Growth Lab

Tackling Discrepancies in Trade Data: The Harvard Growth Lab International Trade Datasets

Sebastian Bustos, Ellie Jackson, David Torun, Brendan Leonard, Nil Tuzcu, Piotr Lukaszuk, Annie White, Ricardo Hausmann, and Muhammed A. Yildirim

Growth Lab Working Paper Series
No. 251

July
2025

GROWTH LAB
HARVARD KENNEDY SCHOOL
79 JFK STREET
CAMBRIDGE, MA 02138

GROWTHLAB.HKS.HARVARD.EDU

Acknowledgements
We would like to thank Timothy P. Cheston for his help with the data validation and contributions to the Atlas of Economic Complexity. We would like to thank Mali Akmanalp and Romain Vuillemot for working on earlier versions of the Atlas data and architecture. We would like to thank the current and former members of the Harvard Growth Lab for continuous feedback on the data.

This paper may be referenced as follows: Bustos, S., Jackson, E., Torun, D., Leonard, B., Tuzcu, N., Lukaszuk, P., White, A., Hausmann, R., Yildirim, M.A. (2025). “Tackling Discrepancies in Trade Data: The Harvard Growth Lab International Trade Datasets.” Growth Lab Working Paper, John F. Kennedy School of Government, Harvard University.

Tackling Discrepancies in Trade Data: The Harvard Growth Lab International Trade Datasets

Sebastián Bustos[*,†] Ellie Jackson[*,†] David Torun[*,†] Brendan Leonard[*]
Nil Tuzcu[*] Piotr Lukaszuk[‡] Annie White[*] Ricardo Hausmann[*,§,||]
Muhammed A. Yıldırım[*,¶,||]

July 24, 2025

Abstract

Bilateral trade data informs foreign and domestic policy decisions, serves as a growth indicator, determines tariffs, and is the basis for financial and investment decisions for corporations. Accurate trade data translates into better decision-making. However, the raw bilateral trade data reported by UN Comtrade suffer from two structural problems: reporting differences between country partners and countries reporting in different product classification systems, which require product-level harmonization to compare data across countries. In this paper, we address these challenges by combining a mirroring technique and a data-driven concordance method. Mirroring reconciles importer and exporter differences by imputing country reliability scores and applying a weighted country-pair average to calculate the estimated trade value. We harmonize product classifications across vintages by calculating conversion weights that reflect a product's market share. The resulting publicly available datasets mitigate issues in raw trade statistics, reducing reporting inconsistencies while maintaining product-level granularity across six decades.

1 Introduction

International trade is a data-rich field with intensive empirical applications. Export and import transactions of goods are classified and recorded by government agencies worldwide, generating a wealth of bilateral trade data. In principle, each transaction should be recorded

[*]The Growth Lab, Center for International Development – Harvard University, Cambridge, MA, United States.
[†]These authors contributed equally.
[‡]State Secretariat for Economic Affairs (SECO), Switzerland. The views expressed in this study are the author's and do not reflect those of SECO.
[§]Santa Fe Institute, Santa Fe, NM.
[¶]Department of Economics, College of Administrative Sciences and Economics - Koç University, Istanbul, Turkey.
[||]To whom correspondence should be addressed.

twice: once by the exporting country and once by the importing country. Moreover, countries are expected to report their trade transactions using the most recent classification system. However, despite significant efforts to coordinate and standardize reporting practices, the quality and consistency of reported trade data vary substantially across countries. Such inconsistencies challenge any attempt to analyze trade over time, gauge market potential, track shifts in comparative advantage, or study evolving value chain dynamics, all of which require data that are complete, reliable, and fully comparable across years and partners. In this paper, we introduce and document the construction of a set of international trade datasets spanning the past six decades. These datasets are designed to address several imperfections in the raw data, including discrepancies in partner reporting and discontinuities across product classification systems.

Two key issues plague the construction and use of trade datasets. First, discrepancies between exporter- and importer-reported values are common. These differences arise not only at the aggregate level of total bilateral flows, but are often magnified at more disaggregated levels, including individual product records. In many cases, trade flows are reported by only one of the two trading partners, and even when both report, the values often differ substantially. This raises the question of which measure more accurately reflects the underlying *true* trade flow.

Second, each country reports its trade using a single product classification, but different revisions or vintages of these classifications have been introduced over time. While some countries often adopt new versions in a timely manner, others tend to lag in implementation, hindering cross-sectional analysis. The coexistence of multiple classification vintages, which must be harmonized using general concordance tables, further complicates the computation of coherent and comparable trade statistics. Although UN Comtrade provides harmonization services,[1] its approach substantially reduces the number of unique products, leading to artificial product disappearances and discontinuities in time series.[2]

This paper introduces new datasets that address key challenges in international trade statistics by systematically reconciling exporter and importer reports[3] and improving conversion across different product classification systems, leveraging the data-driven Lukaszuk-

Torun method.[2] We propose a comprehensive method to tackle major inconsistencies in raw trade data and apply it across multiple vintages of international trade classifications. The result is a collection of bilateral, product-level trade datasets spanning from the 1960s to the present, aligned with the classification vintages used during each period. These datasets have powered the visualizations in the Harvard Growth Lab's *Atlas of Economic Complexity* since 2014, with earlier versions widely used in both academic and policy research.[4–10]

This paper proceeds as follows: First, we document in detail the nature and extent of reporting discrepancies and classification mismatches in existing trade data. Second, we present our methodology for reconciling mirror flows and harmonizing classification correlations. Third, we describe the pipeline used to construct the new, cleaned, and harmonized dataset. Finally, we evaluate the improvements achieved relative to the raw data and existing benchmarks.

