Reconciling International Trade Data

Karam Shaar [a]

Abstract

International trade data are filled with discrepancies–where two countries report different values of trade with each other. I develop a novel trade data quality index for reconciling the discrepancies in bilateral trade data. I calculate the quality for each country's imports and exports separately for every year from 1962 to 2016 and reconcile international trade data by picking the value reported by the country with higher data quality in every bilateral flow. The reconciled data reshape our views on international trade: (a) countries with low data quality under-report their imports and exports: low-quality reporters are 14% more open to trade using reconciled data; (b) corruption, the level of development, and erroneous reporting can explain data quality; (c) importers' data are more accurate; (d) China tends to under-report its exports and over-report its imports, while there is only a small difference between US self-reported and reconciled data.

Keywords: Trade data quality, data discrepancy, trade data reconciliation, international trade

JEL Classification: F01, C20, C18

---

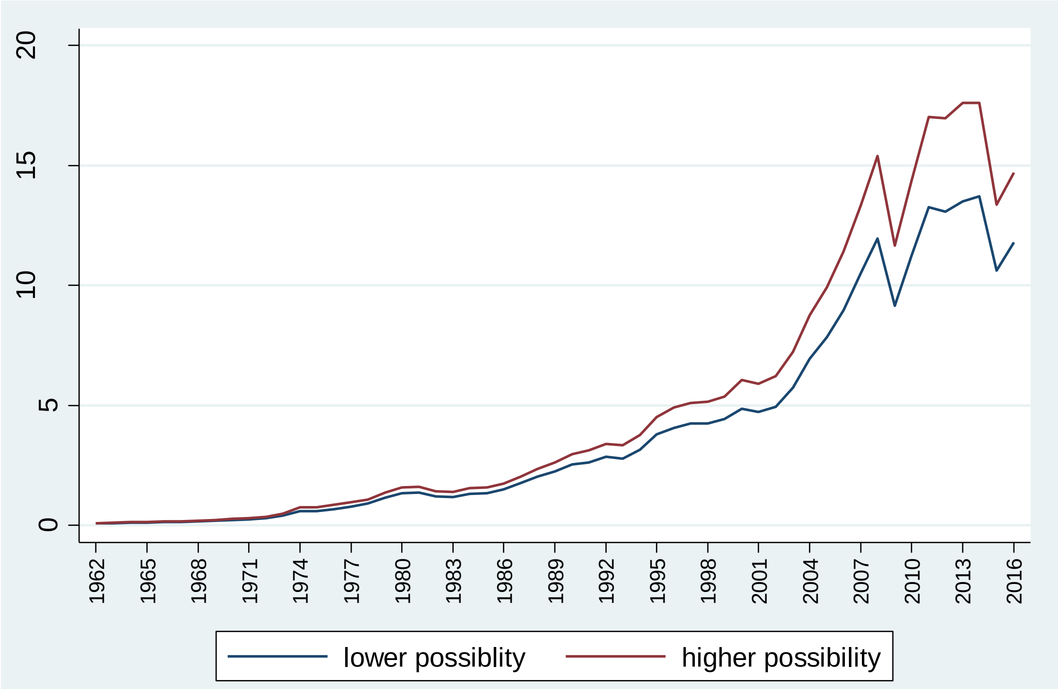
[a]: Victoria University of Wellington, Wellington 6140, New Zealand, PO box 600, karam.shaar@vuw.ac.nz

ORCID: 0000-0003-3088-5156

1. Introduction

When trade flows from country A to country B, A reports its value as exports and B reports it as imports. Therefore, every bilateral trade flow is reported twice. Trade data discrepancy occurs when two countries report different values of trade with each other. To illustrate the scale of discrepancies in international trade data, Figure 1 plots global trade through aggregating bilateral trade flows once using the higher reported value and once using the lower reported value. Which of the two possibilities represents actual global trade?

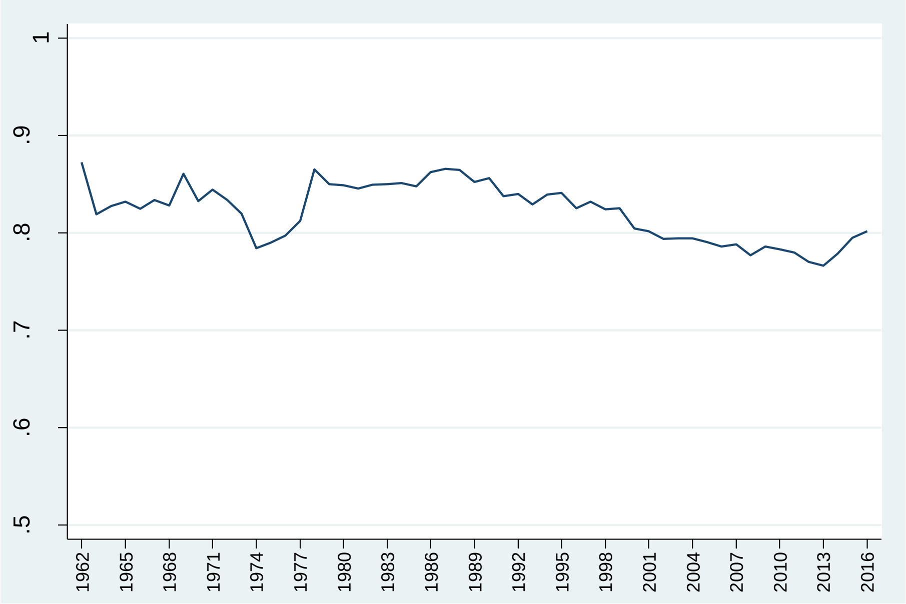
Figure 1: Global trade based on higher vs lower reports in every bilateral relation ($US trillions)



Source: The Standard International Trade Classification (SITC)
Note: all trade data are on Free-on-Board basis (see section 5)

Figure 1 shows that global trade in 2016 can be anywhere between $11.8 trillion and $14.7 trillion depending on the claims about bilateral trade that you think are more reliable. Figure 2 plots the lower ‘possibility’ relative to the higher ‘possibility’: the ratio of possibilities was in decline most of the time since the late 1980s and stood at 0.8 in 2016 (Figure 2).

Figure 2: The lower possibility of global trade divided by the higher possibility



Substantial discrepancies can mean a study that uses data reported by a particular partner would reach a different conclusion than if it used the data reported by the other. For example, Bahmani-Oskooee et al. (2013) report ‘drastically different results for the impact of the exchange rate on trade between the US and South Korea depending on which of the two countries reported the data.’[1] As stated by Gujarati (2004), ‘the researcher should always keep in mind that the results of the research are only as good as the quality of the data.’

Four organizations attempted to reconcile trade data for a large number of countries and over an extended period. I review the methodologies followed by the four institutions in section 4 and highlight the need for an improved method. The main issues with the literature are the use of biased measures of reporter quality, the subjective choices of acceptable quality thresholds, and not accounting for the role of data availability as an essential dimension of reporter quality.

I construct a Trade Data Quality Index (TDQI) based on the following notion: the more a country’s reports on bilateral trade differ from the corresponding reports of its partners, the more likely it is a low-quality reporter. TDQI accounts for insurance and freight costs, the relative significance of each trade flow, and the cases where a country does not report data, to begin with. The index ranges from 0 to 100, and it is calculated for each country’s imports and exports separately in every year from 1962 to 2016. TDQI investigates the quality of more than 98% of global trade, which comprises around 1.3 million observations.

When two countries report different values for the same trade flow, one of the two values must be closer to the truth. TDQI is used for reconciling trade data. In any trade flow, I pick the value reported by the partner less likely to be in error: the partner with higher TDQI. I take the costs of insurance and transportation into consideration depending on whether I use the value reported by the exporter or the importer as the more reliable value for

[1] Bahmani-Oskooee et al. (2013) examined the impact of exchange rates on US-Korea trade in 96 different industries and showed that for the same period of time and using the same econometric specifications, ‘cointegration analysis produces drastically different results depending on the choice of the reporting country. Nonparametric analysis shows that the Korean imports are more sensitive to real exchange rate fluctuations than US exports…’.

reconciliation. The reconciliation does not invent new numbers: the reconciled data come from whichever partner the rest of the world deems more reliable.

I find that over the last five years, global trade using exporters' data was 6.3% lower than using reconciled data; while 3.3% lower using importers' data. On average, global trade is 4.8% higher using reconciled data. The differences are because low-quality reporters under-report their trade (imports and exports) although imports data tend to be more reliable. Erroneous reporting is also prevalent among low-quality reporters. Reconciled and self-reported data for high-quality reporters are similar for two reasons. First, in trade with a low-quality reporter, the value is picked from the country with higher quality: self-reported data are the source for reconciliation. Second, in trade between high-quality reporters, the magnitude of the discrepancies is small.

I show that several factors can explain trade data quality, such as the size of the economy, the level of development, and corruption in the public sector. Although global trade has never been more complex, import and export data quality improved in most countries over the period of the study. For global aggregate trade, however, the quality has been deteriorating over most of the past two decades because of the increasing importance of low-quality reporters.[2] Where China tends to under-report its exports and over-report its imports, I observe trivial differences between reconciled and self-reported data for the US.

TDQI materially improves international trade data quality. I recommend future studies on international trade use the data reconciled in this study as the differences between the reconciled and unreconciled data have important implications for our understanding of international trade. For illustration, I calculate trade openness for every country in 2016, once using the data reconciled here and once using self-reported data.[3] Low-quality reporters are, on average, 14% more open to trade using reconciled data. Countries with medium and high-quality data have roughly the same trade openness using reconciled and unreconciled data.[4] The literature largely depends on self-reported trade data. [5]

The following link provides the results of TDQI, the reconciled and unreconciled international trade data on a bilateral level and for a country's trade with the world combined. It also provides the Stata programming codes used in the calculations.
https://goo.gl/cGVXDk

2. Sources of unreconciled international trade data

Most countries report their commodity trade to international institutions. The institutions, in their turn, aim to achieve unanimous scales and definitions for international trade through adopting commodity classifications and promoting standard reporting practices. Two institutions maintain the most popular commodity classifications. First, the UN Statistics Division, which maintains the Standard International Trade Classification (SITC).[6] The SITC is based on the economic functions of commodities at different stages of development. It is

[2] Importance is measured in terms of the trade share.
[3] Trade openness is defined as (exports+imports)/GDP.
[4] I break the reporters into three groups (high, medium, and low-quality) based on the average quality of their imports and exports data.
[5] To confirm this, I reviewed the top 30 papers in terms of citations on Scopus search engine, where the papers were published since 2015 and contained 'international trade' as a keyword. None of the papers discuss the issue of trade data discrepancy, while 29% of them use a data source that addresses the problem.
[6] The SITC is a part of the Commodity Trade Statistics Database (COMTRADE).

the oldest international commodity classification: the first version was released in 1950; the earliest online version is available from 1962. Second, the World Customs Organization maintains the Harmonized System (HS). The HS is a multipurpose classification that was implemented in 1988. There are correspondence tables between the two classifications (de Saint Vaulry, 2008).

