

## Trade Estimates

International trade plays an important role in some of the most prominent studies aimed at estimating the scale of illicit financial flows. A case in point is one of the first estimates of illicit financial flows in the book by Raymond Baker (2005), who in 2006 founded Global Financial Integrity (GFI), an NGO which is well-known for its own estimates. Mainly on the basis of around 550 interviews with corporate employees, Baker (2005) estimated that more than USD 539 billion flows out of developing and transitional economies each year due to a combination of commercial tax evasion, fraud in international trade, drug trafficking, and corruption; and that international trade abuses account for the largest part. These abuses are due both to criminal (illegal arms trade, smuggling) and illicit (mispricing between unrelated and abusive transfer pricing between related companies, and ‘fake transactions’) activities. On the basis of Baker (2005), Christian Aid (2008) estimates the amount of tax revenue lost to developing countries annually through these two techniques, transfer mispricing and false invoicing, at 157 billion USD.

In contrast with the pioneering estimates by Baker (2005) based partly on interviews (direct, if anecdotal, evidence), most of the more recent approaches to estimation recognise that it is not possible to observe illicit financial flows directly and estimate them indirectly. These approaches are based on the little available economic data that is available about activities potentially related to illicit financial flows. Specifically, the methodologies often focus on exploiting anomalies in the data that may arise from the process of hiding the flows (but can also arise for other reasons—a critical point to which we return when we critically evaluate the estimation methodologies).

The most prominent approaches focus on anomalies in the current account (via misreported or mispriced trade, discussed in this chapter) and in the capital account (through partially unrecorded capital movements, discussed in the next chapter). Some of the authors combine the two approaches, including GFI reports covering most developing countries, and Ndikumana and Boyce who focus on African countries (in these cases we discuss their trade and capital components in the respective chapters). Both

of these group of approaches are reviewed in the few existing reviews of illicit financial flows such as, for example, the edited volume of Reuter (2012) and Johannesen & Pirttilä (2016), who mostly use the term capital flight (a related but distinct concept, which has earlier generated quite a lot of research interest, e.g. Cuddington, 1987, Dooley, 1988, Collier, Hoeffler, & Pattillo, 2001, Beja, 2005). In these reviews and elsewhere, both of these approaches have been subjected to critical evaluations and we discuss them alongside the relevant literature.

Within the trade estimates we distinguish three broadly defined groups of approaches, roughly according to the data used, and we discuss them one by one. (This classification is not perfect with some studies fitting in more, or none, of these groups, but we believe that it does help us to enhance the discussion.) The first subchapter examines estimates based mostly on country-level data (i.e. for each country or country pair we have only one piece of information available), although some of the reviewed studies use more detailed data. The second subchapter discusses studies based on commodity-level trade data. Each of these first two subchapters deals with a specific methodological approach as well. The first subchapter focuses on so called trade mirror statistics, while the second subchapter investigates studies looking at abnormal prices. The studies discussed in these two subchapters, and in the first one in particular, have been subject to evaluation by other researchers, such as Hong & Pak (2017) or Nitsch (2017), that have pointed out the methodological weaknesses such as unrealistic assumptions in these studies (and we discuss critical observations from these evaluations below). Emerging partly as a response to these criticisms, the final subchapter discusses the most recent and, from the point of view of rigour, most promising studies. These studies rely on only recently available detailed transaction-level data. This kind of detailed data is so far available only for a limited number of countries, although their number is increasing.

We thus provide a broad classification of the existing trade data-based estimates of illicit financial flows, of which we provide an overview in Table 2.1. In addition to prevailing method and level and sources of data, Table 2.1 includes examples of recent studies as well as our brief evaluation of the reliability of methodology and availability of estimates in terms of country coverage. This is of course only a quick bird's eye view—the individual studies covered differ from each other within the subchapters and each study has its own pros and cons, including its suitability for estimation of scale of illicit financial flows or for audit purposes.

**Table 2.1.** Broad classification of trade data-based estimates of illicit financial flows

Sub-chapter	Prevailing level of data	Prevailing sources of data	Prevailing method	Recent examples	Reliability of the methodology	Availability and country coverage
2.1	Country (and commodity)	IMF (and UN Comtrade)	Mirror trade statistics	GFI's Spanjers & Salomon (2017)	Not recommended as estimates of scale, perhaps suitable for preliminary identification for audit purposes	Excellent and most of the world
2.2	Commodity (and transaction)	Country-specific (and UN Comtrade)	Abnormal prices	Chalendard, Raballand, & Rakotoarisoa (2019)	Not recommended as estimates of scale, perhaps suitable for preliminary identification for audit purposes	Excellent and most of the world
2.3	Transaction	Country-specific	Systematic differences between intra-firm and arm's length prices	Davies, Martin, Parenti, & Toubal (2017)	Good (estimates of scale and for audit purposes)	Limited and only a few countries

Source: Authors

## 2.1. Country-level Trade Estimates: Mirror trade statistics

### 2.1.1. Overview

The early estimates of illicit financial flows on the basis of trade data (which happen to be also some of the first estimates of illicit financial flows more generally) are based on aggregate country-level international trade data. Most of these studies capture the quantity of illicit flows by contrasting what a country claims it imported from (or exported to) the rest of the world with what the rest of the world states it exported to (or imported from) that given country. The development of this method of—what others call and we are going to call—mirror trade statistics, which compares import and export data for the same trade flow, goes back to Morgenstern (1950, 1974) and Bhagwati (1964, 1974) and was applied, for example, by Beja (2008) for China and by Berger & Nitsch (2012) for five largest importers. On the one hand, we include all of approaches using the logic of the mirror trade statistics method in this subchapter, although some of them, such as Berger & Nitsch (2012) or Ndikumana (2016), have been applied at the commodity level (and not at the country level as the name of the subchapter suggests). On the other hand, we do not discuss in detail literature related specifically to tariff evasion, as pioneered by Bhagwati (1964) and later developed, for example, by Javorcik & Narciso (2008).

We focus in our description on perhaps the two most prominent mirror trade statistics approaches. These are those by the organisation GFI, and by the duo of authors Ndikumana and Boyce. Both combine a trade-related IFF component with estimates based on capital-account data that we discuss in the next chapter. They both assume that traders deliberately misreport trade through faking invoices or other forms of mis-invoicing and we discuss these two in detail below. Before that we briefly explore various motivations why trading partners might mis-invoice the trade volumes or prices. A recent overview of these various motives is provided, for example, by Kellenberg & Levinson (2016) and Nitsch (2017) and they range from tariff evasion to tax evasion. In Table 2.2, together with Nitsch (2017), we distinguish four types of trade mis-invoicing measured along two dimensions—first whether the trade flows are exports or imports and, second, whether these flows are overinvoiced or underinvoiced. We agree with Nitsch's (2017) argument that the broad range of incentives to misreport trade provides a challenge for the empirical assessment of its scale and in the this and the following two subchapters we review how various researchers and methods have dealt with this challenge so far.

**Table 2.2.** Types of trade mis-invoicing

	Overinvoicing	Underinvoicing
Export	Export overinvoicing To take advantage of export subsidies—Celasun & Rodrik (1989a), Celasun & Rodrik (1989b)	Export underinvoicing To evade export restrictions, to circumvent trade restrictions (a misclassification of products or a misdeclaration of the final destination of a shipment) or to avoid product taxes—Fisman & Wei (2009), Kellenberg & Levinson (2016), Kee & Nicita (2016)
Import	Import overinvoicing To misclassify other imports (underreport some imports and thus overreport other imports)—Chalendard, Raballand, & Rakotoarisoa (2019)	Import underinvoicing To reduce the payment of customs duties or to avoid product taxes—Yang (2008), Kellenberg & Levinson (2016)

*Source: Authors on the basis of Nitsch (2017) and other literature*

### 2.1.2. Data

Both GFI's Spanjers & Salomon (2017) and Ndikumana & Boyce (2010) use IMF's Direction of Trade Statistics (DOTS). DOTS covers many countries and has been the preferred source of international trade data because of its superior coverage of countries. DOTS include imports and exports of merchandise goods only (i.e. not services) and this limitation holds also for the other often used data source, UN Comtrade. Also, in both databases imports are usually reported on a cost, insurance and freight (CIF) basis and exports are reported on a free on board (f.o.b.) basis. C.i.f. values include the transaction value of the goods, the value of services performed to deliver goods to the border of the exporting country and the value of the services performed to deliver the goods from the border of the exporting country to the border of the importing country. F.o.b. values include the transaction value of the goods and the value of services performed to deliver goods to the border of the exporting country.