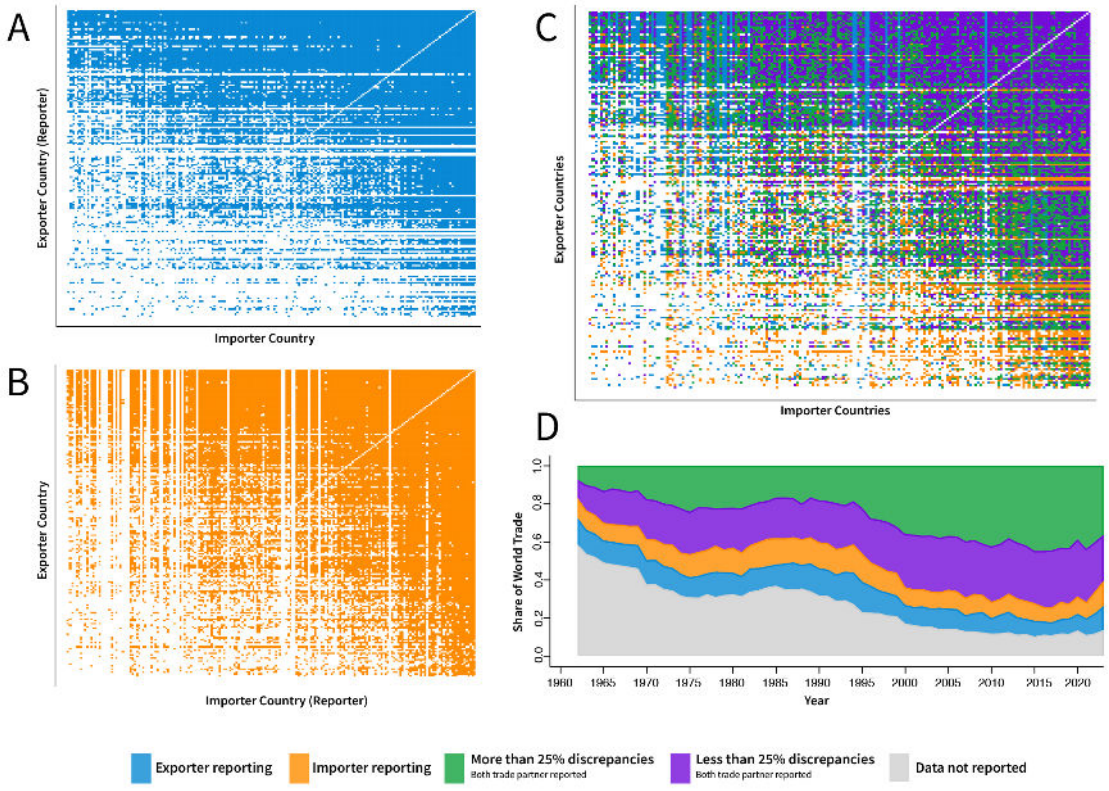
1.1 The Reporting Problem

To illustrate the reporting problem, we abstract from the product dimension and focus solely on total bilateral trade flows. Suppose that goods flow from an exporting country e to an importing country i, with a value denoted by $V_{e,i} \equiv V_{e \to i}$. The exporting country e reports this value as $V^E_{e,i}$, while the importing country i reports it as $V^I_{e,i}$. In a frictionless world, i.e., without trade costs or reporting issues, these reported values should coincide, that is, $V_{e,i} = V^E_{e,i} = V^I_{e,i}$. There are many legitimate reasons why $V^E_{e,i} \neq V^I_{e,i}$. For instance, differences may arise due to costs related to insurance and freight, or because shipments departing at the end of one year may be recorded upon arrival in the following year, leading to temporal mismatches between exporter- and importer-reported values.

However, in practice, the discrepancies between reported flows are often much larger than would be expected from normal statistical variation. Figure 1 shows various differences between exporter- and importer-reported trade values. Panels A and B show whether exporters and importers, respectively, reported trade flows with each possible trade partner in the year 2010 (Note that the choice of the year 2010 is not particularly meaningful; it serves as a representative example). In both panels, countries, whether as exporters or importers, are ranked by their total trade. At first glance, the top-right corners of Panels A and B of Figure 1 ap-

pear almost completely filled, while the bottom-left corners are largely empty. Countries that export and import large volumes tend to trade with most other countries, whereas countries with low trade volumes typically engage with only a few partners. This pattern reflects what the trade literature refers to as the gravity relationship: the idea that trade flows increase proportionally with the economic size of the origin and destination countries. As a result, larger countries tend to engage in higher volumes of trade, while trade between smaller countries is typically minimal or even absent.[11–13]

Figure 1: The Reporting Problem



Notes. This figure illustrates cross-country differences in trade reporting and the extent of discrepancies in reported values. Panels A and B display trade data reported by exporters and importers, respectively, for the year 2010. In both panels, countries are sorted in ascending order by total trade, with exporters on the vertical axis and importers on the horizontal axis. In Panel A, missing rows represent exporters that did not report data, while in Panel B, missing columns indicate non-reporting by importers. Panel C shows the result of combining both sources and highlights the level of discrepancy in bilateral trade flows where both exporter and importer reports are available. In this panel, flows with a discrepancy of less than 25% are shown in purple, while those with a discrepancy greater than 25% are shown in green. Discrepancies are calculated as $|\log(V^E_{e,i}/V^I_{e,i})|$, where V^E and V^I represent the values reported by the exporter and importer, respectively. Panel D summarizes these discrepancies over time, from 1962 to 2023, highlighting the evolution of trade reporting practices and reporting gaps. In this panel, each potential bilateral flow (whether observed or not) is weighted by its expected value based on a simulated gravity model.[11]

What is striking in Panels A and B is the presence of completely blank rows and columns scattered throughout the matrices, indicating that some countries did not report any trade with their partners. In Panel A, the absent rows correspond to countries that did not report any exports to UN Comtrade, while in Panel B, the absent columns indicate countries that did not report their imports. If researchers relied solely on data reported by either exporters or importers, they would often be unable to observe which countries trade and with whom.

What happens when we compare the values reported by exporters and importers for the same transactions? Panel C of Figure 1 displays the level of discrepancy between reported values for each country pair, calculated as the absolute log difference, i.e., $|\log(V^E_{e,i}/V^I_{e,i})|$. For visualization purposes, trade flows are classified into five categories. First, in white, we depict cases where neither country reported any bilateral trade. In 2010, approximately 50% of all possible country pairs fell into this category. Second, the flows shown in blue and orange represent cases where trade was reported only by the exporter (9.3%) or only by the importer (10%), respectively. Third, trade values were reported by both partners for roughly 31% of country pairs. Within this group, we distinguish between two subcategories based on the (log) level of discrepancy: low and high. We set the threshold between these categories at 25%, a level sufficiently large to account for statistical differences as well as additional costs such as freight and insurance. Notably, nearly two-thirds of the jointly reported trade flows exhibit discrepancies greater than 25%, shown in green.

