

Reflections on Methods for Estimating Trade Misinvoicing —Approaching Consensus?

Session on *Illicit Financial Flows through Trade Misinvoicing*
WCO PICARD Conference, 2017

Tunis, Tunisia

26 September 2017, 10.30-12

Matthew Salomon

Senior Economist, Global Financial Integrity (Washington, DC, USA)

Adjunct Faculty, University of Maryland University College (Adelphi, MD, USA)

Contacts: msalomon@gfintegrity.org, matt.salomon@faculty.umuc.edu

Three Key Questions

1. *Scale*

How significant is trade misinvoicing?

2. *Methodology*

How do we detect trade misinvoicing?

3. *Policy*

How can we curtail trade misinvoicing?

Overview of Presentation

- We'll address all 3 questions, but with particular emphasis on question 2 (methodology)
- Defining trade misinvoicing (TM) [slides 4-5]
- Scale? An illustrative estimate [slides 6-8]
- Policy? Some general recommendations [slide 9]
- Methodology? Comparing methodologies [slides 10-24]
- Approaching consensus? [slides 25]
- Of course, the 3 questions are interrelated: progress on any one can mean progress on all

Defining Trade Misinvoicing (TM) & Concepts

- **Trade misinvoicing** (TM) is the willful falsification of the value or volume of an international transaction of goods or services by at least one party to the transaction.
- TM flows are subclass of **illicit financial flows** (IFF) between countries which, broadly, can be defined as illegal movements of value from one country to another.
- Such transnational financial flows can be considered to be **illicit** if the funds are **illegally earned, illegally transferred or illegally utilized**.
- TM may be related to but does not precisely correspond to **trade-based money laundering** (TBML) which has been defined to be “the process of disguising the proceeds of crime and moving value through the use of trade transactions in an attempt to legitimize their illicit origins.” [APG *Typology Report on TBML*, 20 July 2012, p. 9]
- Conceptual definitions of TM are distinct from operational definitions of TM which are implied by particular methods and data used in estimation.

Economic Justifications for Measuring TM—Some Examples

- **“Capital flight.”** Tracking recorded and unrecorded movements between countries in order to identify vulnerabilities and predict macroeconomic outcomes.
 - Numerous studies over decades
 - Net outflows from countries of interest
 - Potential implications for macroeconomic policy
- **Social cost.** There are benefits accruing to and costs incurred by private entities engaging in TM but there are additional costs to societies affected by such activity.
 - “Rent seeking.” Broadly, a redirection of economic resources away from their most productive uses and resulting inefficiencies of allocation and distribution.
 - For social costs, TM flows into and out of particular countries matter (i.e., no netting)
 - Significance? If measurement indicates a proliferation of TM flows into or out of particular countries, that signals the likely presence of rent seeking and social costs affecting those countries (and their trading partners).
 - Potential implications for macroeconomic and microeconomic policies

“Advanced” v. “Developing” Economies

- **Advanced (A) v. Developing (D).** One of many distinctions used in TM research.
- **Motivation.** The limited productive capacities of “developing” economies makes them particularly vulnerable to the social costs of TM.
- Those costs can undermine sustainable growth in living standards in developing countries as well as increase inequality in income and wealth. Those costs can also impede progress on important social goals, such as poverty reduction.
- The International Monetary Fund (IMF) uses three main criteria in determining whether a country is considered “advanced” or “developing” (i.e., developing and emerging markets):
 - 1) a country’s level of per capita income (WEO data),
 - 2) a country’s export diversification (COMTRADE data), and
 - 3) a country’s degree of integration into the global financial system (BOP data).
- The IMF classification is not based on numerical rules for the three factors; rather, they involve additional judgment by the IMF and designations can evolve over time.
- IMF: “The objective is to facilitate analysis by providing a reasonably meaningful method of organizing data.”

<https://www.imf.org/external/pubs/ft/weo/faq.htm#q4b>

Illustration: “Conservative” Estimates for Developing Countries Trading with Advanced Countries, 2005-2014

Assumed CIF/FOB margin	Misinvoicing Propensity = $100 \times \text{TM} / (\text{Exports} + \text{Imports})$
5%	13%
8%	12%
10%	13%
12%	13%
15%	13%

SOURCE: GFI[2017], p. 18, based on bilateral merchandise trade data (DOTS).

NOTE: 149 developing countries trading with 35 advanced countries. Denominator in TM propensity is estimated total developing country trade over the 2005-14 period, totaling \$101,794,724 million, US.