DOTS includes information at country level with trade flows between country pairs available for a subgroup of countries. When available, GFI's Spanjers & Salomon (2017) use DOTS data preferably at bilateral level (around half countries in Europe and Western Hemisphere) and otherwise at aggregate level (two thirds of all countries including a vast majority of countries in Sub-Saharan Africa and most countries in Asia and other regions). They do further adjustments to their trade mis-invoicing estimates (not discussed in

detail below) using data from Hong Kong, Switzerland, South Africa and Zambia with additional data for these countries. Similarly, Ndikumana & Boyce (2010) rely on the IMF's DOTS (using bilateral data for a group of industrialised countries) in trade adjustments of their estimates for trade invoicing.

Some recent research, such as Berger & Nitsch (2012), Kellenberg & Levinson (2016) and Ndikumana (2016), uses UN Comtrade. UN Comtrade data (discussed in some detail below) seem to be in some respects equivalent to IMF's DOTS data (but, importantly, the coverage of countries has been lower in UN Comtrade) and in respect of disaggregation, UN Comtrade seems to be the preferable source: data are available at a product level and for recent years, on a monthly basis. The mirror statistics approach could be applied to (rarely available) transaction-level data as well, but it is only possible if this data is available from two reporting countries so that their bilateral trade could be analysed (since in practice most of the transaction-level data that is currently available is limited to one country only, as we discuss in the final subchapter). This is in contrast with the abnormal prices methodologies discussed in the following subchapter, for which one country data source is sufficient and that might partly explain why most of the research at the frontier on the basis of the transaction-level data discussed in the third subchapter builds on ideas similar to those in the abnormal prices research (rather than mirror trade statistics).

Both IMF's DOTS and UN Comtrade, if used at country-level in particular, provide a very good international coverage in terms of a number of countries for which there is information available. However, the information is highly aggregated, giving the values of imports and exports with disaggregation by trading partner only, and any analysis on the basis of this data is naturally constrained by this aggregate data's limitations. We evaluate the methodologies critically below and it is clear that many of the limitations stem from, or are interlinked with, the nature of the data, its aggregate level in particular. Also, the IMF has expressed concerns about the use of the discrepancies in these international trade datasets. In a consultation for the SDG indicator (Inter-agency Expert Group on SDG Indicators, 2015, p. 280), the IMF argued that official estimates of trade mis-invoicing cannot be derived by transforming trade data from the IMF's DOTS or UN Comtrade either by individual data or in aggregate. Instead, the IMF representative Carol Baker argued that estimates of illicit financial flows should be based on an understanding of specific country's circumstances and on administrative data such as customs reports and we review the leading available estimate based on this type of data in the final sub-chapter.

UN Comtrade database contains the most detailed data available for global analysis of international trade. The most detailed classification of commodities (goods or products) follows the Harmonized System categorisation at the six-digit level. The UN Comtrade database collates, standardises and makes available data from national authorities (typically customs authorities) on the annual quantity, value and trade partner country of commodity trade. Complications in the data collation process mean that there are some limitations to what can be expected of the dataset.

UN Comtrade (2018) lists six limitations in particular, which we summarise as follows. First, confidentiality results in some data on detailed commodity categories not being available, although this is still captured in data on higher-level aggregates. Second, coverage is not complete; that is, while the database runs from 1961 to the present, not all countries report all of their trade for every year. Third, classifications vary—that is, different commodity classifications are used by different countries in different periods, so comparisons cannot always be exact. Fourth, conversion cannot always be precise; that is, where the database includes data that has been converted from one classification to another, these will not always map precisely one on to the other and hence imprecision may result. Fifth, consistency between reporters of the same trade is not guaranteed: that is, exporter and importer country might record the same trade differently due to various factors including valuation (e.g. imports c.i.f. versus export f.o.b.), differences in inclusions of particular commodities or timing. Sixth, country of origin rules mean that the ‘partner country’ recorded for imports will generally be the country of origin and need not imply a direct trading relationship. These limitations are relevant for analysis of prices as well as for mirror trade statistics. We believe that they require caution, but should not prevent careful use of the data.

### 2.1.3. Methodology

The trade mis-invoicing estimates by the GFI, recently reported by Spanjers & Salomon (2017), are based on the assumption that whatever exports or imports are reported by advanced economies, but not equally reported by developing countries, are illicit financial flows (either under-invoicing or over-invoicing). In addition to what they call lower bound estimate using only developing country-advanced economies relationships, their upper bound estimates are scaled up on the basis of assuming that traders mis-invoice with other developing countries at the same rate they mis-invoice with advanced

economies. An earlier similar approach applied by the GFI is named the trade mispricing model, e.g. Kar & Cartwright-Smith (2009), but here we focus on the recently published version by Spanjers & Salomon (2017). (While finalising the book, we note that Global Financial Integrity (2019) just published its most recent set of estimates, for 148 countries for years 2006–2015. The most significant methodological change, according to the report, involves the use of both DOTS and Comtrade data to generate two sets of estimates. We also note the change in assumed trade costs from 10 to 6 per cent.

GFI's Spanjers & Salomon (2017) use the following series of equations to explain their 'bilateral advanced economies calculation.'

$$ID_{jp,t} = \frac{I_{j,t}}{r} - X_{p,t}$$

$$ED_{jp,t} = \frac{I_{p,t}}{r} - X_{j,t}$$

where  $I_{j,t}$  are imports by the developing country  $j$  from the partner country  $p$  at time  $t$ ,  $I_{p,t}$  are partner country  $p$ 's imports from the developing country  $j$  at time  $t$ ,  $X_{j,t}$  are developing country  $j$ 's exports to partner country  $p$  at time  $t$ , and  $X_{p,t}$  are partner country  $p$ 's exports to the developing country  $j$  at time  $t$ . Through the use of  $r$  (assumed to be 1.1) they aim to make the import and export data comparable by converting import data reported as c.i.f. to an f.o.b. basis, in which export data are reported in IMF's DOTS.

Spanjers & Salomon (2017) interpret negative values of ID as import under-invoicing and illicit inflows and positive values as over-invoicing and illicit outflows. In parallel, they interpret negative values of ED as export over-invoicing and illicit inflows and positive values of ED as export under-invoicing and illicit outflows. In their interpretation, they make a number of assumptions that we discuss and, with the help of existing literature, critically evaluate below.

Furthermore, for developing countries for which the bilateral data used in the equations above are not available (almost two thirds of developing countries), Spanjers & Salomon (2017) apply what they call world aggregate calculation—substituting the individual partner countries  $p$  above with one partner, the whole world,  $w$ . Spanjers & Salomon (2017) themselves recognise a number of challenges related to this step. First, it implicitly treats developing country partner trade data as being as accurate as those of advanced economies. Second, it leads to what they call erratic swings in magnitude.

Spanjers & Salomon (2017) apply this approach to developing countries and their partner advanced economies to arrive at what they label low



estimates (they scale down the world aggregate calculation to include only the share of trade with advanced economies using the partner data). For their high estimates, they extrapolate this to the world total, assuming that trade mis-invoicing is as prevalent with other developing countries as it is with advanced economies. When the scale of trade mis-invoicing is summed up across all developing countries, the high estimates are bound to double count flows between developing countries, an issue that the low estimates avoid.

In a separate, but similar stream of studies, Ndikumana & Boyce (e.g. 2010) also estimate trade mis-invoicing. Ndikumana & Boyce (2010) make the trade invoicing adjustment by comparing countries' export and import data to those of trading partners, assuming the industrialised countries data to be relatively accurate and interpreting the difference as evidence of mis-invoicing. They use equations equivalent to those of Spanjers & Salomon (2017) to arrive at values of  $ID_{jp,t}$  and  $ED_{jp,t}$ . They then, in line with GFI's high estimates, extrapolate these estimates for industrialised countries to global totals by dividing each of  $ID_{jp,t}$  and  $ED_{jp,t}$  with the average shares of industrialised countries in the African country's exports and imports, respectively.

An important distinguishing feature of the Ndikumana & Boyce methodology in contrast with of GFI's Spanjers & Salomon (2017), is that for each year and each African country, the values of estimates of export discrepancies and import discrepancies are summed up to a total trade mis-invoicing (which is then added to their total estimate of capital flight). In GFI's labelling we can write the equation of Ndikumana & Boyce as:

$$Trade\ misinvoicing_{it} = ID_{jp,t} + ED_{jp,t} = \frac{\frac{I_{p,t}}{r} - X_{j,t}}{ICXS_i} + \frac{\frac{I_{j,t}}{r} - X_{p,t}}{ICMS_i}$$

where  $ICXS_i$  and  $ICMS_i$  are the average shares of industrialised countries in the African country's exports and imports, respectively. It implies that outflows and inflows can net out at this stage. This would lead to similar estimates to the GFI method when both export and import mis-invoicing estimates have the same sign, but to very different magnitudes where one indicates outflows and the other, inflows. Overall, Ndikumana & Boyce net off their estimates of illicit inflows to obtain a more conservative (and also more volatile) series, while the GFI argues that because 'there is no such thing as net crime', it makes sense to consider gross outflows (summing absolute values to arrive at a sum of illicit financial flows).