This pattern of discrepancies is not unique to 2010; it reflects longer-term trends. Panel D of Figure 1 shows the evolution of the five reporting categories from 1962 to 2023. In this figure, every potential bilateral flow between countries is weighted by predicted trade values from a gravity model. Each potential bilateral trade flow is weighted by its expected value derived from a simulated gravity model (following Section 3.6 and the companion code of Head and Mayer[11]). This approach assigns greater relevance to flows between large trading countries and provides meaningful weights for absent trade, as reflected in zero trade flows observed in the data. An alternative approach—weighting each bilateral flow equally—produces a broadly similar picture, with the main difference being that the equal-weights method gives greater relative importance to the share of flows with zero reported trade. In

Panel D, we also restrict the sample to countries present throughout the entire period to avoid distortions caused by the entry of newly formed states, such as the former Soviet republics, which did not report trade data independently in earlier years.

As global trade has expanded over time and countries have engaged with more partners, the share of non-reporting country pairs has declined. Throughout the period, trade flows reported by importers have consistently accounted for a larger share than those reported by exporters. Among observed trade flows, the proportion of cases reported only by the exporter or only by the importer has also decreased. The most striking pattern in Figure 1 is the rise in flows with large bilateral discrepancies (shown in green). Despite better coverage, the fraction of cases in which the two reports diverge by more than 25 percent has grown, underscoring that inconsistencies in reporting remain a persistent and worsening problem. This trend highlights the value of datasets that explicitly reconcile partner reports and provide an improved, dataset of bilateral trade.

Overall, Figure 1 highlights a key challenge in working with bilateral trade data: for any given flow, researchers must often choose between missing data, when only one partner reports, and conflicting data, when both partners report but with large discrepancies. The best-case scenario, in which both partners report and the discrepancy is small, applies to only about 10% of all country pairs. This figure underscores that combining sources and improving the consistency of trade data is a central concern when analyzing bilateral flows. The problem becomes even more pronounced at the product level, where reporting discrepancies are compounded by differences and inconsistencies in classification systems across countries. These challenges grow increasingly complex over time as classification systems are frequently updated—an issue we explore in the following section. In Section 2.2, we detail our approach for combining reports from both trading partners to construct a more complete and consistent picture of global trade flows.

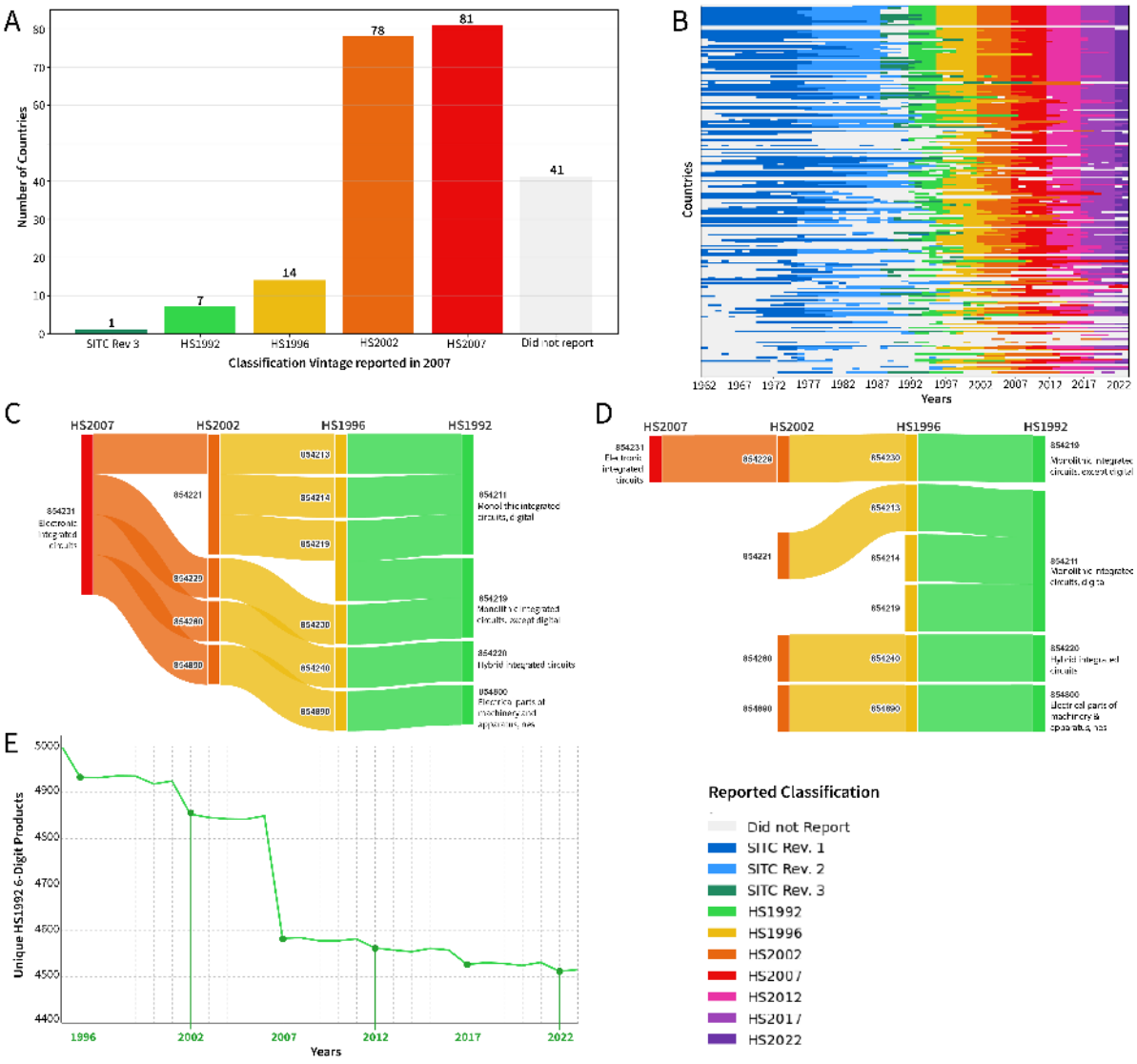
1.2 The Concordance Problem

Bilateral trade is reported at the product level, with each reporting country recording data annually using a single product classification version. There are two main *classification systems*:

the Standard International Trade Classification (SITC), with its first *vintage* ("Revision") introduced in 1962, and the Harmonized System (HS), which became the global standard in the early 1990s. SITC and HS serve distinct purposes and are designed with different objectives in mind. SITC, developed by the United Nations, is structured to facilitate economic analysis and emphasizes product characteristics such as material composition, stage of processing, and end use. This makes it particularly useful for studies on structural change and long-term trade patterns. In contrast, HS, the current standard used by most countries, is defined and regularly updated by the World Customs Organization (WCO). It is primarily designed to facilitate and standardize customs administration and tariff classification. HS organizes products based on their physical and technical attributes to ensure consistency in the reporting and monitoring of international trade transactions. As new technologies emerge and trade practices evolve, new *vintages* of the classification systems are introduced (e.g., HS1992, HS1996, HS2002). Approximately every five years, the World Customs Organization (WCO) releases a new product classification *vintage*. Product classifications change to reflect factors such as technological progress, practical considerations (e.g., improving ambiguous classifications or bundling product components), and policy needs in areas such as environmental and social monitoring.[2]

These revisions introduce two major challenges for trade data analysis.[2] First, in any given year, trading partners may report their bilateral trade using different vintages of the classification system, leading to inconsistent product categorization across countries in cross-sectional analysis. Second, as new vintages are introduced over time, they often split $(1:n)$, merge $(m:1)$, or generally reassign $(m:n)$ products, complicating the comparability of product-level trade data across years. Although correlation tables are provided to harmonize vintages, they describe only the mappings between product codes and do not provide the economic weights needed to accurately distribute traded values when products are split or merged. This critical gap remains largely unaddressed in existing datasets.

Figure 2: The Product Concordance Issue



Notes. Panel A shows the number of countries that reported trade using each classification vintage in 2007, an illustrative year, in which the Harmonized System (HS) underwent a significant revision. Panel B presents the timeline of classification adoption across countries, with each row representing a country and indicating the years in which it reported using each HS vintage. Panel C illustrates, using a specific example, the links in the WCO correlation table connecting three HS2007 product codes to their HS1992 counterparts. The example focuses on code 854231 in HS2007, "Electronic integrated circuits," which maps into four products in HS1992. Panel D displays the subset of these links that are retained in UN Comtrade's default concordance,[1] highlighting that some codes receive no value allocation under Comtrade's method. Panel E shows the number of unique HS1992 product codes in U.S. imports over time in UN Comtrade's harmonized dataset, with vintage release years highlighted.

Figure 2 illustrates the challenges posed by differences in classification vintages across countries and over time. Panel A shows the distribution of classification vintages used by countries in 2007, the year in which the WCO released the HS2007 revision, a major update

that introduced numerous changes, including the splitting and merging of product codes. The data reveal substantial heterogeneity in classification usage: in 2007 alone, countries employed five different vintages to report trade, ranging from the newly released HS2007 to the much older SITC Rev. 3. This variation makes direct trade comparisons between reporters impossible without proper harmonization. Panel B displays the adoption patterns of classification systems over time, with each pixel representing a country-year observation colored by the vintage used to report trade statistics. In essence, Panel A provides a cross-sectional summary for the year 2007 of what Panel B shows longitudinally. These figures highlight that once a new vintage is introduced, it often takes several years to achieve widespread adoption, resulting in persistent heterogeneity in the classification systems used across countries. Although delays were especially pronounced during the SITC era (i.e, 1962–1995), the figure shows that the introduction of the Harmonized System (HS) brought more frequent updates—yet staggered adoption remains a prevalent issue. As a result, in any given year, countries report trade figures using different vintages, and sometimes even different classification systems, further complicating comparisons between trade partners.

Correlation tables provide the mapping of product codes across classification vintages. We illustrate this using the example of "electronic processors and controllers with integrated circuits" (code 854231 in HS2007), which accounted for approximately 1% of global trade in 2007. The two Sankey diagrams in Figure 2 (Panels C and D) compare the WCO's official product code mappings (Panel C) with the practical implementation used by UN Comtrade (Panel D).[1] The Sankey diagrams show that reconciling data reported under HS2007 with an older classification, such as HS1992, requires multiple intermediate steps. Here, for simplicity, we focus on the conversion from HS2007 to HS1992. The same conceptual issues arise when using more recent classifications, such as HS2022, though the number of required steps increases, adding too many layers in the visualization. Panel C reveals the complex many-to-many relationships defined in the WCO's official correlation tables. In this example, code 854231 maps to several product codes in earlier vintages and ultimately splits into four distinct 6-digit HS1992 codes. While the WCO's correlation tables specify the links between product vintages, they do not provide guidance on the weights that should be used to allocate trade values when

harmonizing across classifications.

An often-neglected issue is that, without conversion weights being provided, when performing the conversion Comtrade maps each product in a classification vintage to a single product code in an adjacent classification vintage, and does not allow for $1:n$ or $m:n$ mapping relationships.[1,2] Panel D shows that Comtrade maps product code 854231 in HS2007 to a single HS1992 product code, 854219: "Other monolithic digital integrated circuits (including bipolar and MOS technologies)." Comparing Panels C and D highlights the striking impact that classification correlations have on trade statistics. In our illustrative example, the current practice used by UN Comtrade[1] results in only one product code being retained, effectively disregarding the three additional codes included in the WCO's official mapping.