Scale: Is TM Significant?

- An estimated TM propensity at 13% of total trade appears to be significant in the usual meaning of the word.
- Selected reasons to believe the estimate might be “conservative”:
 - Excludes trade in services for which mispricing is even less detectable than it might be with goods trade, other things equal
 - Use of country-level bilateral trade aggregates “biases” estimate down, relative to identical analysis with underlying commodity detail
 - Technical adjustments (by IMF, by GFI) work to further “bias” down the estimate
 - Method (“macro” trade data) based on irregularities in macro data (aka “gaps”)
 - Does not pick up collusive cheating (two parties, same-invoice faking)
 - Does not pick up fraudulent transfer mispricing (one party to both sides)
- Many other reasons to question the reliability of any particular estimate
- No single study/methodology can claim to be the final word on the significance of TM
- We’ll pause briefly here and return to method/methodology/practice after we define some terms

General Policies for Curtailing TM

- In general, policies to curb illicit transnational flows can help curtail TM as well
 - Increase the amount and granularity of information available to governments concerning primary and intermediary parties to transnational trade and finance
 - Make better use of existing information to enhance enforcement
- Greatly increased transparency in many aspects of trade and finance can work to increase accountability all around.
- Some of the methods we discuss next can help

More details on policies for curtailing IFFs are provided in GFI[2017], especially pp. 21-23

General: Method, Methodology, Practice

Popular discussions of TM research often conflate issues of method, methodology and practice. Hence, a need for concreteness in terms.

Proposed definitions of terms for TM research (as used in research more generally):

- **Method.** A particular research approach to measuring TM.
 - Example: qualitative v. quantitative methods
 - Example : macro v. micro methods
- **Methodology.** Justifications for a particular method for TM measurement, usually defined by the type of data used. Appropriate questions for methodologies include:
 - Are the data used appropriate to the particular research objective?
 - Do (inevitable) enabling assumptions critically limit interpretations of results?
- **Practice.** Assessment of particular studies using one or more methodologies for TM measurement.
 - Assessment criteria are inherited from those attending the parent methods and methodologies in any given study
 - Example: robustness of reported results
 - Example: potential for replication and transparency of reporting

Methods for Measuring TM: Broad Types

- **“Macro” methods**

- Quantitative analyses of aggregate (macro) trade data
- General approach
 - Identify “irregularities” in data (aka, “discrepancies”, “gaps”)
 - Infer possible relationships between “irregularities” and TM

- **“Micro” methods**

- Quantitative and qualitative analysis of direct TM information
- Examples: Customs and FIU reports, field surveys, press reports

Some Strengths & Weaknesses of Methods

“Macro” methods

- Some strengths
 - Allows comparisons of TM measures across countries and over time
 - Such comparisons might be used to identify “risky” trade and traders
- Some weaknesses
 - No direct measurement of TM, just potential characteristics of TM
 - No unambiguous linking of TM measures to motives for illicit activity

“Micro” methods

- Some strengths
 - Direct measurement of TM, with important detail on scale and motives
 - Evidentiary value in prosecutions and guidance to future enforcement
- Some weaknesses
 - Difficult to compare TM measurements across countries and over time
 - Customs/FIU measurements: changes in incidence of TM or changes in enforcement?
 - How representative are particular field surveys and press reports?

Macro Method: Selected Methodologies

- **Method.** Quantitative analysis of aggregate (macro) trade data
- **Aim.** Identifying “irregularities” in data relevant to TM/IFFs
- **Methodologies.** Alternatives have been based on particular data available with varying detail and global scope
 - Bilateral trade data (e.g., *mirror-trade* or *partner-country* (“**P-C**”))
 - IMF (Direction of Trade Statistics, “DOTS”)—country-to-country bilateral merchandise trade flows: all countries, long time series, annual frequency
 - UN (Comtrade)—country-to-country bilateral merchandise trade flows: detailed classes of commodities (up to HS6 classification and by stage of processing), all countries, relatively long time series, monthly frequency
 - Unilateral trade data (e.g., *price-filter* (“**P-F**”))
 - Highly detailed published country-specific trade reports including sub-national information for reporting country on transactions of commodities at HS6 and higher degrees of disaggregation
 - Typically reported by advanced countries (e.g., US, EU, Japan among others)
 - National income & financial account aggregates (TM implicit)
 - Example: financial flows involving secrecy jurisdictions/tax havens