The HS and the SITC were revised several times to account for the changes in the nature of the traded goods. Both classifications publish bilateral data on the commodity as well as aggregate levels. Self-reported commodity trade does not necessarily add up to self-reported aggregate trade.[7] There are marginal differences between bilateral aggregate data published by the HS and the SITC, even across different revisions: the differences are on the commodity level, while this study is concerned with aggregate trade.

The IMF publishes bilateral aggregate trade in its Direction of Trade (DOT) Database. I compare DOT with SITC/HS data for the period 1962–2016 and find the data sets identical in most cases. DOT, occasionally, replaces missing values with partners' corresponding values; it is not clear under what circumstances the replacements are made. DOT uses a 10% share of the reported value to account for insurance and transportation when converting between the data reported by the exporter and the importer to replace missing values.[8] DOT also estimates annual trade values based on available monthly reports. The WTO does not report bilateral trade, and the number of reporting countries is smaller than in the SITC and the HS. The World Bank retrieves its data from the WTO, the IMF, and COMTRADE.

3. Causes of trade data discrepancy

Several studies investigated why some countries might report different values of trade with each other. Yeats (1990) compared sub-Saharan trade data with the data reported by the rest of the world and detected significant discrepancies, stating that mis-invoicing and smuggling are 'apparently responsible.' Confirming the logic of the Trade Data Quality Index proposed in this study, Yeats (1990) shows that although significant disparities exist in data on trade with developed countries, the average differences in intra-African trade statistics are more substantial. Finally, the paper suggests that sub-Saharan African data should not be relied on to indicate 'the level, composition, or even direction and trends in African trade.'

West (1995) focused on US-China trade and claimed that most of the discrepancies in their data are due to two factors. First, re-exportation through Hong Kong. Second, the price mark-ups added by merchants in Hong-Kong after the products had left China. Several studies, later on, focused on US-China trade.

Makhoul and Otterstrom (1998) examined the discrepancies in the Direction of Trade database, managed by the IMF, and found that taking insurance and freight into account when comparing the data of importers and exporters can only explain a small portion of trade data discrepancies. Something, I echo here in this study.

Vincent (2004) used econometric analysis of bilateral trade statistics for Romania and other European countries and found that measurement error, shipment lags, and intentional under-reporting all play a role in explaining discrepancies for two different types of sawn wood.

[7] Due to confidentiality, countries may not report some of their detailed trade.
[8] See section 3 for more details on export and import valuation methods.

Guo (2010) analysed the asymmetries in trade between China and its top five trading partners during the period 1992–2008. The paper listed several reasons for the discrepancies, such as the different pricing systems between imports and exports, the different trade systems among countries, re-exporting, and re-importing. At odds with what I aim to show, Guo (2010) concluded that it is difficult to find a systematic way to correct asymmetries in international trade statistics among countries.

Wang, Gehlhar, and Yao (2010) proposed a mathematical programming model to reconcile trade data discrepancies in the presence of a third re-exporting country. However, the proposed method requires initialization with detailed estimates of actual bilateral trade flows and data reliability indices, which are not observable from the data.

Ferrantino et al. (2012) investigated the direct trade between the US and China and found ‘strong statistical evidence’ of under-reporting exports at the Chinese border to avoid paying value-added tax (VAT). I also confirm this finding. The under-reporting of Chinese exports to the US accounted for approximately two-thirds of the discrepancy over the period 2002-2008.

Simola (2012) assessed trade data quality in Russia and showed that while some discrepancies can be explained by the misclassification of commodities, some discrepancies reflect deliberate misreporting.

Several studies in the literature also showed that trade data discrepancy could be explained by favourable bilateral trade agreements, export schemes, tariffs, and corruption. For example, Fisman and Wei (2009) showed that bilateral trade discrepancies are highly correlated with the corruption level of the exporting country as measured by commonly used survey-based indices and that this correlation is stronger for artefact-rich countries.

Based on the studies reviewed above, I break the causes of trade data discrepancies into four types:

Table 1: Potential causes of trade data discrepancy

Type	Cause	Explanation
lack of uniformity in data compilation methodologies between the partners	what constitutes imports and exports?	the treatment of re-exports, goods in transit, inward and outward processing, and re-imports
	partner attribution	attributing exports to the final known destination or the country of consignment; attributing imports to the country of origin or the country of consignment
	geographical definition of a trading partner	example: whether the Virgin Islands are a part of the United States or a separate entity/country
	conversion to foreign currency	what method is used to express the value of trade in terms of a foreign currency
uniformity in data compilation methodologies	the valuation method of imports and exports	imports are usually reported on cost-insurance-freight (CIF) basis; exports are usually reported as free-on-board (FOB)
	timing effect	shipments are registered at different points of time by both countries since they are registered as they happen (shipment lags)

corruption	mis-invoicing	over-invoicing the value of a shipment to take advantage of specific export support schemes or under-invoicing for duty evasion
	smuggling	not registering a shipment for duty avoidance or due to the illegality of the traded goods by either one of the partners or both
	partner misattribution	attributing trade to another partner (against the stated methodology), to benefit from lower duties
erroneous reporting		failure to value the goods correctly or to attribute the flow to the right partner due to negligence

4. Previous efforts in reconciling trade data

The literature documents multiple attempts investigating the causes of discrepancies among a handful of countries (see section 3). The US-Hong Kong-China trilateral trade received most of the attention. Other countries, such as Mexico, the US, and Canada, set up commissions to reconcile the discrepancies with each other. The review in this section is concerned with attempts at reconciling trade data internationally: reconciliation for a large number of countries and over an extended period. Note that all such efforts reconciled the data on the commodity-level; our attempt reconciles aggregate trade only.[9]

In summary, the main problems with the literature are the use of biased measures of reporter quality, the subjective choice of acceptable quality thresholds, and not accounting for the role of data availability as a dimension of reporter quality.

4.1 Global Trade Analysis Project (GTAP)

Gehlhar (1996) reconciles trade data for GTAP starting from the SITC commodity-level and later aggregates the reconciled bilateral flows into different regions. GTAP data are proprietary.

The logic of reconciliation is similar to the one adopted in this study. First, use the overall similarity between a country's reports and the corresponding reports of all of its partners as a measure of reporter quality. Second, reconcile every bilateral flow through picking the value from the reporter with higher quality.

Gehlhar (1996) uses the following measure of similarity between two corresponding values: |export-import|/import. The problem with the measure is the asymmetry to whether the importer or the exporter misreport bilateral trade: the measure yields different results when the difference in reports is the same. The asymmetry means that the measure is biased; the direction of the bias depends on whether the importer or the exporter over or under-reported trade. Section 6 shows that low-quality reporters tend to systematically under-report their trade, which highlights the importance of a symmetric measure of similarity.

Gehlhar (1996) uses a threshold of 0.2 for the measure of similarity to classify every bilateral flow into 'matched' and 'not matched'. Reporter quality is then defined as the sum of the total value of matched flows as a share of total self-reported trade.[10]

[9] Our method can be later extended to investigate the quality on the commodity-level.
[10] What I call 'reporter quality' is called 'reporter accuracy' in Gehlhar (1996). As his study is on commodity-level, the measured accuracy varies by commodity, country, and flow direction (export vs import).

I use a symmetric measure of similarity and follow a ranking method to avoid setting a subjective threshold. Table 2 provides an illustrative example.

Table 2: An illustrative example of different methodologies of measuring the similarity between two corresponding values

	New Zealand-reported imports from the US (m)	US-reported exports to New Zealand (x)	Gehlhar (1996) similarity measure |x-m|/m	proposed similarity measure |x-m|/(x+m)
scenario 1	33	40	0.21 (not matched)	0.10
scenario 2	40	33	0.18 (matched)	0.10

As in the case of other studies in the literature, Gehlhar (1996) does not account for the role of data availability in measuring quality. This paper argues that reporter quality should be a function of available and unavailable data. A country should not be considered a high-quality reporter because it reported a few values consistent with partners' reports together with many no-trade values that differ substantially from partner's reports.

4.2 The National Bureau of Economic Research-the United Nations (NBER-UN)

Feenstra et al. (2005) reconcile the data for the NBER-UN database. Country-specific knowledge about reporting practices over time is used to the discrepancies in the SITC data on a case-by-case basis. Such knowledge relates to the treatment of re-exports, the geographic boundaries of the partner country, and so on.

However, Feenstra et al. (2005) give primacy to the importers' reports, whenever available, assuming they are more accurate. If the importer's report is not available for a country-pair, the corresponding exporter report is used instead. I argue that primacy should not be given to either side. Although TDQI confirms that importers' reports are more accurate on average, the index also highlights the numerous cases where the opposite is true. The reconciled value should be picked from whichever partner has a higher data quality. As section 6 shows, throughout the period of the study, exporters' data are used 43% of the time for reconciling bilateral discrepancies, while importers' data are used for the remaining 57% of the cases.

The case-by-case nature of the reconciliation means it takes a long time to reconcile the data. Timely reconciliation is essential, especially for policymakers. NBER-UN data are currently available until the year 2000.

4.3 Centre d'Etudes Prospectives et d'Informations Internationales (CEPII)

Two studies from CEPII attempted to reconcile commodity trade. De Saint Vaulry (2008) reconciles the data for the period 1967–2005 using the STIC data. The study breaks each country into four quadrants based on the timeliness as well as the quality of reporting. The reconciliation depends on the quadrants of the partners , and it is achieved using the weighted-average of reports. However, the weights are determined by the author. For example, if a country in quadrant x traded with another in quadrant y, a weighted average of reports of one-third to two-thirds is used to reconcile the discrepancy. The results are likely to be sensitive to the choice of averaging weights.

Gaulier and Zignago (2010) also reconcile the data for CEPII in what is called the BACI database. Reporter quality is estimated using a regression that makes strong assumptions about the distribution of the discrepancies. For instance, the discrepancies of reported data from the truth are assumed to follow a multiplicative log-normal form with a mean of zero. I argue that the difference between reported and actual trade is not likely to be centred around zero (actual trade is expected to be higher due to under-invoicing, and smuggling [see Figures 10 and 11]).

Most importantly, BACI, like du Saint Vaulry (2008), uses the quality-weighted average of reports to reconcile every flow. The problem with this approach is that the reconciled value in a certain trade flow will end up worse than the report of one of the two partners. When two countries report different values for the same trade flow, one of the two values must be closer to the truth. As done by Gehlhar (1996), it is safer to pick the report from whichever partner with higher quality to ensure the reconciled data are an improvement over self-reported data. As the literature shows and TDQI confirms, the values reported by some low-quality reporters mean next to nothing; such reporters should not be given any weight in the reconciliation process. Finally, for the sake of credibility, it is essential not to be seen as making up new numbers for international trade.