Panel E illustrates one of the consequences of Comtrade's approach. Comtrade's practice of enforcing a $1:1$ mapping creates a compounded effect when assembling and harmonizing historical trade data.[2] For example, displaying trade data from 1995 to 2023 using HS1992 requires chaining multiple correlation tables (HS2022 $\rightarrow$ HS2017 $\rightarrow$ HS2012 $\rightarrow$ HS2007 $\rightarrow$ HS2002 $\rightarrow$ HS1996 $\rightarrow$ HS1992). Each conversion step drops product codes due to omitted mapping relationships, and these losses accumulate. Consequently, Comtrade's 2022 data converted to HS1992 contains only around 4,500 codes—approximately 500 fewer than the $\sim$5,000 products in the complete HS1992 vintage. Panel E illustrates this cumulative loss: the more intermediate vintages there are between the target classification and the original data, the more products are dropped. Our conversion method, explained in detail in Section 2.3, overcomes this systematic dropping of product codes.

2 Methods: What Do We Do?

We introduce a new collection of trade datasets that address the limitations of raw trade data by systematically mirroring bilateral flows and improving harmonization procedures. Our comprehensive data pipeline begins with automated data collection using a Comtrade downloader that fetches country trade reports across all available years from the UN Comtrade API.

Figure 3 sketches our pipeline and procedure. Panel A illustrates the mirroring stage, which addresses asymmetric reporting by reconciling conflicting trade values between partners using reliability-weighted averaging. The final output is bilateral, product-level trade data that are both harmonized to a consistent classification vintage and reconciled across trading partners. Panel B illustrates the conversion method, which solves the classification harmonization problem identified earlier. Rather than relying on Comtrade's 1:1 mappings, which omit complex product relationships, this component generates data-based conversion weights when new classification vintages are introduced (approximately every five years). The pipeline then converts raw Comtrade data to the requested classification vintage using these generated weights, preserving the many-to-many relationships that official correlation tables specify but do not quantify. The data harmonized through the process sketched in Panel B serve as the input for the mirroring outlined in Panel A.

2.1 Data Sources and Use

Before describing the methodology in detail, we document the sources of raw data used to construct the bilateral trade data and the mirroring process. These are categorized into three parts:

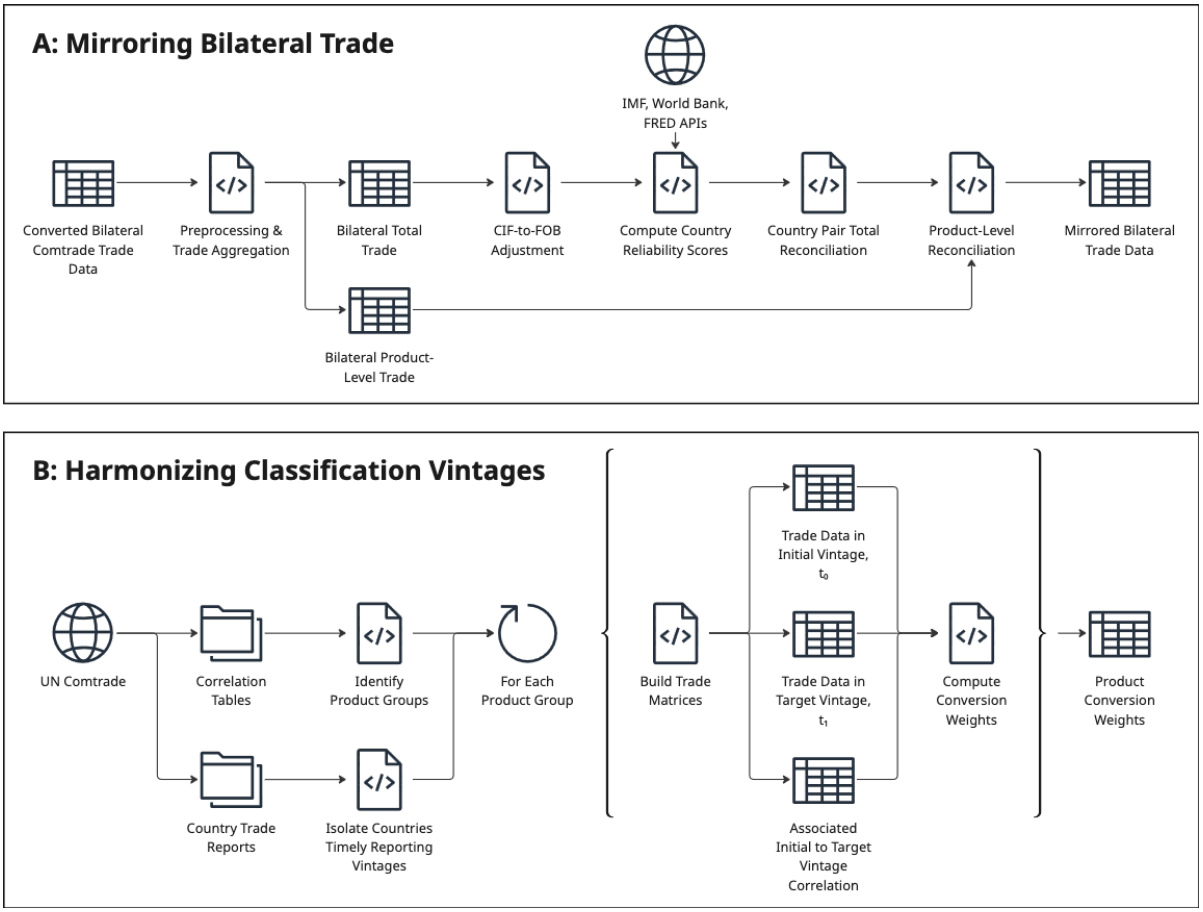
- **Primary Trade Data Sources:** Raw bilateral trade data extracted from UN Comtrade API (`https://comtrade.un.org/api/`) requiring an authentication key covering all import and export records reported by member countries from 1962 to the most recently available reported data available in Comtrade.[14]

- **External Reference Data:** Geographic and economic adjustment data sourced from established databases: (1) CEPII GeoDist providing bilateral distances, contiguity indicators, and landlocked classifications for countries;[15] (2) International Monetary Fund Direction of Trade Statistics;[16] (3) World Bank World Development Indicators;[17] (4) Federal Reserve Economic Data (FRED) accessed via API that requires an authentication key.[18]

- **Supporting Classification Data:** World Customs Organization correlation tables pro-

vided by UN Comtrade linking product codes across classification revisions,[19] supplemented with manually verified mappings for missing connections based on product description analysis and trade pattern consistency.

