets, livers, roes, and other fish meat of heading 0304	61.119	6.125
2014	400241	Rubber; synthetic, chloroprene (chlorobutadiene) rubber (CR), latex, in primary forms or in plates, sheets or strip	155.298	15.567
2016	551411	Fabrics, woven; plain weave, unbleached or bleached, containing less than 85% by weight of polyester staple fibres, mixed mainly or solely with cotton, exceeding 170g/m2	68.420	6.859
2016	060490	Foliage, branches and other parts of plants, without flowers or flower buds, and grasses, mosses and lichens; suitable for bouquets or for ornamental purposes, dried, dyed, bleached, impregnated or otherwise prepared	248.690	24.972
2013	440726	Wood, tropical; white lauan, white meranti, white seraya, yellow meranti and alan, sawn or chipped lengthwise, sliced or peeled, whether or not planed, sanded or end-jointed, thicker than 6mm	62.150	6.244

Note: values in USD million.

Source: author’s calculations based on UN Comtrade and CEPII gravity data.

Table 12: Share of over- and under-invoicing of exports and imports according to economic development level

Panel a. Imports

Year	Country type	Share of over-invoicing
2013	Developed	32.5%
2014	Developed	30.7%
2015	Developed	32.6%
2016	Developed	32.1%
2013	Developing	73.4%
2014	Developing	72.8%
2015	Developing	76.1%
2016	Developing	73.2%

Panel b. Exports

Year	Country type	Share of over-invoicing
2013	Developed	56.1%
2014	Developed	58.8%
2015	Developed	57.6%
2016	Developed	57.5%
2013	Developing	42.2%
2014	Developing	40.6%
2015	Developing	40.9%
2016	Developing	40.5%

Note: the share of under-invoicing consists of 100% - share of over-invoicing.

Source: author's calculations based on UN Comtrade and CEPII gravity data.

Table 13: Descriptive statistics of the estimated coefficients of equation (1) using four-digit HS-level data

Panel a. Imports

Variable	Obs.	Mean	Std. Dev.	Minimum	Maximum
Contiguity	1,228	1.26	.805	-14.87	11.259
Common official language	1,228	.54	.911	-2.772	24.58
Common colonizer	1,228	.763	1.022	-3.135	6.208
Distance (delta_cif)	1,228	-.483	.395	-2.92	1.693
Std. error contiguity	1,228	.167	.229	0	6.506
Std. error common official language	1,228	.201	1.4	.053	48.921
Std. error common colonizer	1,228	.308	.199	0	2.162
Std. error distance (delta_cif)	1,228	.061	.047	.025	.75

Panel b. Exports

Variable	Obs.	Mean	Std. Dev.	Minimum	Maximum
Contiguity	1,228	1.515	0.676	-0.669	10.439
Common official language	1,228	0.647	0.594	-3.704	3.843
Common colonizer	1,228	0.564	1.057	-5.451	8.679
Distance (delta_fob)	1,228	-0.5	0.401	-4.576	0.976
Std. error contiguity	1,228	0.163	0.122	0	2.069
Std. error common official language	1,228	0.159	0.101	0.054	1.311
Std. error common colonizer	1,228	0.296	0.191	0	2.569
Std. error distance (delta_fob)	1,228	0.062	0.056	0.026	1.209

Source: author's calculations based on UN Comtrade and CEPII gravity data.

Table 14: Descriptive statistics of the estimated Delta_cif and of the estimated Delta_fob using four-digit HS-level data

	Delta CIF	Delta FOB	CIF – FOB
Mean	-0.483	-0.500	0.018
Std. Dev.	0.395	0.401	0.239
Variance	0.156	0.161	0.057
Skewness	-0.774	-1.864	1.032
Kurtosis	7.596	18.561	84.368
Minimum	-2.920	-4.576	-3.171
Maximum	1.693	0.976	3.369
Percentiles			
1%	-1.634	-1.664	-0.540
5%	-1.140	-1.144	-0.234
10%	-0.921	-0.916	-0.146
25%	-0.681	-0.692	-0.038
50%	-0.459	-0.468	0.025
75%	-0.252	-0.274	0.081
90%	-0.065	-0.103	0.160
95%	0.072	0.070	0.244
99%	0.408	0.350	0.471
Observations	1,228	1,228	1,228

Source: author’s calculations based on UN Comtrade and CEPII gravity data.