In practice a similar methodological approach can be applied not only at the country level, but also at the commodity level. For example, there are two

pieces of research carried out in the late 2000s that consider the mirror trade statistics and aim to explain the observed gap. Fisman & Wei (2009) focus on the trade in arts and find evidence consistent with smuggling patterns. Kee & Nicita (2016) find that exporters of products that have higher ad valorem equivalents of non-tariff measures tend to have larger trade discrepancies, suggesting firms mis-declare product codes or country of origin to circumvent the cumbersome and opaque non-tariff measures. Berger & Nitsch (2012) use a similar approach for more products and argue that the reporting gaps partly represent smuggling activities.

More recently, and more explicitly focused on illicit financial flows, United Nations Economic Commission for Africa (ECA) & African Union (2015), in the report of the High Level Panel on Illicit Flows from Africa (the 'Mbeki report'), aim to assess illicit financial flows at the country and sector levels through trade mispricing using mis-invoicing. They consider their methodology similar to the trade mispricing model used earlier by GFI, e.g. Kar & Cartwright-Smith (2009), and thus similar to the most recent trade mis-invoicing estimates of Spanjers & Salomon (2017). Although the logic remains the same, the quality of the methodology increases substantially and ECA (2015) is able to address many, but not all, problems of trade mis-invoicing estimates discussed below.

ECA (2015) improves the earlier methodologies in a number of aspects. In contrast with the GFI approach, the Mbeki report by ECA (2015) uses data from UN Comtrade, which provides bilateral trade data at the product-level for more than 5000 products (GFI's preferred IMF data do not contain this detail). ECA (2015) recognises that discrepancies can occur for a number of reasons, including but not limited to illicit financial flows. In line with Ndikumana & Boyce, but in contrast with GFI, ECA (2015) net off the estimates for a given pair of countries for a given product, which helps them avoid the issue of negative illicit financial flows. Rather than assume that c.i.f. values are 10 per cent higher than f.o.b. values, as both Ndikumana & Boyce and the GFI do, ECA (2015) estimate it using the CEPII's BACI database built upon UN Comtrade. Overall, with the application of these improvements ECA (2015) likely achieves more reliable estimates of trade mispricing than the other approaches; but some other important drawbacks of this adaptation of trade mirror statistics method remain. A similar, although modified, approach has been applied by Economic Commission for Latin America and the Caribbean (2016, p. 124) and, most recently, by Kravchenko (2018) of the United Nations Economic and Social Commission for Asia and the Pacific (ESCAP).

The research of Ndikumana (2016) is similar to that of Ndikumana & Boyce, but makes use of a different level and source of data. (Despite these differences, and Ndikumana (2016) using commodity- rather than country-level trade data, we include it here—as with the Mbeki report above—because of its similarities to Ndikumana & Boyce). Ndikumana (2016), in a report prepared, and later partially updated following a critical feedback, for UNCTAD, follows a similar methodological approach as in the research by Ndikumana and Boyce discussed above, but at a more detailed, commodity level. Ndikumana (2016) estimates export mis-invoicing ( $DX$ ), and import mis-invoicing ( $DM$ ), for country  $i$ , product (or commodity)  $k$ , and partner  $j$  at time  $t$ :

$$DX_{i,j,t}^k = M_{j,i,t}^k - \beta \times X_{i,j,t}^k$$

$$DM_{i,j,t}^k = M_{i,j,t}^k - \beta \times X_{j,i,t}^k$$

where  $M_{j,i,t}^k$  stands for imports by country  $j$  from country  $i$  in time  $t$  of commodity  $k$ , and, similarly,  $X_{i,j,t}^k$  for exports by country  $i$  to country  $j$  as reported by country  $i$ , and  $\beta$  is the freight and insurance factor (similarly to  $r$  in the previous subchapter).

As in the research by Ndikumana and Boyce, Ndikumana (2016) argues that positive values of  $DX$  and negative value of  $DM$  provide indications of export and import underinvoicing, respectively, and negative values of  $DX$  and positive values of  $DM$  indicate export and import overinvoicing, respectively. Ndikumana (2016) applies the methodology to selected countries and commodities. Carton & Slim (2018) use a modified version of the mirror trade statistics, applied to the Comtrade data of OECD countries and supplemented by trade intensity index. The methodology of both of these recent papers is also dependent on assumptions similar to those of Ndikumana and Boyce. Most of the criticism discussed below is therefore broadly relevant—in some cases likely to a lower extent because of the more detailed data used. In addition, the critical discussion specific to Ndikumana (2016) has been documented by Forstater (2016a) and Forstater (2016b). For example, Brühlhart, Kukuena, & Dihel (2015) explain the trade gap in Zambia's copper exports by the copper being traded by companies headquartered in Switzerland, but exported to other countries than Switzerland. Ponsford & Mwiinga (2019) document broader concerns using the example of the Zambian government's request for financial models from extractive companies.

While these trade-based estimates of illicit financial flows may include some trade mispricing by multinational enterprises for the purpose of

shifting profits to countries with lower taxation (so called transfer mispricing), trade mis-invoicing is a more crude approach to tax reduction than most of those challenged in the OECD Base Erosion and Profit Shifting action plan, the major international attempt to curtail the problem. The survey conducted by Baker (2005), which found widespread commercial tax *evasion* through trade, relates to an earlier period; and it may be thought likely that the documented explosion in sophistication of multinational tax minimisation practices has seen non-trade-based forms of avoidance become dominant.

Instead, the anomalies now estimated through mirror trade statistics may be more likely to reveal unrelated party transactions that aim to shift part of one party's income into a different country (so called trade mispricing). As GFI, for example, now state on their website (<http://www.gfintegrity.org/issue/trade-misinvoicing>, accessed 1 June 2018)—in contrast to Baker (2005):

*Because they often both involve mispricing, many aggressive tax avoidance schemes by multinational corporations can easily be confused with trade mis-invoicing. However, they should be regarded as separate policy problems with separate solutions. That said, multinational corporations can and do engage in trade mis-invoicing. This activity, however, involves the deliberate misreporting of the value of a customs transactions, and is thus illegal tax evasion, not legal tax avoidance.*

In the mirror trade statistics approach, researchers use mostly country-level trade data to establish anomalies in the declared values of total exports and imports, on the basis that these reveal illicit shifts of value. On one view, these estimates are rather conservative. They are able to pick up only a share of all of trade mispricing or trade mis-invoicing. The data does not pick up, for example, trade transactions where the mis-invoicing is incorporated in the same invoice exchanged between exporter and importer. In addition the data includes only goods and their results thus exclude any scale of mis-invoicing of services and intangibles. On the other hand, the estimates are based on a number of important assumptions and are bound to include much more than trade mis-invoicing, as discussed below.

Overall, the earlier studies succeeded in highlighting the importance of tax havens and illicit financial flows and bringing these issues to wider attention, but there are difficulties with these estimates and some of the individual methods were earlier criticised by, for example, Hines (2010) or Fuest & Riedel (2012) and a number of other chapters in the book edited by Reuter

(2012). Both Forstater (2015) and Reuter (2017) consider the estimates of illicit financial flows as overestimates and as playing a misleading role in the public debate.

Some problems are common to most of the pioneering research in this area (including these trade estimates as well as the capital account estimates in the following chapter). To be able to derive any estimates, most of the methods necessarily rely on strong assumptions, for example, about what the data on trade reflects. Similarly, most of the estimates do not shed more light on specific policy measures—the results may not provide more guidance for policy other than a general recommendation to reduce illicit financial flows; or, in the worst possible case, they could suggest erroneous areas for policy priority, if the broad trade channel is over-estimated. We discuss some of these critiques in detail below.

Because the studies which are critical of the methodologies, are often important contributions in themselves we briefly review their critical points one study at a time below. For each of the selected recent studies, we briefly sum up and evaluate their main points, including their views, if any, on how to improve estimates in the future. None of the reviewed critical studies dispute the existence of trade-based illicit financial flows, but they do raise important reservations about their estimated scale and the methodologies, notably their assumptions. This is underlined by Reuter (2017) in a recent study for the World Bank, who summarises some of the criticisms of, above all, the GFI approach in particular—but importantly also draws two other conclusions. First, he acknowledges that GFI is the only organisation that has consistently studied the phenomenon. Second, he argues that whatever the criticisms of the existing estimates, there is no doubt that illicit financial flows are substantial enough to merit close attention.

One of the few recent papers explicitly aimed at reviewing the methods, Johannesen & Pirttilä (2016) highlight three important conceptual issues (again of some relevance also to estimates based on capital account data). First, these estimates are likely to capture some completely legitimate flows, which the applied methodologies are not able to distinguish. Second, the approaches of GFI as well as Ndikumana & Boyce estimate net illicit financial flows and provide some scope for outflows and inflows to neutralize each other (at transaction, commodity, or country level). Where these result in zero or negative total illicit financial flows, they will complicate interpretation of the estimates. Third, since the illicit financial flows are often estimated as residuals or discrepancies, the resulting estimates will tend to be compounded by measurement errors associated with the trade flows.