Figure 3: Data Pipeline Processes



Notes. These diagrams describe key portions of our overall data pipeline, highlighting important data sources, processing steps, and outputs. Panel A provides an overview of the bilateral trade data mirroring process. Panel B explains the process by which we generate the product conversion weights using the Lukaszuk-Torun (LT) method.[2]

2.2 Mirroring Bilateral Trade

Mirroring refers to the practice of combining trade data reported by both partners in a bilateral trade flow to improve the quality of trade statistics. Our approach addresses data coverage gaps for non-reporting countries and estimates trade values in cases where discrepancies exist between the reports of two trade partners. Mirroring techniques have long been used to

compare or enhance the accuracy of international trade data.[20–38] As illustrated in Panel A of Figure 3, the bilateral mirroring process consists of five computational steps:

Step 1: Pre-processing and trade aggregation

"As-reported" bilateral trade data from Comtrade is preprocessed to address data quality issues and standardize reporting formats. The process filters data by trade flow types and product classification levels, standardizes country codes, and integrates world-level trade totals. An important data quality step involves identifying countries that report a large percentage of trade with an unknown partner, referred to as Areas Not Specified (ANS). When the ANS ratio exceeds 25% of a country's total trade, we subtract ANS from the country's reported trade value to avoid double counting. We assume that the unknown partner most likely reported the trade with country-level detail, and this flow will be captured in the subsequent mirroring step.

To avoid concordance issues in this preprocessing stage, we use data only at the aggregate importer-exporter level. This level of aggregation is sufficient for the next two steps, including the calculation of reliability scores. Once the reliability scores are computed, we revert back to product-level data for mirroring and final assembly.

Step 2: CIF-to-FOB adjustment

Although Comtrade recommends that countries report their trade data—particularly imports—in both free on board (FOB), and cost, insurance and freight (CIF) terms, only a handful of countries actually submit data that fulfill this requirement. Consequently, when comparing the same trade flow as reported by the exporting country (i.e., FOB values) and the importing country (i.e., CIF values), an adjustment is required to account for the portion attributable to insurance and freight in the CIF value to ensure comparability. In the absence of systematically reported data, we are therefore forced to estimate and impute this portion of the CIF values.

Our approach uses reported data from importing countries that provide both FOB and CIF values to estimate the transportation cost component. We then apply these estimates to

adjust the CIF values for countries that do not report both measures. This approach is similar to the CIF/FOB adjustment implemented by CEPII in the construction of the BACI international trade database.[34] A key difference is that we estimate the adjustment regression at the aggregate level, without disaggregating by product. This choice significantly reduces the dispersion and potential noise in the estimates, which can arise from product-level idiosyncrasies and reporting inconsistencies.

The implicit assumption is that transportation costs are proportional to the distance between countries, and that risk-related factors—such as those affecting insurance or security costs—are constant across products and can therefore be captured through fixed effects in a regression framework. To this end, we estimate the following regression:

$$\ln(CIF/FOB)_{e,i} = \alpha + \tau_1 \times \ln(\text{dist})_{e,i} + \tau_2 \times \text{contiguity}_{e,i} + \lambda_e + \lambda_i + \epsilon_{e,i},$$

where α is a constant, $\ln(\text{dist})_{e,i}$ is the log of distance between exporter e and importer i, the binary $\text{contiguity}_{e,i}$ variable captures whether countries e and i share a border, λ_e is an exporter fixed effect, λ_i is an importer fixed effect, and $\epsilon_{e,i}$ is the error term.

Using the estimated parameters, particularly τ, which captures how trade costs scale with distance, we predict CIF/FOB ratios for all countries, including those that do not report imports on an FOB basis. We then apply these predicted ratios to adjust the CIF values reported by importers, yielding estimates that are comparable to the FOB values reported by exporters for the same flows. Two constraints are imposed on the adjusted ratios: they must be non-negative, and they are capped at a maximum of 20%, which limits the influence of potential outliers. Given the estimated distance coefficients, this threshold approximates the maximum plausible CIF/FOB ratio for trade between the most distant country pairs. In practice, the cap applies to fewer than 1% of observations.

Step 3: Compute country reliability scores

Country reporting reliability scores recover an indicator of the accuracy of each country's reported trade data, based on observed discrepancies. While all reported values may contain

some noise, some reporters are systematically more reliable than others. When estimating these reliability indicators, it is essential to account for network effects. We broadly define network effects as incorporating the reliability and accuracy of a country's trade partners when computing its own score.

For example, suppose there is a large discrepancy between the trade values reported by countries e and i. This could arise if (i) country e is an unreliable reporter, (ii) country i is an unreliable reporter, or (iii) both countries are unreliable reporters. If (i) is true, then the values reported by e will show systematic discrepancies with all of its other trading partners. The same logic applies to country i in case (ii), and to both countries in case (iii). By integrating information about each country's own discrepancies and the discrepancies observed in their partners' trade reports, we can construct a reliability score that distinguishes among these cases. Consequently, countries that report more consistently relative to their partners and that maintain broader, more interconnected trade networks receive higher reliability scores.

Therefore, the main idea is to exploit the fact that countries report trade with many different partners. By examining the overall mismatch in reported trade values across all bilateral pairs, we develop a measure of each country's reporting accuracy. To illustrate this intuition, consider a simple example. Suppose that country A is known to report accurate trade data. Then, any discrepancies between its reported values and those of its partners reflect inaccuracies on the part of the partner countries. By aggregating these discrepancies across all country pairs, we can iteratively refine our estimates and arrive at a consistent measure of trade reporting accuracy for each country.

First, we define a measure of discrepancy between the trade values reported by exporting and importing countries. Suppose country j exports goods to country k, with a true (but unobservable) trade value denoted by V. In practice, we observe the value reported by the exporter, V_j^e, and the value reported by the importer, V_k^i. We define the bilateral trade discrepancy between countries j and k as:

$$D_{j,k} = \frac{|V_j^e - V_k^i|}{V_j^e + V_k^i} \in [0, 1]. \tag{1}$$

In equation (1), the numerator of the ratio captures the absolute difference between the trade values reported by the two countries, while the denominator is the sum of their reported values. As a result, the discrepancy measure ranges from 0, when both countries report identical values, to 1, which occurs when only one country reports the trade transaction. However, discrepancies may arise due to inaccuracies in reporting by either country j or country k. Therefore, to assess how accurately a country reports its trade data, we must account for the reporting accuracy of its trading partners. Specifically, evaluating the accuracy of exporters (the j's) requires incorporating information on the accuracy of importers (the k's), and vice versa.