4.4 Organization for Economic Co-operation and Development (OECD)

Fortanier (2016) reconciles the data for the OECD using the HS from 2007 to 2016. Reporter quality is defined as the 'share of bilateral trade for which the absolute difference with the corresponding trade data is less than or equal to 10% of the sum of these two value flows.'[11] Therefore, Fortanier (2016) uses the same similarly measure followed here. However, like Gehlhar (1996), Fortainer (2016) applies an arbitrary threshold for classifying every two corresponding reports into 'matched' and 'not matched'. As argued earlier, the findings are possibly sensitive to the choice of the threshold.

Most importantly, as in the case of Gaulier and Zignago (2010), the study uses the quality-weighted average of reports in reconciling every bilateral flow. As discussed in section 4.3, I argue that this method might end up worsening the quality of trade data than improving it. Finally, Fortainer (2016), like other studies in the literature does not account for the role of data availability in the measurement of reporter quality.

5. Data and Methodology

To assess data quality, I need to rely on raw data where the international publishing institution made no changes to the data reported by countries. The dataset also needs to be comprehensive. The HS and the SITC meet these two criteria (refer to section 2 for more details). Since there are only small differences between bilateral aggregate trade data published by the HS and the SITC, I use the data from the SITC, revision 1, to maximize the period of the study.

The annual data extend from 1962 to 2016 and cover all available bilateral aggregate trade; this comes to 1,632,245 observations. The data are in US dollars and exclude the years 2017–2018 since some countries have not reported their trade yet while some other reports are still subject to revisions by the reporting countries.[12] I downloaded the data from the World Integrated Trade Solutions (WITS) website, administered by the World Bank. Several other online portals offer free access to SITC records.

[11] What I call 'data quality index' is called 'asymmetry measure' in Fortanier (2016).

[12] Countries can update their reported trade as more information become available.

The UN Statistics Division regularly publishes its recommendations to individual countries on preferred reporting practices to enhance comparability. The SITC distinguishes between total and gross bilateral trade:

Gross exports= total exports+ re-exports

Gross imports= total imports+ re-imports

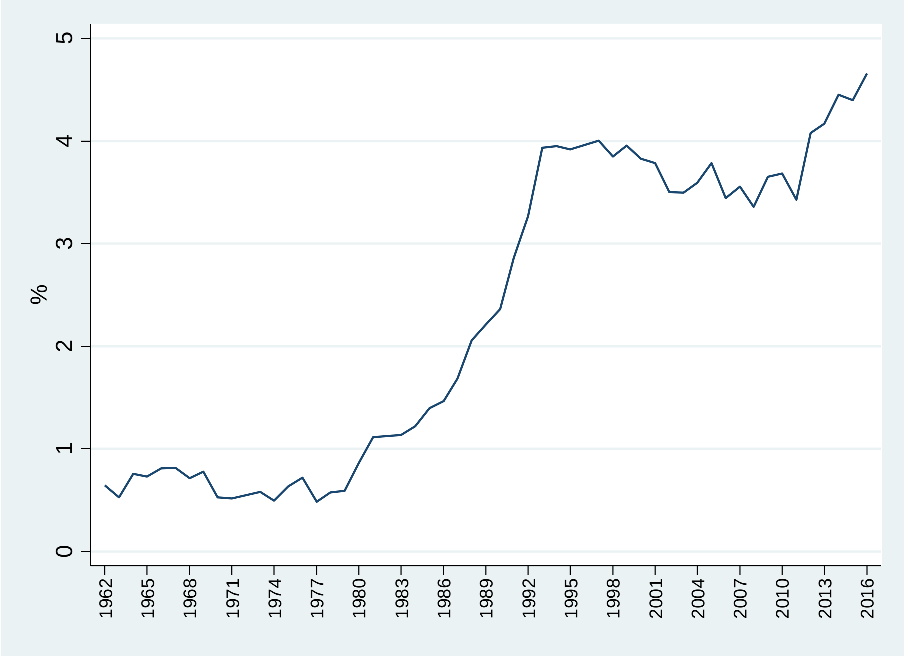
As defined by the UN Statistics Division ('International Merchandise Trade Statistics-Concepts and Definitions ', 2011), re-imports are imports of domestic goods which were previously recorded as exports (returned). Re-exports are exports of foreign goods which were previously recorded as imports and later re-exported without 'substantial changes'. Assuming no re-exports and re-imports take place, total and gross bilateral trade data would be identical. I use 'total' rather than 'gross' bilateral data in constructing TDQI to focus on actual trade from domestic origin; this is in line with de Saint Vaulry (2008). 'Total' trade also enhances the comparability between corresponding trade values.

For all countries, over the period 1962–2016, reported re-imports as a percentage of gross imports is less than 0.5%. To ensure comparability among corresponding trade flows, I assume all countries are equally engaged in re-importing. This allows TDQI to remain comparatively correct, which is what the index aims to achieve as TDQI is used for ranking purposes.

Global re-exports as a percentage of gross exports over 1962–2016 is almost 2.3%. In 2016, as much as 85% of reported global re-exports went through Hong Kong and the US, with 57.2% and 27.9% respectively. According to the UN Statistics Division recommendations, a country should attribute its exports to the final known destination and its imports to the country of origin. Assuming all global re-exports happen through Hong Kong and the US only, and if the rest of the world complies with the recommendations of attribution, TDQI of the imports of Hong Kong and the US would be downwardly biased. TDQI for their exports remains unaffected since re-exports are not included in total exports but included in total imports when they enter a country.

Not all countries, however, are capable of identifying the country of origin or the country of final destination. As discussed below, this requires special treatment for re-exporting countries, especially in the recent three decades in which re-exports as a percentage of global gross exports started to increase (Figure 3).

Figure 3: The share of re-exports in gross exports globally



As stated in Table 1, the third reason for trade data discrepancy is the geographical definition of a trading partner. To this end, I drop any entity or country, which is not considered by its partners as a partner. For example, several large countries do not report any trade with tiny pacific islands, such as Palau and Nauru. To ensure comparability, I drop any observation relating to exporting or importing from them. Appendix B lists the discarded entities/countries. The share of exports reported by the discarded entities in global self-reported exports along the whole period of the study is 1.9%. This leaves us with 1,299,665 bilateral trade observations.

The causes of discrepancy related to timing and conversion to foreign currency are expected to play a minor role. Since the data are annual, the beginning and end of the period are expected balance each other out unless a substantial growth in trade occurred during that period. The UN Statistics Division has clear recommendations on conversion to foreign currency.

The other issue to be addressed is that most countries, as recommended by the UN Statistics Division, report their exports on FOB and their imports on CIF. This recommendation deters us from directly comparing corresponding bilateral trade claims since the importer includes the costs of insurance and transportation in the value of imports and the exporter does not include them in the value of exports. To enhance comparability, I convert all imports from CIF to FOB before constructing TDQI and reconciling trade data. Therefore, all the figures of trade reported and reconciled in this paper are valued on a FOB basis. I use Miao and Fortanier (2017) bilateral relation and time-specific CIF-FOB conversion factors.[13] Miao and

[13] As Miao and Fortanier (2017) bilateral CIF-FOB conversion factors are not available before 1995, I use the average of those conversion factors over the period 1995-2016 to replace the missing values before 1995 for every bilateral relation. This approach is taken as the weighted average of global CIF-FOB conversion factors does not have any discernible time trend. In other words, if CIF-FOB factors between, say, the US and Japan average to 4% over the period 1995-2016, I assume they are also 4.5% for the period 1962-1994.

Fortanier (2017) use a combination of actual shipping and insurance data and gravity models to estimate the conversion factors.

The causes of discrepancy related to corruption and erroneous reporting are discussed in section 6.

For a specific year and trade flow, let N denote the number of trading partners involved in the calculation of TDQI for country i in a given year. For partners $j = 1, \ldots, N$:

M_{ij}: import value of i from partner j, reported by i

X_{ij}: export value of i to j, reported by i

M_{ji}: import value of j from i, reported by j

X_{ji}: export value of j to i, reported by j

$M_{i.}$: the sum of import values of i from all partners, reported by i

$X_{i.}$: the sum of export values of i to all partners, reported by i

The similarity between two corresponding trade values for country i's exports is:

$$S_{ij}^{X} = 1 - \frac{|X_{ij} - M_{ji}|}{X_{ij} + M_{ji}} \qquad [1]$$

The similarity between two corresponding trade values for country i's imports:

$$S_{ij}^{M} = 1 - \frac{|M_{ij} - X_{ji}|}{M_{ij} + X_{ji}} \qquad [2]$$

For all years in our data, I have 453,067 bilateral trade flows for which there are values reported by both partners (906,134 out of 1,299,665 observations) to calculate the similarity measures.[14] The remaining bilateral observations are reported by one side only, which I treat differently, as shown later. The similarity measures are unit-free, and therefore, are comparable over time and across countries without worrying about inflation or trade value. The measures are confined between greater than zero and one. The higher the measure, the higher the similarity. Unlike the method applied by the GTAP, the similarity measure is symmetric (refer to section 4.1 for more discussion).

Now I calculate the relative significance of each similarity measure between country i and each of its trading partners as reflected by the trade shares. Trade shares are unit-free, range from greater than 0 to 1, and sum up to one.

Exports of country i to country j as a share of i's total exports as reported by i:

$$Sh_{ij}^{X} = \frac{X_{ij}}{X_{i}} \qquad [3]$$

Imports of country i from country j as a share of i's total imports as reported by i:

[14] 453,067 X 2=906,134

$$Sh_{ij}^{M} = \frac{M_{ij}}{M_i} \qquad [4]$$

For country i, I calculate the weighted average of the similarities between i's reports on trade and the corresponding values reported by all of its partners j to N. Average similarities for imports of country i:

$$\bar{S}_i^M = \sum_{j=1}^{N} Sh_{ij}^M * S_{ij}^M \; [5]$$

Average similarities for exports of country i:

$$\bar{S}_i^X = \sum_{j=1}^{N} Sh_{ij}^X * S_{ij}^X \; [6]$$

Average similarities for imports and average similarities for exports for each country and year range from greater than 0 to 1. The higher $\bar{S}_i$, the higher the average similarity between country i's available data and the data reported by its partners. The rationale behind the measure of average similarities as a proxy for data quality is straightforward: the more a country's reports on bilateral trade differ from the corresponding reports of its partners, the more likely that country is a low-quality reporter.

Each single similarity measure between two corresponding trade values is included in the calculation of two average similarities: the average similarity for the imports of one country and the average similarity for the exports of the other. Therefore, a country with higher data quality gets its average similarity lowered due to trading with a country with lower quality. As j grows to its maximum N in Equations 5 and 6, all countries are assumed to be equally likely to trade with different combinations of partners with the same average quality. Section 7 argues that this assumption is reasonable.

