Project Overview

Accurate trade data remains central for empirical investigations of international trade and informed formulation of trade policies. However, discrepancies in trade reporting, stemming from reasons such as logistics all the way to deliberate misclassification, pose challenges to obtaining an accurate representation of trade activities. As the importance of these issues comes into sharper focus, an increasing number of studies have suggested that discrepancies in reported trade figures are not just minor blips. They're significant enough to sometimes skew the results of studies, potentially compromising the validity or robustness of obtained results.

Project Aim

The aim of this project is to provide tools and resources for assessing the quality of reported data in international merchandise trade.

Methodology

This study focuses on UN Comtrade, the largest and most widely used database in empirical research on international merchandise trade. The quality of reported trade is investigated in relation to its mirror, i.e the data reported by the trading partner. To measure the level of divergence between the reported data and its mirror, we rely on the Discrepancy Index as a measure of bilateral asymmetry. Our working paper delves further into how the Discrepancy Index is constructed and subsequently applied to investigate the quality of reported trade flows. Our analysis provides comprehensive coverage of discrepancy and reliability measurements across reporting nomenclatures and periods.

Deliverables

Analytical Code:This repository includes the code necessary to replicate our study's findings. It aims to provide interested parties a way to validate and explore our findings in depth.

Indicators of Data Quality: Our repository provides country- and product-level indicators related to the quality of reported data. These indicators are accessible for multiple years and across different reporting nomenclatures.

Profiling Tools: We offer code for generating detailed profiles for countries, products, and trade partnerships, allowing for a more in-depth investigation of reporting practices. These tools are especially useful for case-by-case analysis, where analysts are interested in the reporting patterns of specific reporters.

While direct access to UN Comtrade data or our mirrored dataset is not provided, our repository includes scripts for constructing mirrored data tables, enabling users to apply our methodologies to their data sets.

Project Goals

This project aims to equip trade analysts with the necessary tools to make informed decisions regarding the use of a country's reported data or its mirror. As such it provides a variety of tools intended to visualize and explore reporting discrepancies. These tools can be applied to assess the quality and reliability of reported data, identify different reporting problems faced by countries, explore potential drivers of recording gaps and aid reconciliation procedures for trade flows. Through this initiative, we aim to contribute to the improvement of trade data accuracy and, consequently, the formulation of more reliable trade policies.

Repository Content

1. Data

This section provides detailed documentation on the data sources used in our working paper.

Data File	Source	Note	Provided	Data URL
dianapublic.unsd_wits_annual	UN Comtrade	Detailed global bilateral trade data in the Standard International Trade Classification (SITC) and Harmonized System (HS) product classification.	Yes	comtradeplus.un.org
diana.income_region_country_list	World Bank	Historical data for income classification, based on GNI per capita in current USD, using the	Yes	datacatalog.worldbank.org

Data File	Source	Note	Provided	Data URL
		Atlas method.		
diana.wdi_gdp	World Bank	GDP per capita, PPP (constant 2017 international $).	Yes	databank.worldbank.org
default.spi_index	World Bank (Dang et al., 2023)	Statistical Performance Indicators that assess the maturity and performance of national statistical systems.	Yes	github.com/worldbank/SPI
diana.cepii_gravity_202211	CEPII (Conte et al., 2022)	Different measures of bilateral distances.	Yes	cepii.fr
diana.productlist	World Customs Organization	The product names and structure of the Harmonized System method of classifying traded products.	Yes	wcoomd.org

2. Results

This directory contains all the necessary code and instructions to replicate the primary results presented in our working paper. It aims to provide interested parties a way to validate and explore our findings in depth.

3. Reporting profiles

In this section, we introduce specialized code that enables users to generate reporting profiles for different countries, products and trade partnerships. Users can input specific parameters such as reporter country, year, nomenclature, and partner country to obtain customized reporting profiles. We demonstrate their application through the case of Senegal and the France - Madagascar trade relationship, examples highlighted in our working paper.

4. Supplementary materials

This section includes additional resources and materials that supplement the main content of the repository. It may contain additional datasets, technical documentation, or extended analyses that provide further context or support for the research presented.

Database Availability Statement:

The Discrepancy Index and the aggregate indicators were generated using UN COMTRADE database (https://comtradeplus.un.org/). World Bank has access to this database and we replicated the data in DataBricks using the Bulk API. To obtain UN COMTRADE data, please visit https://shop.un.org/databases to check how you can access the data. Please note based on UNSD terms you may have to pay a subscription to get access to this data. The output is avaiable via World Bank Data Catalog and the links are provided below. All databases are open and available using license Creative Commons Attribution 4.0. ● Discrepancy Index Annual for H5 nomenclature (https://datacatalog.worldbank.org/search/dataset/0064901/Discrepancy-Index-H5) ● Discrepancy Index Annual for H4 nomenclature. (https://datacatalog.worldbank.org/search/dataset/0064897/Discrepancy-Index-H4) ● Discrepancy Index Annual for H3 nomenclature (https://datacatalog.worldbank.org/search/dataset/0064895/Discrepancy-Index-H3) ● Discrepancy Index Annual for H2 nomenclature (https://datacatalog.worldbank.org/search/dataset/0064890/Discrepancy-Index-H2) ● Discrepancy Index Annual for H1 nomenclature (https://datacatalog.worldbank.org/search/dataset/0064889/Discrepancy-Index-H1) ● Discrepancy Index Annual for H0 nomenclature (https://datacatalog.worldbank.org/search/dataset/0064887/Discrepancy-Index-H0) ● Trade reporting Indicators all nomenclatures (https://datacatalog.worldbank.org/search/dataset/0064900/Reporting-Index-HS) ● Trade reporting Indicators by commodity all nomenclatures (https://datacatalog.worldbank.org/search/dataset/0065794/reportingindexproducths)

Authors

- Sonja Mitikj - github/email
- Siddhesh Kaushik - github/email

Assumptions:

- You have access to UN COMTRADE database.
- You have access to Databricks, if not you may have to change the code from PySpark to Python