We choose to decompose the discrepancy $D_{j,k}$ into components attributable to each country in the pair, in order to identify the underlying source of reporting error. The simplest approach is to estimate the following regression:

$$D_{j,k} = \alpha_j + \alpha_k + \varepsilon_{jk}, \tag{2}$$

where α_j and α_k are j- and k-specific dummy variables that represent the reporting inaccuracy of countries j and k, respectively. This specification has an intuitive interpretation: if $D_{j,k}$ is large and country k is known to report trade accurately (i.e., has a low α_k), then the discrepancy is likely due to inaccuracies in country j's reporting. Conversely, if k has low reporting accuracy (a high α_k), then the discrepancy may not reflect poor reporting by j. To ensure consistency with the idea that each country independently contributes to discrepancies, we impose a non-negativity constraint on all α parameters. Negative estimates would suggest that a country reduces overall discrepancies when included as a trading partner, an interpretation that is inconsistent with the assumption that countries report their trade independently of their partners.

The regression in equation (2) incorporates the network structure of international trade. This trade network is defined as a graph in which nodes represent countries and edges represent trade flows between them. Specifically, an edge from exporter j to importer k, denoted by $x_{j\to k}$, corresponds to a reported trade flow. For simplicity, we refer to a generic trade flow

edge with x, with the understanding that each edge is uniquely associated with a direction and a pair of countries: an exporter and an importer. Let N_x denote the total number of trade flows (i.e., bilateral trade transactions), and N_c the number of countries. Since each observation in equation (2) corresponds to a trade edge, we can re-index the model in terms of edges as follows:

$$D_x = \alpha_{x(exporter)} + \alpha_{x(importer)} + \varepsilon_x.$$

This equation can be written in matrix form as follows:

$$\mathbf{D} = \mathbf{B}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}, \tag{3}$$

where $\mathbf{D}$ is the vector whose x^{th} element is D_x (with x denoting the edge between nodes j and k), and $\boldsymbol{\alpha}$ is the vector of country-specific reliability parameters α_j. $\mathbf{B}$ is a matrix of size $N_e \times N_c$ that maps edges (trade transactions) to nodes (countries, either exporters or importers), with elements defined as:

$$\begin{aligned} B_{x,j} = B_{x,k} &= 1 \quad \text{if edge } x \text{ is between nodes } j \text{ and } k, \\ B_{x,j'} &= 0 \quad \text{if edge } x \text{ does not connect to node } j'. \end{aligned}$$

In other words, $B_{x,j}$ and $B_{x,k}$ are equal to 1 if edge x represents a trade transaction between exporter j and importer k. Note that there may also be another edge x' corresponding to the reverse transaction—from exporter k to importer j—which would have the same row values as x in the matrix $\mathbf{B}$. We estimate equation (3) using ordinary least squares (OLS) by expressing the model in matrix form:

$$\hat{\boldsymbol{\alpha}} = (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{D} + \boldsymbol{\varepsilon}. \tag{4}$$

Here, $\mathbf{D}$ is the vector of observed discrepancies, $\boldsymbol{\alpha}$ is the vector of country-specific accuracy parameters, and $\mathbf{B}$ is a binary incidence matrix linking edges to countries in their roles as exporters or importers.

Let us now analyze what the estimation of equation (4) reveals about the structural elements of the international trade network. The term $\mathbf{B}'\mathbf{D}$ returns a vector of length N_c, where

each element j represents the sum of all trade discrepancies in which country j is involved. The matrix $\mathbf{B}'\mathbf{B}$ is a square matrix of size $N_c \times N_c$ that is closely related to the adjacency matrix of the trade network. Formally, let us define $\mathbf{A} \equiv \mathbf{B}'\mathbf{B}$ to express this relationship explicitly. The elements of the matrix $\mathbf{A}$ are:

$$
\begin{aligned}
A_{j,k} &= 2 \text{ if both } j \text{ reports exports to } k \text{ and } k \text{ reports exports to } j \text{ and } j \neq k, \\
A_{j,k} &= 1 \text{ if only one of } j \text{ reports exports to } k \text{ or } k \text{ reports exports to } j \text{ and } j \neq k, \\
A_{j,k} &= 0 \text{ if neither } j \text{ reports exports to } k \text{ nor } k \text{ reports exports to } j \text{ and } j \neq k, \\
A_{j,j} &= \sum_{k \neq j} A_{j,k}.
\end{aligned}
$$

The $(j,k)^{\text{th}}$ element of matrix $\mathbf{A}$ is equal to 2 if both reported values are available for j exporting to k and k exporting to j, and it is equal to 1 if only a single trade transaction is reported between j and k. Zero elements in this matrix indicate the absence of reported trade transactions between the corresponding country pairs. The matrix $\mathbf{A}$ is symmetric, and the sum of its off-diagonal elements in any row or column equals the corresponding diagonal element. Notably, $\mathbf{A}$ is closely related to the Laplacian matrix of the trade network, except that the signs of the off-diagonal elements are not negated in this formulation. With these insights, we see that the OLS estimator effectively traces discrepancies back to individual countries by inverting a matrix derived from the structure of the trade network—namely, $\mathbf{A}$.

A limitation of the OLS estimation is that, when attempting to estimate all α values simultaneously, the matrix $\mathbf{A}$ becomes singular and non-invertible. This occurs because each diagonal element equals the sum of the corresponding row or column's off-diagonal elements, making $\mathbf{A}$ rank-deficient. A common solution is to fix the α value of one country to zero and interpret the remaining estimates relative to this baseline. However, this raises the question: which country should be chosen as the reference? To address this, we implement an iterative strategy: we perform the OLS estimation multiple times, each time omitting a different country as the baseline (i.e., setting its α to zero). We then evaluate the goodness of fit using the R^2 statistic and select as the baseline the country that yields the highest predictive power.