The main concern about the average similarities measure as a proxy for data quality arises when a country does not report all of its bilateral trade. This would cause the weighted average similarities measure to be incorrect due to the bias of the trade shares used in constructing it: a country might appear to have higher quality because it reported a few values that happened to be in line with partners' reports. Other studies on reconciling international trade data do not take data availability into account when measuring reporter quality (see section 4). To account for the impact of data availability, I first develop a Trade Data Availability Index (TDAI).

For country i in a specific year, let:

m_i^M: number of countries which reported exports to country i while i did not report any imports from them

n_i^M: number of all countries which reported exports to country i

m_i^X: number of countries which reported imports from country i while i did not report any exports to them

n_i^X: number of all countries which reported imports from country i

Imports TDAI for country i is:

$$TDAI_i^M = \left(1 - \left(\frac{m_i^M}{n_i^M}\right)\right) * 100 \qquad [7]$$

Exports TDAI for i:

$$TDAI_i^X = \left(1 - \left(\frac{m_i^X}{n_i^X}\right)\right) * 100 \qquad [8]$$

TDAI ranges from 0 to 100. The higher the TDAI, the more available the bilateral trade data.

SITC recommends that countries report any bilateral trade flow exceeding $1000 US dollars. However, not reporting bilateral trade with a particular partner or partners might be the right practice from country i because such trade did not happen. Assuming all countries are equally susceptible to this scenario, TDAI remains comparatively correct, i.e. biased downwards for all countries equally. I argue that this bias in the measures of data availability, if exists, is quite small as several countries have a TDAI of 100 (see Appendix A).

I use TDAI to penalize a country for not reporting data at all:

$$TDQI_i^X = \bar{S}_i^X * TDAI_i^X \qquad [9]$$

$$TDQI_i^M = \bar{S}_i^M * TDAI_i^M \qquad [10]$$

As average similarities range from above 0 to 1 and TDAI ranges from 0 to 100, the product of the two ranges from 0 to 100: a country that does not report data at all, has a quality of zero. TDQI is a function of available and unavailable data. The higher the index, the higher the quality of a country's trade data.

TDQI is built to simply tell us which partner in any bilateral trade flow is more likely to be more accurate, which serves as the basis of data reconciliation. In any bilateral trade flow, I reconcile the discrepancy in claims by picking the data from the partner with higher TDQI. Following Gehlhar (1996), I do not use the average of claims weighted by quality as a basis of reconciliation. Section 4.3 brings several reasons for not using the quality-weighted average of reports as a reconciliation method.

The most serious issue with using TDQI for reconciling trade data happens due to re-exportation. As discussed earlier, most reported global re-exports happen through Hong Kong and the US with more than half global re-exports going through the former only. The impact of re-exports on TDQI depends on whether the country of origin, the re-exporting country and the country of final destination comply with the UN Statistics Division recommendations of partner attribution. Otherwise, it is not possible to determine who traded what. The comparatively high TDQI for imports and exports for the US (around 90) along the period of the study indicates that the countries of final destination, mostly Canada and Mexico, correctly attribute their imports to the country of origin; perhaps because the US properly informs them as the three countries reconcile their trilateral data before reporting to the SITC.

Hong Kong's TDQI for imports is also high. The problem of using TDQI for data reconciliation for Hong Kong happens at the side of exports as some high-quality reporters wrongly report Hong Kong as the country of origin while the goods actually originate from China. This might occur due to the change in goods ownership once in Hong Kong and before re-exportation (Ferrantino et al., 2012). Since this information might be only available to Hong Kong, it is hard for its partners to identify the correct country of origin.

I believe Hong Kong's exports reports are more accurate than the reports of the importers from it, even when these importers have a higher TDQI for their imports. This position is supported by two facts. First, Hong Kong's TDQI for exports started declining only after it started engaging actively in re-exportation. Second, Hong Kong has very low corruption, as indicated by the Corruption Perceptions Index from Transparency International. Therefore, mis-invoicing, smuggling, and intentional partner misattribution are not common.

Hong Kong's exports are the only case where I do not reconcile the data using the partners' imports but using Hong Kong's exports regardless of the TDQI scores. In other words, this is the only case where I believe a country is right and the rest of the world is wrong.

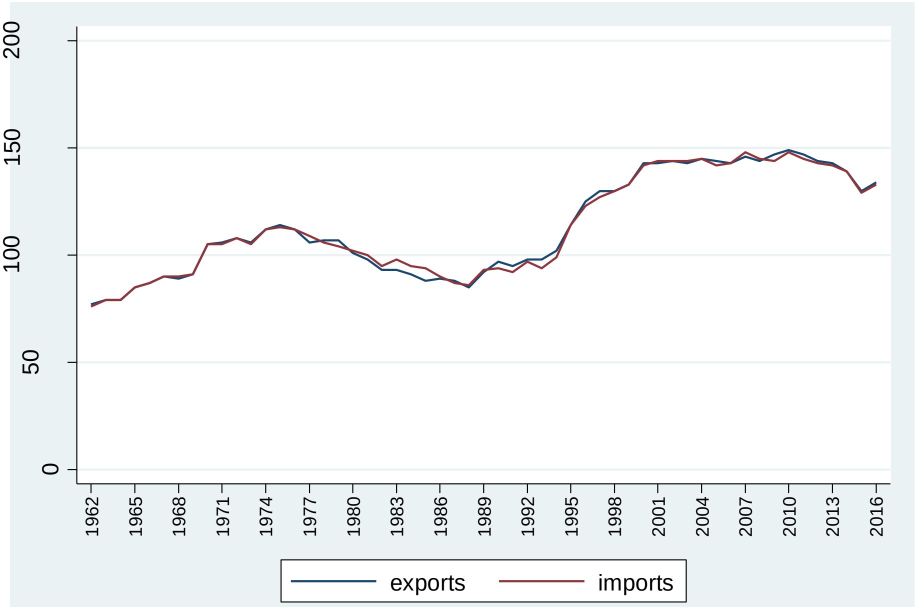
The uncertainty surrounding the trade of re-exporting countries and their partners using unreconciled data is transferred to the reconciled data. TDQI cannot address the role of re-exports given the available information and the ambiguity of partner attribution practices in each country. The uncertainty is extended to countries known to be active in re-exportation despite not reporting re-exports to the SITC, such as the Netherlands and Singapore.

6. Results and discussions

More countries are reporting their trade over time

The number of countries reporting bilateral trade increased over time: 77 countries reported bilateral exports to at least one partner in 1962 compared with 134 in 2016. For imports, the number rose from 76 to 133 (Figure 4).

Figure 4: Number of countries reporting trade to at least one partner

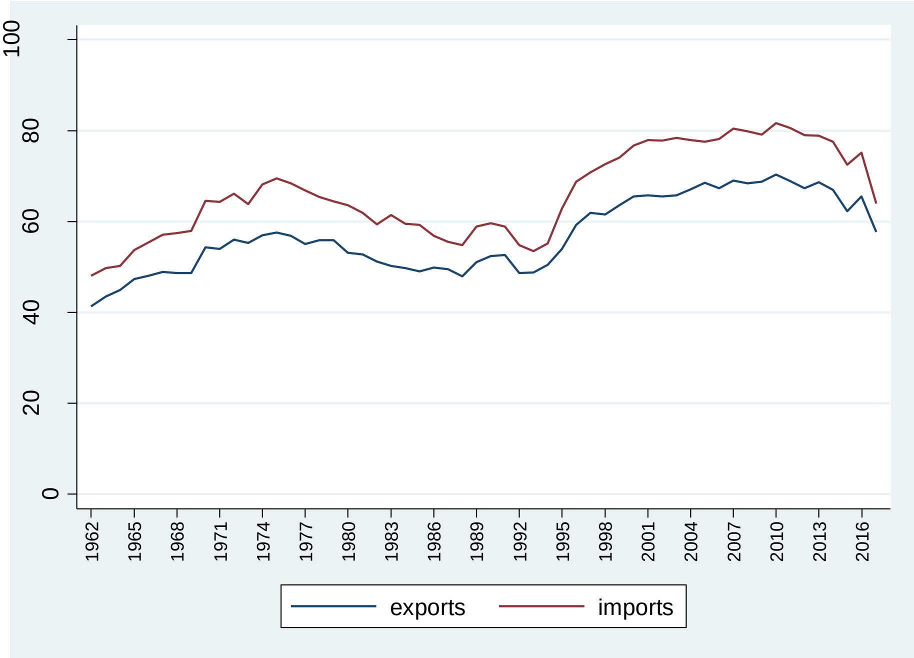


Note: Some of these countries cease to exist; others became countries after 1962.

Trade data are more available for imports than for exports

Figure 5 calculates the unweighted average of the data availability index (TDAI) for all countries in a given year. TDAI has been systematically higher for imports than for exports. There are two possible factors why trade data are more available for imports: First, customs care more about determining the origin of imports than the destination of exports as import tariffs tend to be higher. Second, it is easier to identify the origin of imports than the final destination of exports at the time registration.

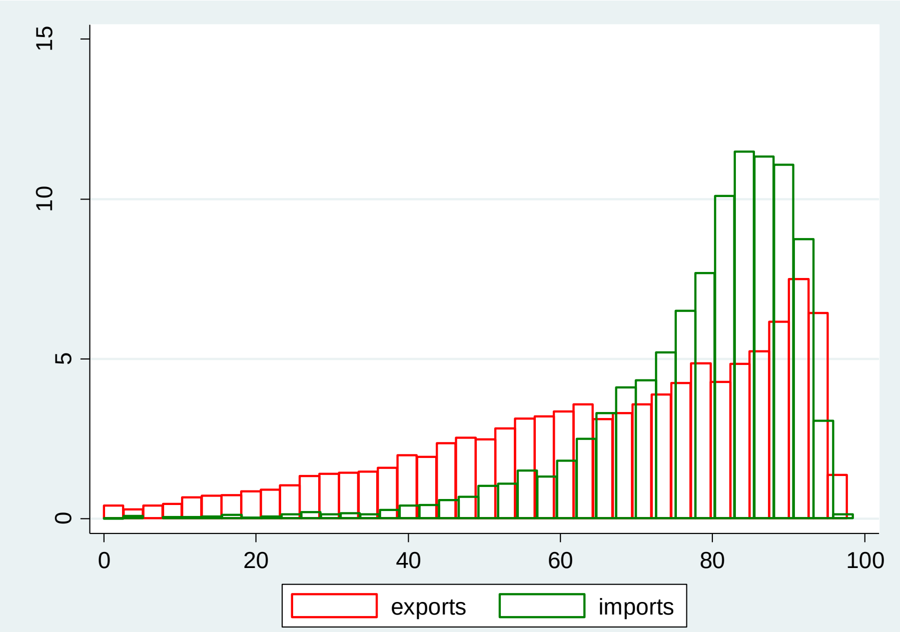
Figure 5: Annual average of TDAI for all countries



Importers' data are slightly more reliable

Figure 6 plots the distributions of the data quality index (TDQI) for imports and exports for all countries and years.