Hong & Pak (2017) are concerned that the GFI estimates overestimate how much developing countries lose due to illicit financial flows. Hong & Pak (2017) focus on a specific assumption of what they call the partner-country trade data comparison method (what we prefer to call here trade mirror statistics). They argue that the assumption of no mis-invoicing in partner countries cannot be supported and raise doubts about the reliability of the method. Hong & Pak (2017) argue that the advanced economies trade data cannot serve as a counterfactual to the developing countries' trade data. Hong & Pak (2017) convincingly show that advanced economies also likely suffer from trade mis-invoicing, and therefore that the results of the trade mirror statistics approach are biased. Unfortunately, given the data limitations, the scale of bias is hard to determine. As a more promising alternative approach to mirror trade statistics, Hong & Pak (2017) consider the abnormal prices research that we discuss in the following subchapter.

Being critical of the methodologies as well as the excessive attention trade-based illicit financial flows might receive at the cost of other types of flows for which similar estimates do not exist, Forstater (2016a) focuses on the trade mirror statistics approach, while Forstater (2018) discusses tax and development more generally; and Forstater (2015) discusses profit shifting by MNEs (and we discuss her views on this in the later chapter focused on this type of illicit financial flows). Forstater (2016a), as well as her blogs, focuses on criticising empirical methodologies of illicit financial flows and their interpretations. For example, Forstater (2016a) provides some detailed criticisms of the Ndikumana (2016) report by Ndikumana. She proposes four areas for further work—understanding domestic realities, measuring international progress, commodity value chains and the role of multinational companies. The brief paper by Forstater (2016a) is accompanied by a comment by one of the GFI economists, Matthew Salomon, who agrees that focusing only on trade mis-invoicing as representative of all illicit financial flows would be too narrow, but asserts that trade mis-invoicing is an important area of further research and that even when there are detailed administrative data available, illicit financial flows remain unobservable and assumptions are needed to estimate them.

In a series of contributions—Nitsch (2012), Nitsch (2016), and Nitsch (2017)—Nitsch discusses the limitations of the trade-based methodologies. For example, Nitsch (2016) critiques the GFI methodology, focusing in particular on deficiencies in the use of mirror trade statistics to quantify the extent of capital outflows due to trade mis-invoicing. He identifies what he believes to be arbitrary assumptions, mixed methodologies and skewed

sampling to argue that their estimates have no substantive meaning. Nitsch (2017) observes that a highly disaggregated transaction-level data is usually not available to researchers and mis-invoicing behaviour is thus often identified from more aggregate trade information, which introduces two types of problems. First, at a more aggregate level, discrepancies in mirror trade statistics from mis-invoiced trade transactions may cancel each other out. Second, for the analysis of aggregate data, the set of assumptions that is used for the identification of mis-invoicing practices typically becomes even more restrictive and we discuss these assumptions below. An additional complication is that the accuracy of trade mis-invoicing estimates is unknown, since, as Nitsch (2017) argues, only an unknown fraction of all misreported trade activities is identified from official statistics.

Building on his earlier critical assessment in Nitsch (2012), Nitsch (2016) provides insights into pitfalls of mirror trade statistics and how the problems might be overcome (albeit he does not seem to be very optimistic on this topic in Nitsch (2017)). Nitsch (2017) presents similar critical points to Nitsch (2016) but makes somewhat more strident conclusions about existing methodologies ('a matter of faith') without providing much new guidance for improved methodologies in the future.

Below we focus on the discussion of assumptions by Nitsch (2016). Nitsch (2016) observes that the trade mirror statistics approach is in principle a credible methodology only if a few restrictive assumptions hold: for example, if it was applied on transaction-level data with information on the transactions from both countries, and the mis-invoicing affected only one side of the transaction. The latter is a crucial implicit assumption of the trade mirror statistics approach as applied by the GFI: the trade statistics of the two countries are assumed to be affected differently, with one a perfect reflection of reality (the transaction is recorded and is recorded correctly) while the other is deliberately mis-invoiced. While Fisman & Wei (2009) make the assumption explicit and argue why it is likely to hold in the case of antiques and cultural property, it is not clear from the GFI and other similar research how often trade mis-invoicing is carried out in this way (whether none, one, or both of the countries' statistics should be affected). Given these assumptions, not only it is hard to estimate the scale of illicit financial flows, but also hard to know the accuracy of these estimates.

Focusing on the deficiencies of the mirror trade statistics approach as applied by GFI in particular, Nitsch (2016) identifies four crucial assumptions of the GFI approach and some of these relate to other applications of mirror trade statistics. First, GFI assumes that the differences between export and



import values are homogeneous across countries at the rate of 10 per cent of transportation costs. He documents the sensitivity of this assumption as well as that it is not consistent with the observed values. A similar robustness check has been recently carried out by Erskine (2018), who shows systematic differences in mis-invoicing for landlocked and coastal African countries (a good proxy for a relative scale of transportation costs), providing further support for country-specific approach, as exemplified by ECA (2015). Second, all discrepancies in countries' trade statistics (other than these transportation costs) are assumed to be a result of trade mis-invoicing and thus illicit financial flows—which seems bound to lead to overestimates.

This assumption has been addressed by GFI to a degree. Since 2013, following a critical analysis by Kessler & Borst (2013), GFI take into account the transit trade of Hong Kong, which is important for China in particular, and this should make the estimates somewhat more realistic. They also made a few similar adjustments for other countries. But there are a number of countries that serve as transit jurisdictions and their role in trade might cause trade gaps ('Rotterdam and Antwerp effect'), as argued, for example, by Herrigan, Kochen, & Williams (2005).

Third, only discrepancies that lead to (positive) outflows out of developing countries are considered. A number of assumptions could explain this methodological position—either there are only outflows out of developing countries or only outflows are worth their focus or the method works well when outflows are estimates and not so well when inflows are estimates. At least one form of this assumption seems to be reflected in that GFI adds a particular flow to the overall sum only if it is an outflow from developing countries (any estimates that might indicate an inflow to developing countries are set at zero). And in a more recent report by GFI, Spanjers & Salomon (2017) also provide the estimated inflows in developing countries.

The fourth assumption, identified by Nitsch (2016), is that GFI assumes that countries' aggregate trade with the world is representative about trade mis-invoicing of country's partners. This aggregation enables the inflows and outflows to cancel out each other and thus the estimates based on comparison with the world are lower-bound estimates. This fourth assumption applies only to a part of GFI estimates since 2013, when they started using bilateral data for a share of the developing countries. GFI still partly, in their high estimates, relies on extrapolation, or scaling up, on a sample of advanced economies partners for the whole trade of developing countries—if advanced economies are likely to be the destination of more illicit financial flows than other countries, this extrapolation biases the estimates upwards.



There is another reason why this extrapolation likely leads to upwards bias. Any use of trade mirror statistics faces the challenge of attributing observed discrepancies to one of the partners since import overinvoicing in one country is equivalent to export underinvoicing in its trading partner. Without any decision, both of these were counted and thus double counted in the total. GFI solves this by focusing on outflows from developing countries. However, by this extrapolation, estimated trade mis-invoicing related to the trade among developing countries is counted twice. Furthermore, Kellenberg & Levinson (2016) find evidence of trade misreporting in both developing and developed countries, with only a few detected differences, and Hong & Pak (2017) make a similar point. Given the importance of these assumptions and changes in methodology, it is perhaps not surprising that the estimates published by the GFI are not very consistent over time. Also, Nitsch (2017) notes that the country-level estimates for some countries vary by orders of magnitude over the years.

Among his other, perhaps more minor, comments, Nitsch (2016) notes that although GFI has been transparent about the use of the data and methodologies, the fact that they often make changes in their methodologies makes any subsequent analysis difficult. He also observes that in GFI's first report on illicit financial flows out of developing countries, Kar & Cartwright-Smith (2008) start combining trade and capital-account data based estimates, but that they do not sufficiently discuss how the two overlap or complement each other.

Overall, Nitsch (2016) acknowledges that given the nature of illicit financial flows and data available, there is no first-best solution and he provides suggestions for a more nuanced approach in three areas. His first call for more micro evidence—perhaps focused on a small number of trading relationships important for a given country—is partly already being answered, as we review the recent research in our third subchapter. He hopes that this could shed more light on the relative importance of trade mis-invoicing in illicit financial flows. Second, for a global estimate he suggests to focus on a few large countries responsible for a majority of illicit financial flows. Third, he sees a potential in the use of the trade mirror statistics approach, especially at the product level and when institutional knowledge about practices of trade mis-invoicing is absent.