Macro Trade Methodologies: Notations

- Our notation indexes countries/locations in reporter-partner order: $\{i,j\}$ will indicate a trade flow involving location j as reported by location i .
- A mirror-trade pair is denoted by $\{V_{ijkt}^M, V_{jikt}^X\}$ where
 - V_{ijkt}^M denotes the value of imports (M) of good k from j at time t reported by i
 - V_{jikt}^X denotes the value of exports (X) of good k to i at time t reported by j
- Some datasets also provide trade volumes (Q) allowing further elaboration on a mirror-trade pair (with $P \equiv V/Q$, unit value):
$$V_{ijkt}^M \equiv P_{ijkt}^M Q_{ijkt}^M$$
$$V_{jikt}^X \equiv P_{jikt}^X Q_{jikt}^X$$
- Similar notations for country i exports and country j imports, etc

Step 1: Data

- The P-C and P-F methodologies typically employ data that differ in detail and scope
- P-C uses bilateral trade databases of varying commodity and temporal detail and global scope (e.g., DOTS, Comtrade)
- P-F uses trade databases with considerably more commodity detail than P-C, more location detail than P-C, but typically for only one or a few countries (less than global scope of P-C)

Step 1: Irregularities in P-C Methodology

- Because they are based on different types of data, it's not surprising that the P-C and P-F methodologies would use different benchmarks in defining irregularities
- P-C: The benchmark used for identifying irregularities in a country's trade reporting is the partner country's report. Some examples:
 - Import gap (arithmetic): $\Delta^M_{ijkt} \equiv V^M_{ijkt} - V^X_{jikt}$
 - Import gap (proportional): $\delta^M_{ijkt} \equiv V^M_{ijkt} / V^X_{jikt}$
 - Import gap (normalized): $Z^M_{ijkt} \equiv \Delta^M_{ijkt} / \frac{1}{2} (V^M_{ijkt} + V^X_{jikt})$
- Each of the P-C alternatives (and there are many more) have particular advantages for certain types of analysis and disadvantages for others

Step 1: Irregularities in P-F Methodology

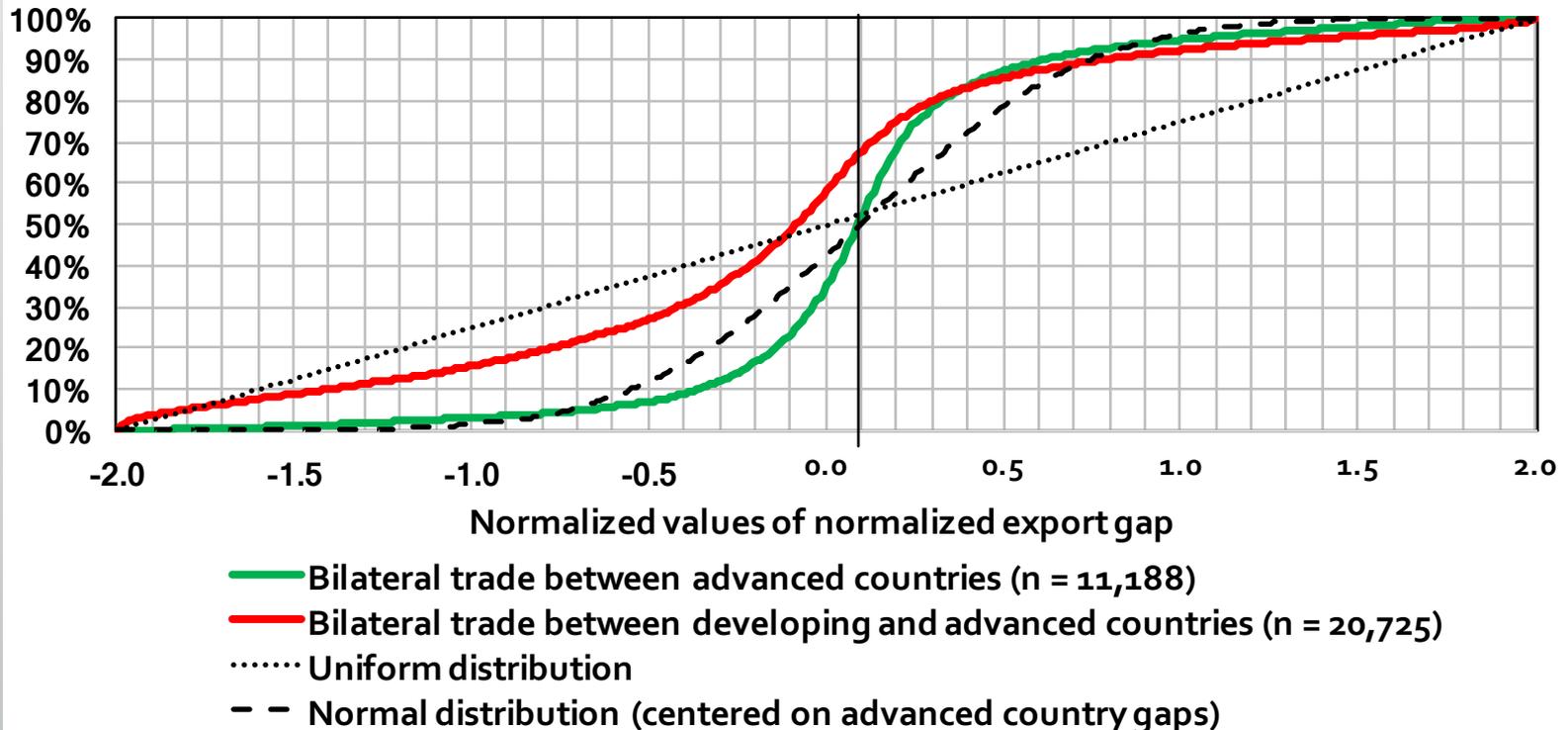
- P-F focuses on the distribution of trade prices observed in the database of the reporting country (say “A”)
- For country A’s imports, irregular prices in $\{P^M_{Ajkt}\}$ might be defined to be:
 - (a) prices lying outside the interquartile range (IQR) of $\{P^M_{Ajkt}\}$
 - (b) prices “distant” from some global reference price for k at t
- Again, as with P-C, each measure has advantages and disadvantages in particular empirical contexts