Figure 6: Distribution of TDQI for all countries and years



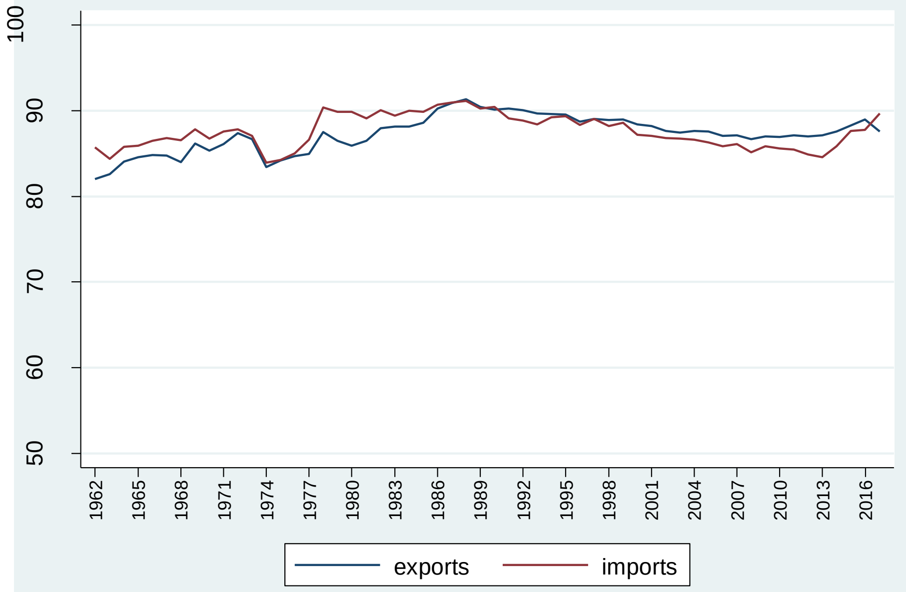
Note: the Figure excludes observations with zero TDQI to improve presentation.

The average of TDQI is 66 for exports and 78 for imports. Average TDQI is higher for imports because both components of TDQI (the average similarities measure and the TDAI) tend to be higher for imports. As a result, it is more likely to pick the claim reported by the importer when reconciling a bilateral discrepancy. Several studies in the literature believe the data reported by importers are more accurate: Feenstra et al. (2005) rely solely on importers' claims in the reconciliation (refer to section 4.4). TDQI confirms that import data are relatively more accurate; however, the index also highlights the numerous cases where the opposite is the case. In fact, throughout the period of the study, exporters' data are used 43% of the time for reconciling bilateral discrepancies, while importers' data are used for reconciling the remaining 57% of the cases.

Countries with low data quality have increased their share in global trade

Figure 7 shows that average TDQI for imports and exports are roughly the same over time if I account for the trade share of each country in global trade. Global trade data quality using weighted average TDQI is in line with the results obtained using the method of lower-higher possibilities ratio explained in section 1 (Figure 2); the similarity serves as a robustness check for TDQI. I conclude that global trade data quality has been mostly deteriorating starting from the early 1990s. The decline in quality coincides with the increase in global trade and re-exportation.

Figure 7: Annual average of TDQI for all countries weighted by trade shares



Note: The trade shares used in calculating the weighted averages are using self-reported data.

The deterioration in global trade data quality does not hold using the unweighted average of TDQI. I calculate the difference between TDQI in the last and the first years a country had trading partners reporting trade with it. For the case of exports, 115 countries had positive differences while 42 had negative differences; 100 compared with 56 for imports. Therefore, the quality of trade data improved in more countries than worsened over time. Combining these numbers with the conclusion from Figure 7 means that the countries which have deteriorated in terms of TDQI have increased their share in global trade over time.

The quality and availability of data are strongly correlated

Appendix A lists TDQI and TDAI for imports and exports for the last year covered in this study (2016). The table is sorted by TDQI for exports from highest to lowest. Table 3 reports the correlations between all four indices. Correlations between TDAI and TDQI for imports and exports are all above 80%. The high correlations between TDQI and TDAI are expected as countries with higher data quality are more likely to report their trade. Similarly, high correlations between import and export measures of data quality and availability indicate that countries which report reliable export data are expected to do the same for their imports, and vice versa.

Table 3: Correlations between TDQI and TDAI for the year 2016

	TDAI exports	TDAI imports	TDQI exports	TDQI imports
TDAI exports	100%			
TDAI imports	94%	100%		
TDQI exports	94%	83%	100%	
TDQI imports	96%	97%	89%	100%

All correlations are statistically significant at 1% (t-test).

Countries on top of the list in Appendix A are mostly developed and transparent. On the other hand, countries with low data quality are, in general, those where corruption is endemic and border controls are mostly non-existent. The findings echo those in Yeats (1990) paper: ‘On the accuracy of economic observations: Do sub-Saharan trade statistics mean anything?’ The answer to Yeats’s question is: no. I add that this is not peculiar to Sub-Saharan countries.

Several factors can explain trade data quality

As discussed in the literature (section 3), several factors might be able to explain why the data reported by certain countries are more reliable than the data reported by others. To investigate this, I regress domestic TDQI on three different possible explanatory variables. First, the size of the economy, proxied by real GDP. Second, by controlling for population, I effectively examine the impact of the level of development as proxied in GDP per capita. Third, public corruption proxied in the Corruption Perceptions Index released annually by Transparency International. The higher the Transparency, the less corrupt the country. As stated in Table 1, in the context of trade data discrepancy, corruption can refer to smuggling, partner misattribution, and mis-invoicing.

Table 4: Possible determinants of trade data quality

Specification	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: TDQI for exports			Dependent variable: TDQI for imports		
Independent variables						
Ln(GDP)	0.10***	0.11***	0.06***	0.07***	0.08***	0.04***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Ln(population)		-0.06***	0.03***		-0.06***	0.02**
		(0.01)	(0.01)		(0.01)	(0.01)
Ln(transparency)			0.21***			0.19***
			(0.01)			(0.02)
Constant	-1.79***	-2.01***	-1.76***	-0.97***	-1.20***	-0.98***
	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Observations	2,078	2,078	2,078	2,078	2,078	2,078
R-squared	0.41	0.42	0.48	0.24	0.26	0.31

The regressions control for country-specific and year effects
Period: 2000–2015
Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

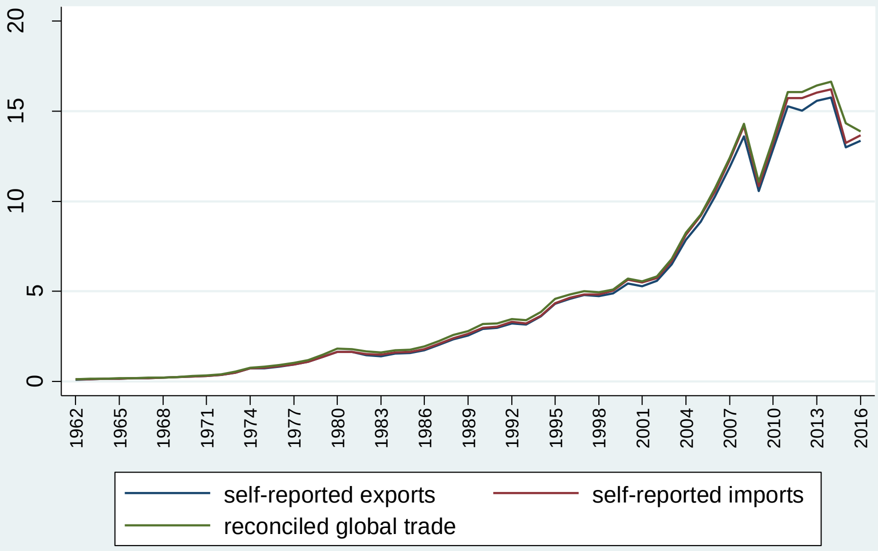
Specifications 1 and 4 suggest that larger economies tend to have higher data quality. Adding the regressor of population in regressions 2 and 5 indicates that after controlling for the size of the economy, countries with a smaller population, tend to have higher data quality. Put differently, countries with higher GDP per capita tend to have higher trade data quality.

Specifications 3 and 6 confirm several earlier studies in the literature, such as Fisman and Wei (2009), showing that lower corruption, i.e. higher public transparency, is associated with higher data quality. As shall be seen later, the fact that low-quality reporters tend to under-report their imports and exports, suggesting smuggling or tariff evasion, supports a causal relationship between corruption and quality. However, more proper identification might be needed to confirm this relationship. Countries in the lowest quartile of average export and import data quality have an average of corruption perceptions index of 25.5 while countries in the highest quartile of data quality have an average of corruption perceptions index of 57.5.

Global trade is under-reported

Following the methodology explained in section 5, I use TDQI to reconcile international trade data. Figure 8 plots global trade using self-reported and reconciled values.

Figure 8: Global trade using self-reported and reconciled data ($US trillions)



Note:
Self-reported exports and imports are summed up to calculate global trade whether the partner reported trade or not. All data are on FOB basis.

Almost along the whole period, global trade using reconciled data is higher than that using self-reported data: global aggregate trade is under-reported. Over the last five years combined, global trade using exporters' data was 6.3% lower than that using reconciled data; while 3.3% lower using importers' data (4.8% difference on average between self-reported and reconciled data). As shown next, the differences are because low-quality reporters under-report their trade.

Low-quality reporters under-report their imports and exports

To capture how reconciled and self-reported data differ on country-level, I calculate the ratio of the two for every country for the period 1962-2016. The ratio is calculated for a country's trade with the rest of the world combined as follows:

Reconciled-unreconciled ratio = (reconciled-unreconciled)/(reconciled + unreconciled)*100 [11]

The ratio is symmetric to differences in reconciled and unreconciled values and is confined between 0 and 100. Figures 9 and 10 report the distribution of the ratios for reporters with high, medium, and low quality (tertiles of countries by data quality).