#### 2.1.4. Results

It is necessary to consider the results with a high degree of caution in the light of the critical evaluation of the methodologies above. This includes the results

estimated by the GFI. In the most recent GFI analysis of illicit financial flows to and from developing countries between 2005 and 2014, Spanjers & Salomon (2017) estimate the illicit financial flows (or outflows) from developing countries in 2014 at between \$620 billion and \$970 billion. In this report they publish such a range for the first time. Also for the first time, they put equal emphasis on inflows and estimate them in 2014 at between \$1.4 and \$2.5 trillion. These and earlier estimates of the GFI had arguably had an impact on media and public debate, with, for instance, *The Economist* (2014) using their results and linking them, among other examples, with money laundering through trade mis-invoicing by Mexican drug gangs. Focusing on trade-based money-laundering, Gara, Giammatteo, & Tosti (2018) provide a recent application of the method for Italy. Nitsch (2016) looks at the estimates of the GFI reports over time and observes two patterns: the estimated illicit financial flows increase over time, while estimates at the beginning of the sample period have been mostly revised downwards. He also points out the high variance of some of GFI's country-level estimates over the years, with some country estimates differing substantially from year to year (in some cases due to changes in methodology).

In the most recent report, GFI still combine capital-account and trade approaches to estimating illicit financial flows (we describe the former in the next chapter). In their lower bound estimates of outflows, trade mis-invoicing is responsible for two thirds of the total, while what they call unrecorded balance of payments flows (using net errors and omissions from the capital account as a proxy for these, which we discuss in detail in the next chapter) accounts for the remaining third. They estimate that sub-Saharan Africa suffers most in terms of illicit outflows. Sub-Saharan Africa is also the focus of the series of papers by Ndikumana & Boyce (e.g. 2010). Ndikumana & Boyce tend to publish only overall estimates of capital flight including trade mis-invoicing, and we thus cannot discuss their estimates here in detail. Instead, we discuss their overall estimates in the following chapter that focuses on estimates using capital account data.

In a section devoted to estimates of trade mispricing, United Nations Economic Commission for Africa (ECA) & African Union (2015) estimate these trade-based illicit financial outflows from Africa at \$242 billion for a period between 2000 and 2008. Making use of their product-level data, ECA (2015) estimate that around 56 per cent of these outflows come from oil, precious metals and minerals, ores, iron and steel, and copper. They highlight the most affected countries (such as Nigeria and Algeria for oil, Zambia for

copper) as well as the trading partners involved. In an update to ECA (2015), Economic Commission for Africa (2018a) estimate that net IFFs between Africa and the rest of the world averaged \$73 billion per year during the period 2000–2015 from trade reinvoicing alone. Economic Commission for Latin America and the Caribbean (2016) estimates that outflows from countries in Latin America and the Caribbean through international trade price manipulation have increased in the last decade, representing 1.8 per cent of regional GDP (totalling US\$765 billion in the period 2004–2013). In 2013, illicit outflows climbed to US\$101.6 billion and the associated tax losses stood at about US\$31 billion (0.5 percentage points of GDP) as a result of foreign trade price manipulation. This amount represents between 10 per cent and 15 per cent of the actual corporate income tax take. Mexico and Costa Rica are estimated to be among the most severely affected.

Taking a similar, but somewhat more general approach, Kellenberg & Levinson (2016) observe the differences in mirror trade statistics and find that gaps between importer- and exporter-reported trade at the country level vary systematically with GDP, tariffs and taxes, auditing standards, corruption, and trade agreements, suggesting that firms intentionally misreport trade data. Using the example of Cameroon, Raballand, Cantens, & Arenas (2012) present the use of mirror trade statistics as a useful tool to help identify customs fraud. Similarly, for Madagascar, Chalendard, Raballand, & Rakotoarisoa (2019) use mirror trade statistics at the individual transaction level to identify discrepancies and then products and importers in which customs fraud seems to be likely.

### 2.1.5. Conclusions

The influential illicit financial flows estimates by the GFI and Ndikumana and Boyce are based on country-level trade data, and are subject to well-argued critical evaluations of their methodology and results. The GFI estimates in particular have had their share of both media attention and criticism, and the latter remains largely relevant despite some methodological revisions over time. These limitations, coupled with the increasing availability of commodity-level trade data for a number of developing countries (e.g. through the UN Comtrade database), indicate a gap in research that could result into more reliable trade-based estimates. We investigate how existing research has made use of the advantages (as well the disadvantages) of these possibilities in

the next subchapter, before turning to the state of the art studies based on transaction-level data in the final subchapter.

## **2.2. Commodity-level Trade Estimates: Abnormal prices**

### **2.2.1. Overview**

Having discussed misreported trade volumes in the previous subchapter on the mirror trade statistics approach, we now turn to a discussion of misreported trade prices. The relevant studies here are based on trade data that allow identification of ‘abnormal’ prices at the commodity level. As the previous subchapter documents, some of the early trade data approaches in the literature on illicit financial flows use international trade data and the more recent study often used data at commodity (or product) level to study trade mispricing. Trade mispricing occurs when transactions between both related and unrelated parties are mispriced to avoid tariffs, taxes or achieve similar, illicit or other, objectives (in contrast to a more narrowly defined transfer mispricing that describes only transactions between related parties within a multinational corporation). In other words, trade mispricing (among unrelated trade partners) and transfer mispricing (among related partners such as affiliates of the same MNE), consist of inflating (or deflating) prices in order to shift income or profits from one country to another to take advantage of tax or other differences. For illicit financial outflows, trade mispricing enables shifting income or profits out of countries mainly either through import overinvoicing or export underinvoicing, although there are some plausible motivations for import underinvoicing or import overinvoicing, as summed up in Table 3 above.

In this subchapter we focus on studies making use of abnormal prices. These studies usually examine the normality or extremeness of trade prices, which are most often derived as unit values by dividing trading amount in currency with the corresponding amount of trade—weight in kilograms. The prices can be estimated as unit values only with these more detailed, commodity-level, data, rather than the country-level data used often in the studies based on trade mirror statistics approach. Already some of the above discussed research uses commodity level trade data, but its focus is on trade mirror statistics and thus trade that is not being recorded by one of the trade partners. In this subchapter we focus on trade mispricing and thus illicit financial flows that are being observed in the data.

### 2.2.2. Data

Much of the research by Simon Pak, John Zdanowicz and colleagues, such as de Boyrie, Pak, & Zdanowicz (2005), uses data from the United States Merchandise Trade Data Base of the United States Department of Commerce, Bureau of Census, which is a reliable source of detailed data, but only for one country's trading relationships, the United States. The US trade data is available on a monthly basis since 1989. Some studies combine this data source with other data sources—for example, Christian Aid (2009) also uses monthly Eurostat data for EU countries, which dates back to 1988. For both data sets used by Christian Aid (2009), even when some products have no defined measure of units and are thus not included in the analysis, the total number of observations per year is in millions (more than 10 million for the US during 2005–2007 period, while over 80 million observations for the EU in 2007). Some of this work uses the United Nations UN Comtrade database, discussed—including its limitations—in the previous subchapter.

### 2.2.3. Methodology

A number of studies have used trade data to study abnormal prices in order to estimate the scale of capital flight or illicit financial flows, with a duo of authors Pak and Zdanowicz carrying out pioneering work in this area (Christensen, Kapoor, Murphy, Pak, & Spencer, 2007; Zdanowicz, 2009) with their early study from 1994 (Simon J. Pak & Zdanowicz, 1994) and with perhaps a latest similar study published in 2018 (Cathey, Hong, & Pak, 2018). A number of these studies, such as de Boyrie et al. (2005), Zdanowicz, Pak, & Sullivan (1999), Pak, Zanakakis, & Zdanowicz (2003) use detailed transactions data from the United States Merchandise Trade Data Base of the United States Department of Commerce, Bureau of Census. Cathey, Hong, & Pak (2018), Pak in a report for Christian Aid (2009) and Pak (2012) use Eurostat data for EU countries in addition for the US data.

All of these and a number of other papers make use of a price filter approach or some variation of it and we describe it below. The objective of this method is to construct a price matrix from which normal prices are derived and compared with the actual prices to identify 'abnormal' prices and thus estimate the scale of related capital flows. The prices are constructed as unit values by dividing the financial amounts by physical weights. This approach reflects a hypothesised assumption that unit values for a given product

category should vary only within a relatively narrow interval. It implies that any outliers, abnormal prices, might suggest mis-invoicing and we discuss critical assumptions below.