Statistical Considerations Relevant to TM?

- Both P-C and P-F databases are large, but we must expect some relevant information to be located in the tails of distributions
- Illustration for P-C: normalized export gaps (DOTS) for distinct subclasses of countries show some indication of heavy tails and skewness with differences in degree by categories of countries trading (graph)
 - A-A reflects advanced-country reporters trading with advanced-country partners
 - D-A reflects developing-country reporter trading with advanced-country partners
- Illustration for P-F: very large database (relative to a single country subsample of Comtrade, for example) but sub-samples for individual price distributions $\{P^X_{Ajkt}, P^M_{Ajkt}\}$ can be quite small.
- Statistical measures for identifying irregularities should be robust
 - Location--median more robust than mean
 - Scale—Median absolute deviation (MAD) widely used and is preferred to sample variance (which is scale dependent!). More robust measures can be based on statistics developed in Rousseeuw-Croux[1993]
 - Skewness—Adjustments to IQR to account for skewness in outlier detection can be based on Brys-Hubert-Struyf[2004]

Example of Heavy Tails and Skewness in P-C Irregularities (CDF is cumulative distribution function)

Illustration: Empirical CDFs for Export Gaps, A-A & D-A Trade, 2005-2014
Percent of observations at or below particular values of the normalized export gap



SOURCE: Author's estimates using data from the International Monetary Fund, Direction of Trade Statistics.

NOTE: The observed export gaps were normalized by dividing the gap (reporting country exports minus partner country imports) by the average of the two mirror trade flows. By construction, that normalization yields values that are bounded below by -2 and above by 2 for positive trade flows that are not both zero.

Step 2: From Irregularities to TM (P-C & P-F)

- How does an irregularity translate into TM?
- The irregularities identified in macro trade methodologies (P-C and P-F) could arise for reasons unrelated to TM
- Simply assuming the irregularities (P-C or P-F) correspond directly to TM is always unrealistic and, in some contexts, misleading (Hong-Pak[2016])
- Step 2 is necessary and the trade data alone do not indicate a clear answer
- Some probability statement is required (and desired)

Step 2: Some Issues for P-C Methodology

- Observed P-C irregularities could arise for a multitude of reasons, legitimate or illegitimate.
- Leading examples of well-documented (mostly) legitimate reasons:

Valuation. By convention, exports are reported on an FOB basis while imports are reported on a CIF basis that includes margins for transport costs, insurance, etc. We generally don't have direct data on the CIF/FOB margin and must make some assumptions, often using other data.

Timing. Trade takes time. Exports are shipped on a different date than the imports are received (the two "t"'s don't match reality). Even in annual data, some transactions will straddle one year to the next.