Figure 9: Distribution of reconciled-unreconciled ratio for exporters

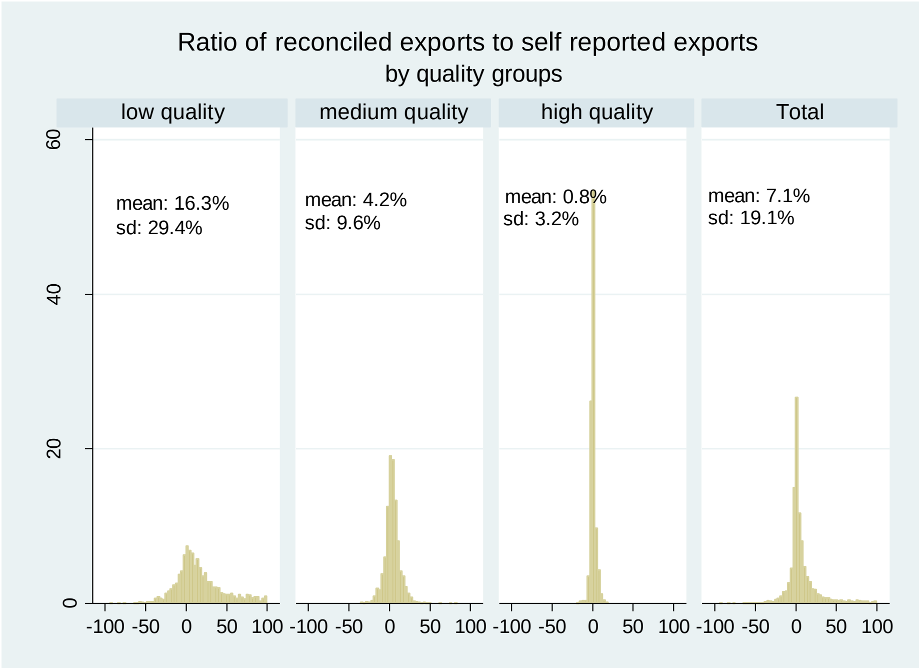
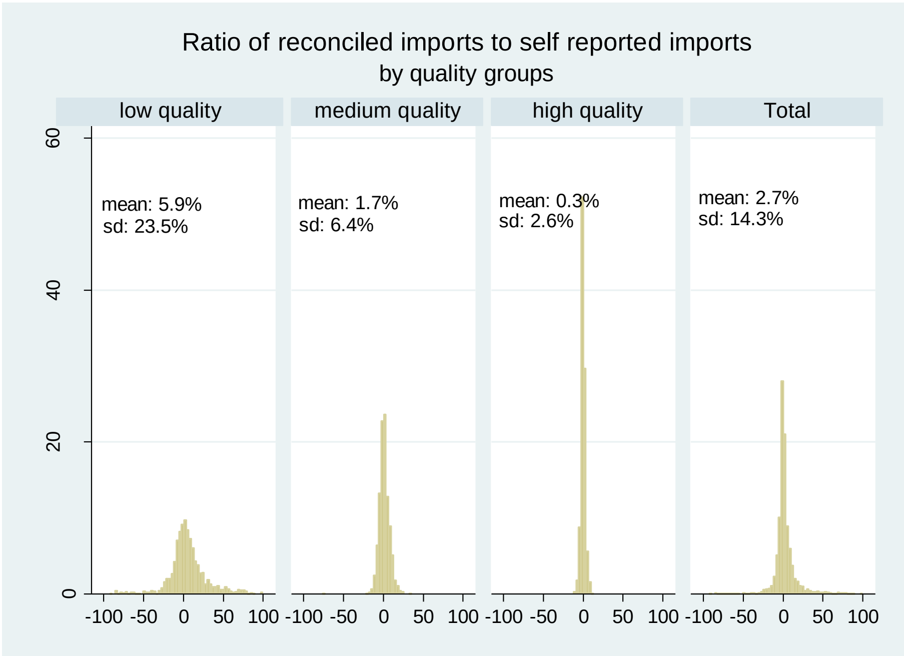


Figure 10: Distribution of reconciled-unreconciled ratio for importers



The histograms show that the deviations of reconciled from unreconciled data for high-quality reporters are relatively small for two reasons. First, in trade with low-quality reporters, the data are picked from the country with higher quality: self-reported figures are used for reconciliation. Second, when reconciling the data between high-quality reporters, the differences in reports tend to be smaller than the differences between countries with low-quality data: high-quality reporters tend to report similar values of trade with each other. The low standard deviations indicate homogeneity among high-quality reporters in terms of the similarity between reconciled and unreconciled data.

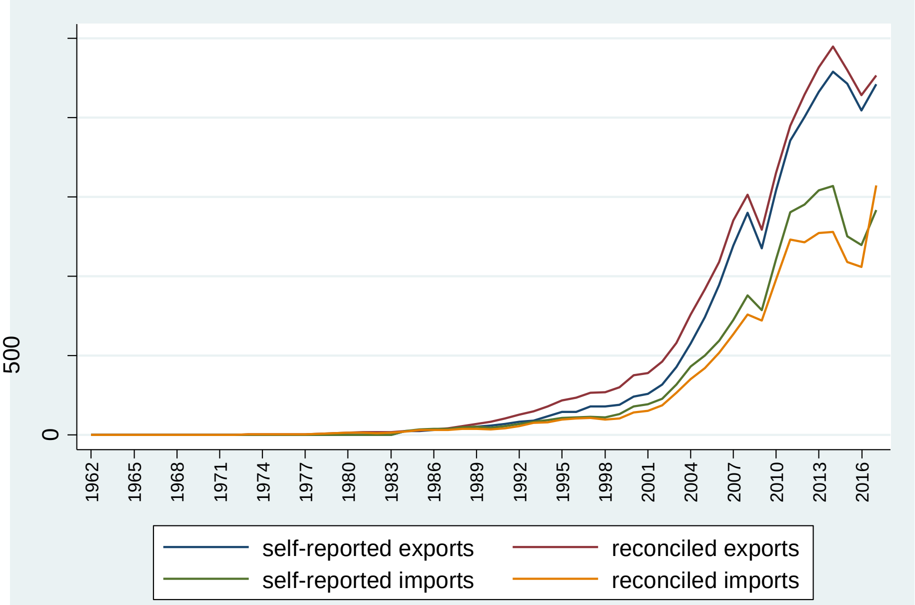
Low-quality reporters tend to under-report their trade as the positive means of the distributions suggest, which adds more evidence that corruption can explain a share of the discrepancies as this indicates smuggling or under-invoicing. The large standard deviations, on the other hand, suggest that there is a significant idiosyncratic component in the discrepancies, which implies erroneous reporting.

China tends to over-report imports and under-report exports

According to SITC data in 2016, the largest two traders in the world were China followed by the US; this holds using self-reported or reconciled data. I now take a look at these two countries.

China's exports are generally under-reported, although the gap has been closing recently (Figure 11). This is not related to the fact that a share of China’s exports goes through Hong Kong since the plotted data are for China’s trade with the rest of the world combined. Ferrantino, Liu and Wang (2012) show strong statistical support for Chinese Value Added Tax evasion for the commodities traded directly between the US and China.

Figure 11: China’s trade with the rest of the world ($US billions)



Note: China started reporting data to SITC in 1984; reconciled data starts from 1962.

The less investigated issue in the literature is China’s data discrepancy on the side of imports. China's reconciled imports are consistently smaller than self-reported throughout most of the period of the study. China tends to over-report its imports from many high-quality reporters. However, this seems to have changed in 2016. Table 5 shows the difference in claims on China’s imports for the year 2014.

Table 5: China’s claims on imports vs its partners’ claims on exports to it in billions of dollars in 2014

Partner	partner's TDQI for exports	China's TDQI for imports	partner's claim	China's claim	reconciled bilateral trade (FOB)	more reliable reporter	partner claim/China's claim
Korea, Rep.	91	79	145.3	190.1	145.3	Korea, Rep.	0.76
Japan	91	79	126.2	162.8	126.2	Japan	0.78

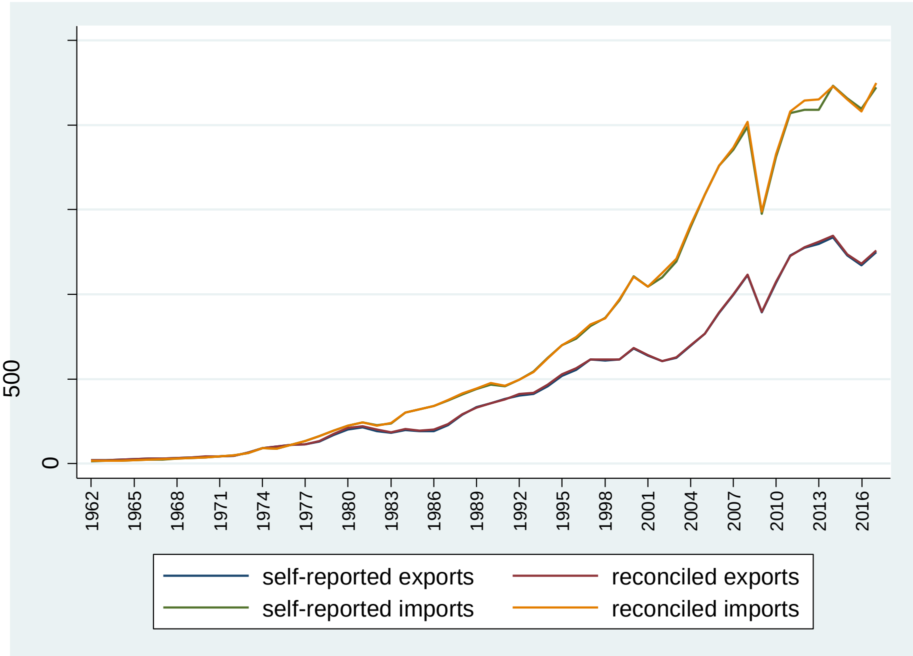
United States	92	79	115.2	159.8	115.2	United States	0.72
Germany	95	79	99.2	105	99.2	Germany	0.94
Australia	82	79	70.1	97.7	70.1	Australia	0.72
Malaysia	86	79	28.2	55.7	28.2	Malaysia	0.51
Brazil	87	79	40.6	51.7	40.6	Brazil	0.79
Saudi Arabia	46	79	7.1	48.5	46.1	China	0.15
South Africa	82	79	8.7	44.6	8.7	South Africa	0.19
Russian Federation	74	79	37.4	41.6	39.5	China	0.95

Note: The list includes the top 10 exporters to China (sorted by China's reports).

US self-reported and reconciled trade are similar

US trade data remains relatively unchanged after reconciliation since US TDQI is comparatively high: US data are used for reconciliation in most cases. In the cases where US partners' data are used for reconciliation, the differences in claims are rather small (Figure 12).

Figure 12: US trade with the rest of the world ($US billions)



Using reconciled trade is alike for understanding international trade

Bahmani-Oskooee et al. (2013) show how substantial data discrepancies can mean a study that uses data reported by a particular partner would reach a different conclusion if it used the data reported by the other (discussed in section 1). Using reconciled data in a measure that depends on trade in its construct makes a non-trivial difference as well. For illustration, I calculate trade openness for every country in 2016, once using reconciled data and once using self-reported data.[15] Table 6 presents the findings, were high, medium and low quality refer

[15] Trade openness is defined as (exports + imports)/GDP. GDP data are in real USD and are retrieved from the World Bank. I break the reporters into three groups (high, medium, and low-quality) based on the average quality of their imports and exports.

to the tertiles of countries in terms of average TDQI for their imports and exports. The total sample size is 133 countries: the subsamples are 45 or 44 each.

Table 2 6: Trade Openness calculated using reconciled and unreconciled data

reporter quality	data source	average trade openness
high-quality reporters	unreconciled	59%
	reconciled	59%
medium-quality reporters	unreconciled	67%
	reconciled	71%
low-quality reporters	unreconciled	43%
	reconciled	57%

The less reliable third of countries are on average 14% more open to trade than thought in the literature, which largely depended on self-reported figures.[16] Trade openness for countries with higher data quality does not change materially using reconciled or unreconciled data.