In the detailed description of methodology, we focus on one of the papers, de Boyrie, Pak, & Zdanowicz (2005), in which they estimate the magnitude of abnormal pricing in international trade between the US and Russia. They use transactions data for over 15 thousand import harmonized commodity codes and over 8 thousand export harmonized commodity codes with detail over 18 million import transactions and 13 million export transactions per year for the period between 1995 and 1999. The fact that they focus on one country, Russia, enables the authors to provide detailed overview of the relevant literature, with Tikhomirov (1997) identifying Cyprus, the UK, Switzerland, the Netherlands, Germany and Denmark as the countries, additional to the US focus by de Boyrie, Pak, & Zdanowicz (2005), used to export capital from Russia.

Their price filter analysis relies on determining some transaction prices as abnormal. Importantly, they consider Russia-US transaction prices normal only when they are within the inter-quartile range of prices of (i) transactions between Russia and the US, or, alternatively, (ii) transactions between the US and all countries in the world. We capture this approach in the following equation for an example of capital flight resulting from over-invoiced exports from Russia to the US in year  $t$  for a commodity  $k$ :

$$KF_{i,j,t}^k = \left( P_{i,j,t}^k - \text{lower quartile of } P_{\text{world},j,t}^k \right) \times X_{j,i,t}^k$$

where  $i$  is Russia,  $j$  is the US,  $t$  is a year,  $k$  is a selected commodity. The equation for under-invoiced imports would follow a similar logic (using upper quartile instead of lower quartile), and similarly in alternative specifications with the use of  $P_{\text{Russia},j,t}^k$  instead of  $P_{\text{world},j,t}^k$  and median price instead of quartiles. For each alternative benchmark prices (i.e. world-US or Russia-US, quartile or median), they arrive at estimates of total capital flight by summing over-invoiced exports and under-invoiced imports together and across all commodities. In addition to estimating the scale of capital flight, they use econometric models by Cuddington (1987) to test whether the capital flight is due to money laundering, tax evasion or portfolio consideration.

Pak has adjusted this methodology for a larger set of countries for Christian Aid (2009). It uses the same data source for the US, and the detailed Eurostat data of 27 then members of the EU. As in de Boyrie, Pak, & Zdanowicz (2005), Pak in Christian Aid (2009) assumes that the price range between an upper quartile price and a lower quartile price for the most detailed product

classification is the arm's length price range. In contrast with de Boyrie, Pak, & Zdanowicz (2005), the trade data used are grouped at product level classification which is likely to result—with some overpriced and some underpriced transactions—into underestimation of the amount of mispricing. Also, Christian Aid (2009) notes that the fact that partner data from other countries are not used in this analysis and that large transactions that are only slightly mispriced might go undetected and contribute to underestimation. Other reasons, such as the product homogeneity assumption discussed above or the volatility of prices during the studied time periods (years used for the price quartiles), could support overestimation.

Using the example of Madagascar, an African country with one of the lowest income per capita and lowest shares of taxes per GDP, Chalendard, Raballand, & Rakotoarisoa (2019) use detailed statistical data from both Madagascar confidential database and UN Comtrade. Chalendard, Raballand, & Rakotoarisoa (2019) used the abnormal prices approach to indicate product misclassification. Specifically, they used inconsistent unit value as indicative of customs fraud—unit values of rice and fertilizers (products exempted from value added tax) were much higher than corresponding world prices.

Naturally, there are limitations to this methodological approach, some of which are common to trade mirror statistics discussed in the previous subchapter and some of which are new. For example, when deliberate trade mispricing does occur, it might be possible to detect it only when the mispricing is extreme and almost impossible when the mispricing is only slight. As The Economist (2014) argues, money launderers, who curb their greed and invoice goods up or down by, say, 10 per cent only, will probably continue to get away with it. We discuss the limitations, including the assumptions that determine price abnormality, below, and we focus here on the critical evaluation of the main Pak and Zdanowicz price filter approach, e.g. Carbonnier & Zweynert de Cadena (2015) and Nitsch (2012).

One important set of assumptions is about the role of prices used in the estimation. For such estimation of mispricing, one would ideally like to have a measure of what the price was if it was an arm's length transaction. This approach to estimating trade mispricing is similar to what recent studies at the frontier of research are estimating, but here the lack of persuasive counterfactual normal prices is substituted with quartile range thresholds. This most often used interquartile price range is endogenous and does not seem to be an objective basis for an arms' length price range. In addition, when product categories are used since transaction- or product-level data are not usually available, each category includes goods with a different degree of



heterogeneity. Pak & Zdanowicz (1994) argue that the use of inter-quartile range is supported by US regulation on transfer prices in international trade and they use two versions (US-Russia trade and US-world trade) and median prices as alternative benchmarks of normal prices. Still, these thresholds are understandably criticised, e.g. by Johannesen & Pirttilä (2016), as arbitrary. Pak in Christian Aid (2009) acknowledges it, actually using the same word (page 52). In addition to the arbitrariness of setting the interquartile range as the norm, Nitsch (2012) highlights that implementation of such a definition is sensitive to the number of observations—with a small number of relevant data points, as is often the case, potentially leading to biased results. In addition, variations in prices might be caused by (unobserved) differences in the timing of carrying out and/or recording trade transactions.

The required assumption of this approach is that there is a way to determine which prices are abnormal, but in reality the available data do not provide other options than the inevitably arbitrary statistical definitions such as interquartile ranges. Generally, there is no reliable guidance on what price is normal or not. As a potential remedy, in addition to average or other statistical distributions of unit values being used as the control prices or proxies for arms' length prices (in the inter-quartile method by Pak & Zdanowicz, 1994), also prices available from the markets can be used, as in the pioneering research by Hong, Pak, & Pak (2014), in which the authors use the import price of bananas reported by UNCTAD almost on a monthly basis. However, the market prices for many goods and product categories are not readily available, and some data sources might be actually subject to the similar challenges as the international trade unit values.

Nitsch (2012) points out that the data usually used are for product categories rather than products and that information is limited in respect of homogeneity of these product categories, including in quality, that might lie behind some observed differences in unit values. He argues that many of the product categories (around half) are catch-all with the word 'other' in their names.

It follows that one important assumption of this approach, which is partly shared even with the more recent studies in the following chapter, is that the products within the identified detailed product-level categories are homogeneous. This homogeneity assumption enables the authors to make abnormality responsible for the deviation from the prevailing prices of the product category defined by inter-quartile price range or median price. In the case of de Boyrie, Pak, & Zdanowicz (2005), they use harmonized commodity codes in the international price matrix, which are specific product classifications more detailed and arguably more useful than industry classifications (such as standard industrial classification codes). These harmonized commodity codes



are arguably the most detailed publicly available trade classification (the more confidential sources of more detailed data are discussed in the next subchapter). Still, if this assumption does not hold, for example in the case of quality differences, the method is to overestimate the extent of mispricing. An additional complication is that the identification of abnormal prices through unit values assumes that trade mis-invoicing is occurring exclusively via abnormal prices rather than weight, in case of which the identification of abnormal prices is inaccurate and, furthermore, the extent of this inaccuracy is unknown.

Partly to counter similar critique, de Boyrie, Pak, & Zdanowicz (2005) in the discussion of their results emphasise that their analysis identifies only potentially abnormally priced trades (for example, to help investigators pre-select cases for auditing) rather than proving that they are abnormal. They acknowledge that when the number of transactions is small for a certain commodity, their identification may not be reliable. Given the discussed assumptions, also other researchers using this approach argue for its use not for estimation of scale of trade mis-invoicing, but as a tool for detecting suspicious transactions from detailed trade data, for example, for auditing purposes by tax and legal authorities (Hong & Pak, 2017). Indeed, this is similar to what some economists at the research frontier do as we discuss in the following subchapter on studies using transaction-level data.

Relatedly, the World Customs Organisation (2018) presented its study report on IFFs and trade mis-invoicing to the Development Working Group of the G20 in July 2018. The multi-co-authored report argues that estimates of both partner country trade statistics and price filter methods are not sufficiently robust and should not be understood as a reliable quantitative measurement of the scale of IFFs, but rather as a risk indicator, which can be useful in comparing the risk of IFFs across commodities, countries and over a longer time period. The World Customs Organisation (2018) also makes the important point that rather than disputing the accuracy of individual assessment mechanisms, attention should instead focus on actions to combat IFFs, the existence of which is indisputable; the estimates of which, however, are dependent on the methodologies used.

## 2.2.4. Results

Academic studies have used trade data to study trade mispricing (Pak, 2007; Zdanowicz, 2009), and these types of methods have been also often applied by non-governmental organisations such Tax Justice Network (2007), Hogg et al. (2009), or Hogg et al. (2010). They all broadly support the view that tax

indeed motivates trade pricing decisions. However, the important assumptions needed and the partially aggregated nature of the data pose methodological limitations that lead us to interpret these results with caution.