Commodity misidentification. It may happen that a commodity is identified one way at one port and another way at the other port (the two "k"'s don't match reality). This could happen due to honest error or due to deliberate misinvoicing (e.g., a commodity label is misstated to avoid import duty).

Re-exports/transshipments. ("Rotterdam effects") Some goods are reported as imports by a country that are only at an intermediate stage in their transit toward another final destination (recorded $\{i,j\}$ may not reflect reality). Similarly, for exports.

- There are ways of mitigating such problems in practice, though none is perfect.

Step 2: Some Issues for P-F Methodology

- Similarly, observed P-F irregularities could stem from a multitude of factors, legitimate or illegitimate.
- Some examples:
 - Heterogeneity
 - Seasonality
- As with P-C, there are ways of addressing such problems in practice, though, again, none is perfect.

In the End: A Lingering Question

- The world is imperfect. So are its data.
- Ideal data. Imagine a data set that has enough detail to address the particular challenges of the various methodologies and methods.
 - In such a data set, the data currently underlying the P-C and P-F methodologies would reflect no more than alternative slices of the same distribution.
- Lingering question. Even using such ideal data, it would still be impossible to identify the direction of TM by analyzing the irregularities alone.
 - The irregularities in the ideal data would conflate possible TM with statistical error
 - Even if we could extract a component of the irregularities that correspond to something other than statistical noise (i.e., fit for TM study), the data alone would not tell us much about the direction of TM in any case
 - Does a defensible statement that a particular value of $\{P_{ijkt}^M\}$ is relatively high, for example, imply:
 - importers in i are “overpricing” their goods received from j ?
 - exporters in j are “underpricing” their goods sent to i ?
 - some combination of both? (or even neither?)

Spotlight on the Probability Statement

- What leads researchers focused on particular cases to conclude that identified irregularities may correlate with TM?
- Example: decades of research advances on “gravity” explanations of bilateral trade have yielded some recent studies that correlate bilateral trade gaps (P-C) with tax and corruption-related and other time series
 - Carrère-Grigourio[2015]
 - Kellenberg-Levinson[2016]
 - Research underway at such official institutions as the Italian Financial Intelligence Unit and the Bank of Italy (among others)
- Economic theory can be used to articulate optimizing models of “cheating”...those notional models can help identify the kinds of data that might be useful in studying certain types of TM (e.g., tax wedges in tax-related TM)

Approaching Consensus? Conclusions

- The world is imperfect. So are its data. Researchers must do their best.
- Lessons from signal theory may be useful here: confronted with an array of noisy indicators:
 - it's always optimal to use all the indicators in some way
 - it's almost never optimal to choose one and ignore the others
- Approaching consensus? Should we? Can we? Are we? Yes to all.
- Growing body of research worldwide is improving the quality of trade data & empirical research
- Methodologies (like P-C and P-F) oriented toward identifying risky trades and traders can be examined for coherence
- In studying particular countries, there is no alternative currently to using all methodologies (to the greatest extent possible)

Works Cited

Brys, G., Hubert, M. and Struyf, A., "A robust measure of skewness," *Journal of Computational and Graphical Statistics*, Vol. 13, 2004, pp. 996-1017.

Carrère, C. and Grigoriou, C., "Can mirror trade data help to capture informal trade?" *Fondation pour les Études et Recherches sur la Développement International*, Working Paper 123, March 2015.

GFI, *Illicit Financial Flows to and from Developing Countries: 2005-2014* (Washington, DC, Global Financial Integrity, April 2017) http://www.gfintegrity.org/wp-content/uploads/2017/05/GFI-IFF-Report-2017_final.pdf.

Hong, K., Pak, C.H. and Pak, S., "Measuring abnormal pricing—an alternative approach: The case of US banana trade with Latin American and Caribbean countries," *Journal of Money Laundering Control*, Vol. 17 No. 2, 2014, pp. 203-218.

Hong, K. and Pak, S., "Estimating Trade Misinvoicing from Bilateral Trade Statistics: The Devil is in the Details," *International Trade Journal*, July 2016.

Kellenberg, D. and Levinson, A., "Misreporting Trade: Tariff Evasion, Corruption, and Auditing Standards," (manuscript, 24 May 2017) <http://faculty.georgetown.edu/aml6/pdfs&zips/TradeEvasion.pdf>.

Rousseeuw, P.J. and Croux, C., "Alternatives to the Mean Absolute Deviation," *Journal of the American Statistical Association*, Vol. 88, No. 424 (Dec. 1993), p. 1273-1283.