7. Robustness checks

A fundamental assumption for TDQI is that all countries are equally likely to trade with different combinations of partners with the same average of quality. However, what if a country trades extensively with low-quality reporters? It would appear as a low-quality reporter itself (low TDQI). From the Melitz (2003) model of trade, extended to a gravity framework by Helpman, Melitz and Rubinstein (2008), firms have to pay a fixed cost to export. Since this fixed cost differs across trading partners and is often higher in more 'difficult' destination countries, it is reasonable to expect that firms in a country like Iceland may not be able to incur the fixed cost of exporting to a country like Mali, while firms in the US might be able to. This means that we would observe the US trading with Mali, while Iceland would not. The TDQI for the US would then be lower than Iceland's if Mali was a low-quality reporter.

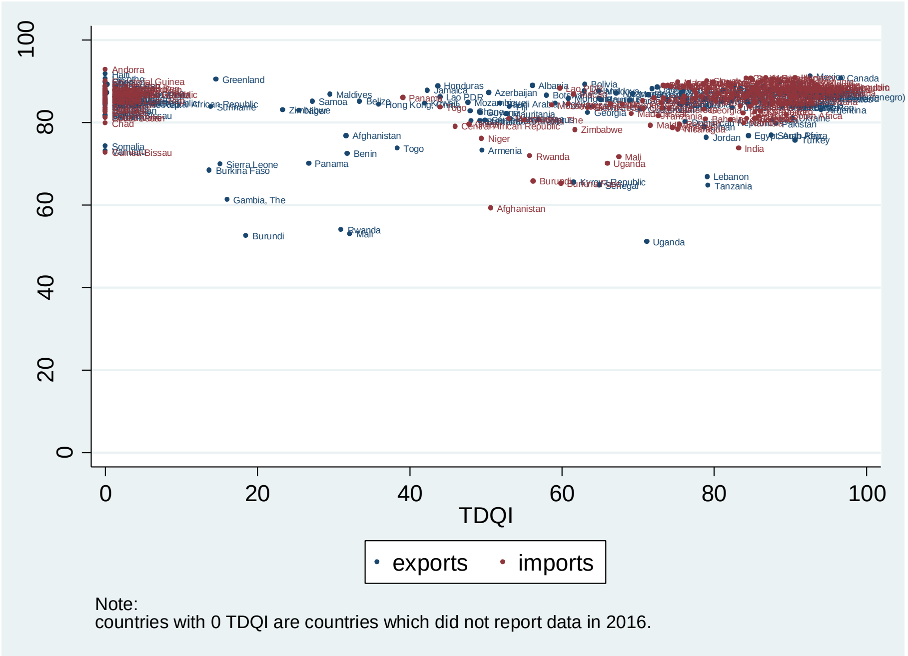
So how reasonable is our assumption that all countries are equally likely to trade with different combinations of partners with the same average of quality? To answer this question, I look at domestic TDQI for each trade flow for each country in the year 2016 and compare it with the weighted average of TDQI for its trading partners. Figure 13 plots domestic TDQI for importers against the weighted average of their partners' TDQI for exports, and vice versa. The plot shows that the weighted averages of partners' quality for every country are clustered in a particular region (clustered horizontally). That is, whether countries are high-quality or low-quality reporters, the weighted average of their partners' TDQI is relatively the same, suggesting our assumption is reasonable.

But how about the few countries with low partner TDQIs? As the plot shows, the countries with low partner data quality are mostly countries with small shares in global trade. Most importantly, what are the implications of having some countries that have lower weighted averages of their partners' data quality? Back to our example of the US and Iceland and their trade with Mali. This means this study is more likely to pick Iceland-US trade data from Iceland, instead of the US. As shown in Figures 9 and 10, as both countries are high-quality

[16] From Table 6: 57% – 43%=14%

reporters, the discrepancies between their claims are small, which explains why US total trade with the rest of the world is roughly the same using reconciled or unreconciled data (Figure 12).

Figure 13: Reporter data quality and the corresponding weighted average of partners' quality (2016)



The same applies to years other than 2016. Figure 14 plots the distributions of the weighted average of partners' data quality for different brackets of exporter quality. The similarity in all three distributions suggests that all exporters, regardless of their export data quality, have similar weighted averages of partners' data quality. Figure 15 confirms the same for the case of importers.

Figure 14: Weighted average of partners' data quality conditional on exporters' data quality (all years)

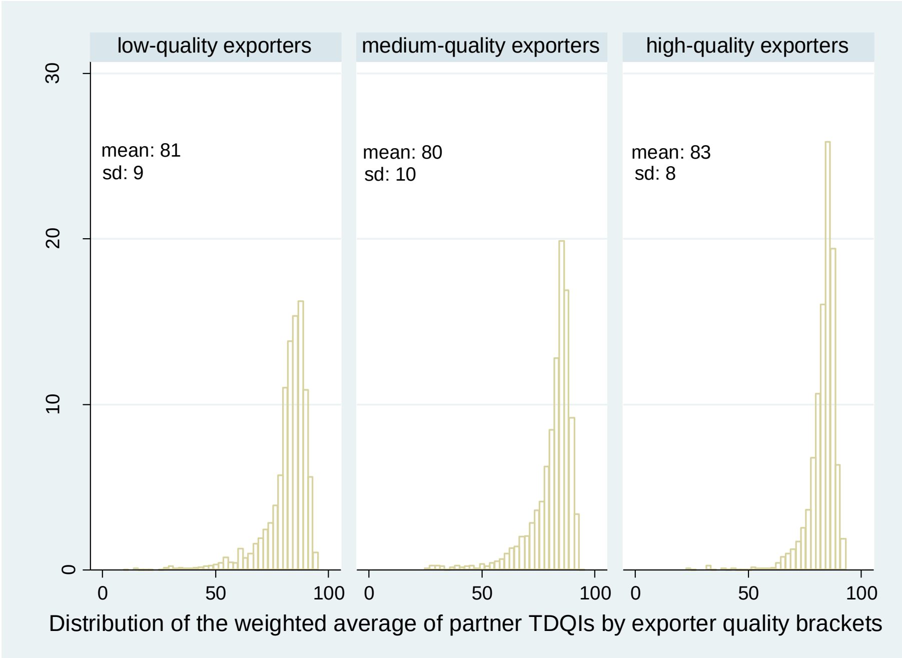
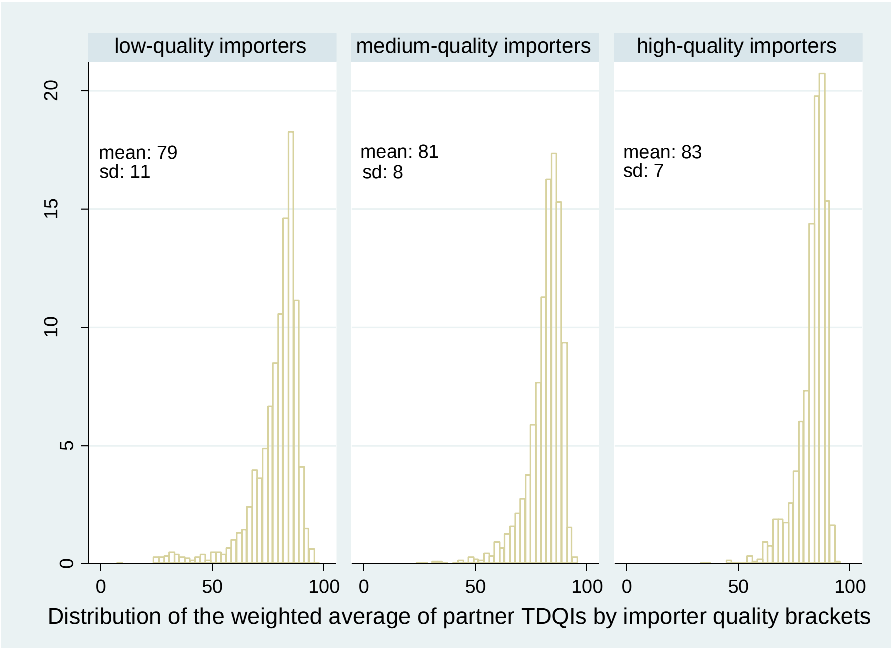


Figure 15: Weighted average of partners' data quality conditional on importers' data quality (all years)



8. Conclusion

When two friends, John and Tony tell different stories about what happened between them, it is hard to find out quickly which claim is more reliable. If a third friend, Sarah, tells me she usually agrees with Tony but not with John, evidence starts to mount against John as now two people have independently claimed he is not telling the truth. If I follow the same logic with many more friends, I should be able to rank everybody in terms of their reliability. When I want to write the history of what happened between each pair of these friends, I should rely on the stories told by the more reliable half of them.

I use this logic to rank countries in terms of their trade data reporting quality and later reconcile trade data by picking the value reported by the more reliable partner in every bilateral relationship.

I confirm the wide belief that importers' data tend to be more reliable. I further show that low-quality reporters tend to under-report their imports and exports, giving evidence of under-invoicing and smuggling. While China tends to under-report its exports and over-report its imports, the US self-reported and reconciled data are roughly the same.

Using reconciled trade is vital for understanding international trade. To illustrate, I calculate trade openness for every country in 2016, once using the data reconciled here and once using self-reported data. Low-quality reporters are, on average, 14% more open to trade using reconciled data.

The results are made available for public access.

References

– ,– "International Merchandise Trade Statistics–Concepts and Definitions," New York: Department of Economic and Social Affairs, Statistical Division, United Nations, No. 52, Series M (2011).

Bahmani-Oskooee, Mohsen, Scott Hegerty, and Hanafiah Harvey. "Exchange-rate sensitivity of commodity trade flows: Does the choice of reporting country affect the empirical estimates?." *The Journal of International Trade & Economic Development* 22, no. 8 (2013): 1183-1213.

De Saint Vaulry, Alix. "Base de données CHELEM-commerce international du CEPII" *Centre d'études prospectives et d'informations internationals*, (2008).

Feenstra, Robert C., Robert E. Lipsey, Haiyan Deng, Alyson C. Ma, and Hengyong Mo. "World trade flows: 1962-2000." NBER working paper No. w11040, 2005.

Ferrantino, Michael J., Xuepeng Liu, and Zhi Wang. "Evasion behaviors of exporters and importers: Evidence from the US-China trade data discrepancy." *Journal of International Economics* 86, no. 1 (2012): 141-157.

Fisman, Raymond, and Shang-Jin Wei. "Tax rates and tax evasion: evidence from "missing imports" in China." *Journal of Political Economy* 112, no. 2 (2004): 471-496.

Fisman, Raymond, and Shang-Jin Wei. "The smuggling of art, and the art of smuggling: Uncovering the illicit trade in cultural property and antiques." *American Economic Journal: Applied Economics* 1, no. 3 (2009): 82-96.