The one study of Pak, Zdanowicz et al that we describe in the methodology section in detail, de Boyrie, Pak, & Zdanowicz (2005), attributes flows through trade mispricing to money laundering and tax evasion. For US-Russia trade data, they estimate the amount of capital shifted through abnormal prices from Russia in 1995 at 3 per cent and 6 per cent of total trade for exports and imports, respectively. They estimate annual capital flight from Russia to the US to range from a low of 0.2 billion USD in 1997 to a high of 0.6 billion USD in 1999 when compared to US-Russia transactions, and, alternatively, to range from a low of 1 billion USD in 1998 to a high of 5 billion USD in 1999 when compared to the US-world trade.

In a combination of mirror trade statistics and mispricing methods, Chalendard, Raballand, & Rakotoarisoa (2019) estimate for Madagascar that undervaluation and product misclassification, each roughly accounting for a half of the total, are responsible for potential revenue losses of almost 100 million USD, which represented 30 per cent of total non-oil revenues collected by customs in 2014. Clothing and telephones are most often undervalued, while fertilizers and rice are often misclassified.

Interestingly, Hong, Pak, & Pak, (2014) apply the abnormal pricing method with market prices for their main results, but compare it with estimates based on the interquartile price filter as well as trade mirror statistics. They show that imports are undervalued by 54 per cent on average between 2000 and 2009 using market prices as a benchmark in the case of US banana imports from Latin American and Caribbean countries; while using the other two, more common methods they find little evidence of either under- or over-valuation of US banana imports—suggesting, perhaps, that the methodological limitations of the common methods may tend to bias results *against* uncovering illicit activity in commodity-level data. Most recently, in a working paper of a Swiss-based research network focused on Laos and Ghana, Carbonnier & Mehrotra (2019) discuss results indicating economically significant estimates of abnormal pricing in Swiss commodity imports.

### 2.2.5. Conclusions

The existing evidence based on commodity level data is useful in highlighting the specific commodities and countries most vulnerable to trade mispricing,

but the results are of limited reliability for estimating the scale and are superseded in their credibility by estimates based on transaction-level data.

One area of promising future research could be to use compare the results achieved with the relatively detailed commodity-level data reviewed in this subchapter with the results using the methods at the frontier of research discussed in the following subchapter. It might be possible to calibrate estimates using UN Comtrade, on the basis of more reliable transaction-level data for countries for which both are available. This would provide evidence of not only the scale of potential bias of UN Comtrade-based commodity-level studies, but also indicate whether and to what extent UN Comtrade can be relied upon when there are no transaction-level trade data available.

### **2.3. Transaction-level Trade Estimates: Research frontier**

#### **2.3.1. Overview**

There is an increasing number of research papers that use detailed trade data at the level of transactions and, with this, methodologies that deliver more credible results. Their most obvious disadvantage in contrast with the studies discussed in the previous two subchapters is that they are limited in geographical coverage, usually focusing on one country only (namely, the source of the unique data). Most of the existing evidence is for major, high income economies such as the United States, France or the United Kingdom, but there are also recent preliminary results for South Africa by Wier (2017)—a first study using such detailed data and providing evidence for transfer mispricing for a developing country, and future research is likely to provide evidence for smaller and lower-income countries. The current difficulties in obtaining consistent, high-quality data of this type mean that the leading global estimates at present rely instead on national-level data—and serious criticisms, including of the GFI approach discussed above, have been raised and we discussed them above. While most of the studies below do not explicitly mention illicit financial flows, they are natural follow-ups to the previous two subchapters in estimating the scale of transfer and trade mispricing.

Below we discuss the earlier evidence for the US by Clausing (2003) and Bernard, Jensen, & Schott (2006), two influential empirical research papers on transfer mispricing for the US from the 2000s. Clausing (2003) provides one of the first empirical pieces of evidence consistent with theoretical predictions regarding tax-motivated income shifting behaviour. Bernard,

Jensen, & Schott (2006), in their well-cited working paper, developed a new method for identifying transfer mispricing and applied it to detailed data of US-based MNEs. There is also more recent evidence for the United States by Flaaen (2017), who uses transaction-level data to find profit-shifting behaviour by US MNEs via the strategic transfer pricing of intra-firm trade.

We also discuss the perhaps most persuasive recent evidence by Davies, Martin, Parenti, & Toubal (2017) as well as by Vicard (2015), both of which rely on detailed data for France. Vicard (2015), in a Banque de France working paper, provides evidence of transfer pricing and its increasing role for France over time. Using similar French data to Vicard (2015) but for one, earlier year only (1999), Davies, Martin, Parenti, & Toubal (2017) arrive at a somewhat lower estimate, most of which is driven by the exports of 450 firms to ten tax havens. We also discuss recent research for Denmark, in which Cristea and Nguyen (2016) use firm-level panel data on Danish exports to find evidence of profit shifting by MNEs through transfer pricing. We note that there is also recent evidence for the United Kingdom, although similarly to Wier (2017) for South Africa, we do not discuss below these recent research contributions. Liu, Schmidt-Eisenlohr, & Guo (2017) use detailed data on export transactions and corporate tax returns of UK MNEs, and conclude that firms manipulate their transfer prices to shift profits to lower-taxed destinations.

### 2.3.2. Data

This research area, which has been intensively developing in the last few years, uses data that are typically at the transaction level, and are confidential but sometimes made available through a collaboration with the country-specific source responsible for collection of the data and for its use for research purposes.

In one of the earliest contributions to this literature, rather than transaction-level data, Clausing (2003) uses monthly data on import and export product prices collected by the Bureau of Labor Statistics from 1997 to 1999 that differentiate between intrafirm and arm's-length transactions (in total, 425,000 observations of monthly prices 33 per cent of these for exports and 38 per cent for intrafirm trade). Bernard, Jensen, & Schott (2006) use the Linked/Longitudinal Firm Trade Transaction Database which links individual trade transactions to specific firms in the United States. It contains detailed foreign trade data, including whether the transaction takes place at arm's length or between related parties, assembled by the U.S. Census Bureau and the

U.S. Customs Bureau which captures all U.S. international trade transactions between 1993 and 2000.

There are two recent papers using transaction-level data. Vicard (2015) uses detailed firm level export and import data by origin, destination and product to estimate revenue impact of profit shifting through transfer pricing. He exploits the panel dimension of data and provides estimates for years 2000–2014. Also using French firm-level data, Davies, Martin, Parenti, & Toubal (2017) make use of 1999 information on the prices of products and whether they are arm's length or intrafirm transactions. They also employ the data to estimate the counterfactual arm's length prices of an intra-firm transaction. Furthermore, they argue that France's relatively simple exemption system of international corporate income taxation provides a more suitable system for studying tax-motivated transfer mispricing than the more complicated US system that aimed then to tax worldwide income of MNEs resident there. Similarly to Davies, Martin, Parenti, & Toubal (2017) for France, Cristea and Nguyen (2016) argue that Denmark is an interesting case study because of its territorial taxation system, in which only income earned from activities performed by Danish residents gets taxed. Cristea and Nguyen (2016) use a firm-level dataset of exports from Denmark between 1999 and 2006.

### 2.3.3. Methodology

To indicate whether there is evidence of tax-motivated transfer pricing in US intrafirm trade prices, Clausing (2003) applies a regression analysis to observe the relationship between export or import prices with tax rate, and includes a dummy variable to indicate when trade is intrafirm. Other similarly indirect evidence to Clausing (2003) that we do not discuss below includes Swenson (2001), who used firm-product level data to show that variations in the reported customs values of US imports from five major economies during the 1980s are consistent with the transfer pricing incentives created by taxes and tariffs. Also for the US, Neiman (2010) uses transaction-level data to show that intra-firm prices are less sticky and have a greater exchange rate pass through than arm's length prices. For the value added manufacturing data from across the OECD countries, Bartelsman & Beetsma (2003) disentangle the income shifting effects from the effects of tax rates on real activity and find evidence consistent with transfer pricing. Similarly, Overesch (2006) uses German MNEs' data to show that intra-firm sales are related to corporate tax rates.

The research discussed below with truly transaction-level data estimates the extent of transfer mispricing as the difference between the so called comparable uncontrolled prices and the actual MNEs' prices multiplied by the quantity traded:

$$\text{IFF by transfer mispricing} = (\text{Comparable uncontrolled prices} - \text{Actual MNEs' prices}) \times \text{Quantity traded}$$

Most of the research below uses this equation implicitly or explicitly in one form or another, but varies substantially with regard to details and especially how they estimate the prices and what control groups or variations in tax rates and other variables they make use of in their empirical strategies.

Bernard, Jensen, & Schott (2006) use a theoretical model to show that the difference between arm's-length and related-party prices depends on firm, product and country characteristics. In their empirical part, they estimate arm's-length-related-party price wedge as the difference between the log comparable uncontrolled price (a proxy for arm's-length price that they estimate on the basis of detailed data at the country, firm, month and transport mode level) and the log related-party price (which they directly observe). They regress firms' price wedges on destination-country tax rates and destination-country product-level import tariff rates as well as proxies of product differentiation and firm market power.

In his empirical strategy, Vicard (2015) uses the price wedge between arm's length and related party trade on a market (defined by destination country and product) and its correlation with the corporate income tax rate of each partner country compared to France as a systematic evidence of transfer mispricing.

In their theoretical framework, Davies, Martin, Parenti, & Toubal (2017) show that due to the concealment costs of transfer mispricing, only some MNEs might choose to do it, with the probability increasing with the tax differential between home and host countries and the amount of exports. In their framework, Davies, Martin, Parenti, & Toubal (2017) also recognise that intra-firm prices could systematically deviate from arm's length prices not only because of tax avoidance stressed by most of the other literature, but also because of pricing to market behaviour (which implies that exporters adjust their prices to the prices that prevail in the export markets). In their empirical approach, they control for pricing-to-market determinants (transport costs, tariffs, GDP per capita) to capture only the tax avoidance effects. In contrast with existing literature, the methodology and data of Davies, Martin, Parenti, &

Toubal (2017) provide evidence of the impact of tax rates and tax havens on transfer prices themselves rather than evidence suggestive of transfer pricing more generally. Furthermore, they use the somewhat ad hoc and outdated classification of tax havens proposed by Hines & Rice (1994), which results in ten tax havens indicated in their data sample: the Bahamas, Bermuda, the Cayman Islands, Cyprus, Hong Kong, Ireland, Luxembourg, Malta, Singapore, and Switzerland.

For the Danish export data, Cristea and Nguyen (2016) use triple difference estimations to exploit the response of export unit values to acquisitions of foreign affiliates and to changes in corporate tax rates. They estimate the extent to which MNEs manipulate both transfer prices to affiliates and arm's length prices to unrelated firms in order to reduce their global tax payments. They further argue that by ignoring the MNEs' manipulation of arm's length prices and using these as comparable uncontrolled prices, tax authorities and researches underestimate the extent to which the MNEs manipulate prices in order to shift profits.

### 2.3.4. Results

For the US trade data Clausing (2003) finds a strong relationship between countries' tax rates and the prices of intrafirm transactions. Controlling for other variables that affect trade prices, as country tax rates are lower, US intrafirm export prices are lower, and US intrafirm import prices are higher. Her results indicate that a 1 per cent drop in taxes abroad reduces US export prices between related parties by 0.9 to 1.8 per cent. This finding is consistent with theoretical predictions regarding tax-motivated income shifting behaviour. Bernard, Jensen, & Schott (2006) find that the prices exporters set for their arm's-length customers are substantially larger than the prices recorded for related-parties. The difference is smaller for commodities than for differentiated goods, is increasing in firm size and firm export share, and is greater for goods sent to countries with lower corporate tax rates and higher tariffs.

For French trading companies Vicard (2015) shows that the price wedge between arm's length and related party transactions varies systematically with the corporate tax rate differential between France and its trading partner. He estimates that this profit shifting decreased France's corporate tax base by 8 billion USD in 2008, and that the related missing tax revenues amount to 10 per cent of the corporate tax paid by multinational groups located in France that trade with a related party. He also finds that the scale is increasing over

time. He estimates the semi-elasticity of corporate profits to tax differentials at 0.5: that is, a 10-percentage point increase in tax differential would increase the pre-tax income reported by the affiliate by 5 per cent. This is based on transfer pricing in goods trade only and is thus relatively high in relation to other estimates on balance sheet data, which he challenges.

Estimates of Davies, Martin, Parenti, & Toubal (2017) suggest that export prices decrease with corporate tax rate only for intra-firm transactions, and only for countries with very low tax rates and especially tax havens (which they consider to combine low tax rates with other characteristics including banking secrecy). Davies, Martin, Parenti, & Toubal (2017) arrive at a somewhat lower estimate than Vicard (2015), most of which is driven by the exports of 450 firms to ten tax havens. Indeed, they find no evidence of tax avoidance once they disregard tax haven destinations. Still, they consider their estimates of tax avoidance through transfer pricing—at 1 per cent of total corporate tax revenues in France—as economically sizable.

Looking at the sensitivity of exports to tax rates, Cristea and Nguyen (2016) estimate that Danish MNEs reduce their export prices by 6 per cent in response to a 10 percentage point decrease in the tax rate of a country with lower rates than Denmark, which corresponds to a tax revenue loss of around 3 per cent of Danish MNEs tax returns. The responses in export prices are higher for differentiated goods (7 per cent) and for MNEs who establish new affiliates during the sample period (9 per cent).

### 2.3.5. Conclusions

The expanding number of research papers providing evidence consistent with trade or transfer mispricing in an increasing number of countries suggest that this is a universal phenomenon. One implication might be that it warrants global solutions. One such solution, for multinationals at least, would be the abandonment of the arm's length principle in favour of a unitary taxation approach (as is now under consideration at the OECD). Before any reform happens, it should be beneficial to see similar empirical analyses for other countries, if only to provide a preliminary basis for potential detailed audits by tax authorities or guidance on the type of regulation that is needed to limit tax avoidance, or to increase awareness and pressures for a reform.

These studies derive their credibility from and build on detailed, country-specific data and, therefore, cross-country estimates are usually not available. The shift in data availability that would allow comparable cross-country



analysis with substantial worldwide coverage, would be dramatic—however desirable—and feels distant at best. For the time being, the low number of countries with similar analysis and the diversity of data available and thus methodologies applied, do not enable a credible comparison of results across countries or the estimation of the global scale of the mispricing.

## 2.4. Conclusions on Trade Estimates

We end this chapter on trade estimates with conclusions drawn from our understanding of the large body of research and policy literature summarised in the preceding sub-chapters. First, international trade is an active channel for illicit financial flows and the research leading to trade estimates has been useful in a number of respects. From numerous case studies as well as indicatively from a number of aggregate studies reviewed here we learn about the use of trade mispricing to transfer funds illicitly across borders. The trade estimates have been helpful in shedding light on international trade data discrepancies. Also, many of the relevant studies have proven useful for customs officials in highlighting cases suitable for more detailed audit, and for policy makers in underlying areas of potential concern. We consider most of the recent transaction-level studies credible for estimation of the scale of trade-based illicit flows. In contrast, the estimates based on the trade mirror statistics approach and country-level data might have been helpful in the past for raising awareness about these issues, but we do not consider them sufficiently credible to inform us about the scale of illicit financial flows over time. We consider some of the abnormal pricing estimates useful as indicators for audit and other purposes, but we would not rely on them for estimates of overall scale.

Second, we observe improvements in the methodology applied by the GFI and other researchers in their quest to provide more reliable trade data-based estimates of illicit financial flows. Despite the related research usefulness in other respects and its recent advances, the employment of trade estimates for the SDG target is not straightforward. We recognise that much research has been carried out recently on trade mis-invoicing and on trade as a channel of illicit financial flows for many countries, and that there is an argument for its inclusion in the indicator of the target as discussed by, among others, the United Nations Economic Commission for Africa (ECA) (2015), Economic Commission for Latin America and the Caribbean (2016) and, most recently, by Kravchenko (2018) of the United Nations Economic and Social Commission for Asia and the Pacific (ESCAP). However, we find that their estimates are

still not of sufficient reliability, and allow for a wider interpretation than illicit financial flows. In addition, an increasing number of trade estimates from the frontier of research reach strong conclusions based on relatively high-quality data and methods. We judge the quality of these frontier estimates as sufficient, but their country coverage is poor and it does not seem feasible to extend them to many more countries in the near future. Indeed, there seems to be a trade-off for the trade estimates—either they are available for many countries but less credible, or they are of relatively high quality but available only for few countries (and, furthermore, it is presently difficult to compare the estimates across the few countries). There remains a gap to be bridged between the two subgroups of trade estimates, to achieve both sufficient quality and coverage. Clearly, more research in this area is required. For the time being, no indicator from the group of trade estimates seems to be workable as the indicator of the SDG target.

Third, while we identify a number of promising areas of further research, none seem sufficiently promising in the medium term to enable their inclusion as the SDG target indicator. One option is to improve the current methods, either at the country level—as exemplified by GFI’s recent changes or Kellenberg & Levinson (2016)—or at more detailed, commodity-level such as ECA (2015). Another promising area of future policy-relevant research is extending the current transaction-level methods to more countries, while making sure that they are comparable, ideally, across both countries and years. Even more reliable than the current one-country, one-data-source studies would be estimates based on customs data from both countries of the trading pair involved in any given transaction examined for illicitness. Before transactions-level data are available in most countries, to reach near-global coverage it might be worth trying to adapt these methodologies for trade data sets with less detailed data but better country coverage, as Kellenberg & Levinson (2016) have done with the trade mirror statistics method and UN Comtrade data. But so far, given the data limitations, a better country coverage can be attained to some extent only at the expense of methodological rigor.