Fortanier, Fabienne. (2016). "Towards merchandise trade statistics without asymmetries." *OECD, The Statistics Newsletter* (Issue No. 64, March 2016).

Gaulier, Guillaume, and Soledad Zignago. “BACI: International Trade Database at the Product-Level (the 1994-2007 version).” *Centre d'Etudes Prospectives et d'Informations Internationales*, (2010).

Gehlhar, Mark. “Reconciling bilateral trade data for use in GTAP.” *GTAP Technical Papers* no. 11. (1996).

Gujarati, Damodar. “*Basic econometrics*.” Mc-Graw Hill, 2004.

Guo, Dong. “Mirror Statistics of International Trade in Manufacturing Goods: The Case of China. Research and statistics branch” *Vienna, United Nations Industrial Development Organization, working paper no. 19/2009* (2010).

Helpman, Elhanan, Marc Melitz, and Yona Rubinstein. “Estimating trade flows: Trading partners and trading volumes.” *The quarterly journal of economics* 123, no. 2 (2008): 441-487.

Makhoul, Basim, and Samuel M. Otterstrom. “Exploring the accuracy of international trade statistics.” *Applied Economics*30, no. 12 (1998): 1603-1616.

Melitz, Marc J. “The impact of trade on intra-industry reallocations and aggregate industry productivity.” *Econometrica* 71, no. 6 (2003): 1695-1725.

Miao, Guannan, and Fabienne Fortanier. “Estimating Transport and Insurance Costs of International Trade”, OECD Statistics Working Paper 2017/04 (2017); http://dx.doi.org/10.1787/8267bb0f-en

Simola, Heli. “Russian import statistics in the mirror of world exports.” BOFIT. *Institute for Economies in Transition working paper no.* 2, 2012.

Wang, Zhi, Mark Gehlhar, and Shunli Yao. “A globally consistent framework for reliability-based trade statistics reconciliation in the presence of an entrepôt.” *China Economic Review* 21, no. 1 (2010): 161-189.

West, Loraine A. “*Reconciling China's trade statistics*. International Programs Center, Population Division.” US Bureau of the Census IPC Staff Paper no.76, 1995.

Yeats, A. J. (1990). “On the accuracy of economic observations: Do sub-Saharan trade statistics mean anything?” *The World Bank Economic Review, 4*(2) (1990): 135-156.

APPENDIX A

Table A1: TDAI and TDQI for imports and exports for the year 2016

Reporter	TDAI exports	TDAI imports	TDQI exports	TDQI imports
Canada	100	100	97	88
Poland	100	100	95	82
Austria	100	100	94	85
Italy	100	100	94	93
Argentina	99	97	94	89
New Zealand	98	100	94	85
Germany	100	100	94	90
Spain	100	100	94	91
United States	100	100	93	93
Bulgaria	98	97	93	90
France	100	100	93	85
Mexico	98	100	93	87
Japan	100	100	92	85
Sweden	100	100	92	91
Finland	99	100	92	87
Korea, Rep.	100	100	92	86
Romania	100	96	91	93
Indonesia	100	99	91	87
Czech Republic	100	100	91	84
Turkey	100	99	91	89
Switzerland	100	99	91	85
Portugal	100	99	91	91
Slovenia	99	100	90	86
United Kingdom	100	100	90	94
Sri Lanka	99	100	90	85
Hungary	99	98	90	91
Ukraine	99	99	90	86
Thailand	100	100	90	88
Greece	99	97	89	86
Denmark	99	99	89	94
Australia	99	100	89	93
Brazil	100	99	89	91
Ireland	99	100	89	77
Slovak Republic	99	100	89	79
Chile	96	97	89	90
China	100	100	88	86
Pakistan	98	98	88	75
India	100	100	88	83
Serbia	97	99	88	79

South Africa	99	100	88	89
Netherlands	100	100	87	86
Morocco	98	98	87	87
Luxembourg	99	100	86	86
Peru	92	100	85	92
Russian Federation	100	100	85	89
Egypt, Arab Rep.	95	95	84	80
Qatar	96	87	84	71
Estonia	98	100	84	85
Malaysia	98	98	84	85
Norway	100	100	83	80
Belgium	100	100	83	92
Philippines	100	98	83	81
Israel	99	99	83	79
Latvia	100	98	83	85
Tunisia	89	96	82	86
Croatia	93	100	82	94
Macedonia, FYR	88	98	81	75
Belarus	96	100	81	89
Tanzania	90	99	79	73
Lebanon	94	99	79	87
Jordan	90	96	79	82
Myanmar	94	96	79	78
Oman	89	92	79	48
Lithuania	99	100	76	88
Colombia	94	100	76	93
Dominican Republic	81	100	76	91
Bosnia and Herzegovina	83	100	76	74
Nepal	84	98	76	82
Vietnam	86	87	76	77
Uruguay	91	99	76	82
Namibia	82	94	75	84
Malta	94	98	74	73
Costa Rica	86	100	74	90
Singapore	100	100	73	86
Algeria	83	96	73	88
Cambodia	89	97	72	78
Ecuador	88	97	71	88
Uganda	82	97	71	66
Guatemala	78	97	70	84
Madagascar	80	97	69	69
Paraguay	90	96	69	73
Nicaragua	76	95	67	75
Brunei	72	93	65	76

Mauritius	74	100	65	88
Senegal	84	94	65	73
Moldova	86	100	65	77
Montenegro	77	99	64	79
Georgia	78	90	63	79
Bolivia	74	99	63	76
Iceland	80	97	63	74
Kyrgyz Republic	78	99	62	61
Mongolia	66	99	61	90
Nigeria	70	97	59	78
Botswana	88	99	58	93
Cyprus	95	100	57	82
Albania	78	100	56	88
Kazakhstan	88	97	53	83
Mauritania	64	91	53	53
Saudi Arabia	89	94	52	84
Azerbaijan	85	92	50	76
Cameroon	70	96	50	74
Armenia	73	99	49	74
Bahrain	78	99	49	79
Guyana	69	96	49	85
United Arab Emirates	98	99	48	82
Ghana	75	96	48	73
Mozambique	78	100	48	59
Lao PDR	56	76	44	60
Honduras	65	92	44	68
Jamaica	54	83	42	72
Kuwait	88	99	42	84
Togo	53	94	38	44
Hong Kong, China	98	100	36	91
Mali	56	92	32	67
Benin	70	94	32	66
Afghanistan	35	69	32	51
Rwanda	79	97	31	56
Maldives	41	94	30	72
Panama	72	73	27	39
Niger	60	91	25	49
Zimbabwe	60	95	23	62
Burundi	46	89	18	56
Gambia, The	44	93	16	55
Sierra Leone	29	90	15	63
Suriname	65	91	14	80
Burkina Faso	64	98	14	60
Central African Republic	24	82	6	46

Iraq	11	0	3	0
El Salvador	79	100	0	85
Kenya	0	0	0	0
South Sudan	0	0	0	0
Haiti	0	0	0	0
Guinea	0	0	0	0
Bhutan	0	0	0	0
Turkmenistan	0	0	0	0
Yemen	0	0	0	0
Liberia	0	0	0	0
Djibouti	0	0	0	0
Somalia	0	0	0	0
Syrian Arab Republic	0	0	0	0
Libya	0	0	0	0
Korea, Dem. Rep.	0	0	0	0
Bangladesh	0	0	0	0
Comoros	0	0	0	0
Malawi	0	0	0	0
Chad	0	0	0	0
Congo, Rep.	0	0	0	0
Cuba	0	0	0	0
Tajikistan	0	0	0	0
Uzbekistan	0	0	0	0
Cote d'Ivoire	0	0	0	0
Zambia	0	0	0	0
Sudan	0	0	0	0
Congo, Dem. Rep.	0	0	0	0
Angola	0	0	0	0
Venezuela	0	0	0	0
Guinea-Bissau	0	0	0	0
Papua New Guinea	0	0	0	0
Gabon	0	0	0	0
Iran, Islamic Rep.	0	0	0	0
Lesotho	0	0	0	0

Note: Some countries have 0 TDQI because they did not report data in 2016.

APPENDIX B

The list of dropped reporters along the whole period of the study is:

Aruba, Anguilla, Netherlands Antilles, Antigua and Barbuda, Bahamas, The, Bermuda, Barbados, Cook Islands, Cape Verde, Cayman Islands, Dominica, Faeroe Islands, Guadeloupe, Grenada, Kiribati, Saint Kitts-Nevis-Anguilla-Aruba, Saint Kitts and Nevis, Saint Lucia, Montserrat, Martinique, Mayotte, New Caledonia, Other Asia not elsewhere specified, Palau, Occupied Palestinian Territory, French Polynesia, Reunion, Solomon Islands, Saint Pierre and Miquelon, Sao Tome and Principe, Seychelles, Turks and Caicos Islands, Tonga, Trinidad and Tobago, Tuvalu, Saint Vincent and the Grenadines, Virgin Islands (U.S.), Wallis and Futura Islands, Eritrea, and Ethiopia (excludes Eritrea). Note that we include Ethiopia (including Eritrea).

The list of dropped partners along the whole period of the study is:

Anguilla, Netherlands Antilles, Antigua and Barbuda, Bahamas, The, Bermuda, Barbados, Bunkers, Curaçao, Cayman Islands, Dominica, Free Zones, Gibraltar, Guadeloupe, Grenada, Saint Kitts and Nevis, Saint Lucia, Montserrat, Martinique, Other Asia not elsewhere specified, Sao Tome and Principe, Turks and Caicos Islands, Trinidad and Tobago, Unspecified, St. Vincent and the Grenadines, British Virgin Islands, Marshall Islands, Special Categories, Saint Pierre and Miquelon, Holy See, Aruba, American Samoa, Antarctica, French Southern Territories, Bouvet Island, Cocos (Keeling) Islands, Cook Islands, Cape Verde, Christmas Island, Western Sahara, Falkland Island, Faeroe Islands, Federated State of Micronesia, Guam, Heard Island and McDonald Island, British Indian Ocean Territories, Kiribati, Northern Mariana Islands, Mayotte, New Caledonia, Norfolk Island, Niue, Nauru, Pitcairn, Palau, Occupied Palestinian Territory, French Polynesia, Reunion, South Georgia and the South Sandwich Islands, Saint Helena, Solomon Islands, San Marino, Seychelles, Tokelau, Tonga, Tuvalu, United States Minor Outlying Islands, Virgin Islands (U.S.), Wallis and Futuna Islands, Neutral Zone, US Miscellaneous Pacific Islands, Saint Kitts-Nevis-Anguilla-Aruba, Former Panama Canal Zone, Ryukyu Island, Pacific Islands, Saint Barthélemy, Bonaire, Saint Maarten (Dutch part), British Antarctic Territory, Sikkim, Eritrea, and Ethiopia (excludes Eritrea). Note that we include Ethiopia (including Eritrea).