Note that we estimate a single reliability score per country. That is why we cannot simply

select the minimum estimated value and subtract it from each reliability score. If instead the equation was $D_{j,k} = \alpha_j - \alpha_k + \varepsilon_{jk}$, then subtracting the same value from every estimate would not change the result. In our equation, we also do not observe every (j, k) pair. Therefore, it is not clear how the choice of normalization propagates through the network. That is why we re-run the estimation by selecting a different base country every time, setting its value to zero, and ultimately maximizing the R^2 statistic.

A second limitation arises from our requirement that the α values be non-negative. To enforce this constraint while maintaining computational efficiency, we simply round any negative predicted α values to zero. We deliberately avoid non-linear estimation methods with inequality constraints in order to keep the computation tractable.

Algorithmically, our procedure consists of the following steps:

1. Select first country as the base country.
2. Run the OLS estimation.
3. Convert all negative estimated values to 0.
4. Calculate the R^2 of the estimation.
5. Go back to step 1 and select the next country as the base country. Repeat Steps 1-4 for all countries.
6. Determine the country that gives the highest R^2 as the ultimate base country.
7. Report the accuracy of countries as one minus the estimated values of αs.

Step 4: Country pair totals trade reconciliation

Bilateral trade weights are calculated using a softmax transformation that converts each country pair's exporter and importer reliability scores into complementary probabilities. These are country-pair-specific measures that determine how to combine conflicting reports for individual bilateral trade flows. For each flow, a higher exporter reliability relative to importer reliability results in a greater weight being assigned to the export-reported value, and vice versa.

Country reliability thresholds are set at the 10th percentile of the reliability score distributions for both exporter and importer roles. Countries with reliability scores above these thresholds are classified as reliable reporters in their respective roles, while those below are flagged as less reliable. This weighting system acts as a quality filter, prioritizing trade value estimates from more reliable reporting countries in the subsequent estimation process and systematically reducing the influence of the least reliable 10% of reporters on the final reconciled trade values. In practice, when combining the reports of two trading partners, we disregard the data from unreliable reporters and rely solely on the values reported by the more reliable partner.

The core idea behind most mirroring methods is to combine the two reported values using a weighted average: $V^F_{e,i} = (1 - w_{e,i})V^E_{e,i} + w_{e,i}V^I_{e,i}$ where $V^E_{e,i}$ is the value reported by the exporter, $V^I_{e,i}$ is the value reported by the importer, and $w_{e,i} \in [0, 1]$ is the weight assigned to the importer's report. In this formulation, weights $w_{e,i}$ are specific to each country pair, and we abstract from the product dimension for simplicity. While mirroring methods differ in the way weights are computed, the general idea is to assign more relevance to partners that are more reliable in their reporting. A comparison of different methods used in the literature to compute the mirroring weights using synthetic data highlights this method as the best performing.[3] The weights we use in our mirroring procedure correspond to the reliability scores—specifically, one minus the estimated values of α, as described above in Step 3.

However, mirroring has important limitations: the approach is information-intensive and relies on data availability from both trading partners. As shown in Figure 1, a substantial share of trade flows are affected by missing information. In recent years, mirroring has been applicable to only about 30% of possible bilateral flows (i.e., country pairs), while for another 30%, data is available from only one reporter—either the exporter or the importer. In these cases, where only a single observation is available, the procedure effectively reduces to assigning full weight to the available report from Comtrade. This is equivalent to giving full weight to the reporter when only one reporter is available and the reporting country is found to be reliable.

Step 5: Product-level trade reconciliation

Following country-level reconciliation, these reliability-weighted estimates are applied to disaggregated product-level trade data (e.g., 6-digits in HS). Each bilateral trade flow is disaggregated, with country-level weights and reliability scores preserved across all products traded between the same country pair. The final product-level values are calculated using identical hierarchical rules, ensuring consistency between aggregate and disaggregated estimates.

A reweighting procedure reconciles any discrepancies between the sum of product-level estimates and the country-level totals. Trade flows are proportionally adjusted to match reconciled country totals while preserving relative product composition. Large unexplained discrepancies that exceed predetermined thresholds ($>$ 20% deviation or $>$\$25 million absolute difference for flows $>$\$100 million) are classified as "trade data discrepancies" and assigned to a separate commodity code "$XXXX$" to maintain accounting consistency. As a result, the datasets at the total level and at the product level show the same aggregate values of trade.

2.3 Harmonizing Classification Vintages Using Weights Based on Observed Trade Data

We harmonize the numerous product code vintages implementing the data-driven Lukaszuk-Torun method[2] (from hereafter referred to as the LT method). As discussed above, when a new vintage is introduced (e.g., changing to HS2007 from HS2002), we need to convert one revision into other previous vintages to compare disaggregated product-level trade flows over time. The method is based on the empirical regularity that product-level trade flows are highly persistent over time. The idea is simple: If imports of a product look very similar from one year to the next when there are no classification updates, this should also apply when there is a classification update. For this purpose, we calculate conversion weights that tell us how to harmonize product codes; for instance, how much of the trade value reported in a given 6-digit product of the HS2007 vintage belongs to a product code in the HS2002 version.

The core objective is to infer these implicit weights from observed trade data by comparing the distribution of trade flows reported under the old classification in a given year with those reported under the new classification in the following year. Conceptually, the method relies on

two key assumptions: (i) trade patterns do not change significantly between the two adjacent years, and (ii) customs authorities accurately apply the old and the updated classification systems.

An alternative approach to address the issue of classification changes is to reclassify all the product codes connected directly or indirectly via concordance tables into a common synthetic code,[39] which has have been adopted in several studies.[40–42] However, this approach aggregates trade data into substantially fewer products when applied to multiple revisions over time, with a single synthetic cluster covering roughly a quarter of total trade when converting across long time horizons.[2]

As illustrated in Panel B of Figure 3, the algorithm to generate the conversion weights consists of three main steps.

Step 1: Group products into networks using mapping from correlation tables

First, we identify groups of product codes that are interconnected through classification changes according to the unweighted correlation tables provided by the WCO. These groups represent networks where codes from the source vintage map to codes in a target vintage through various relationship types (1:1, $1:n$, $m:1$, and $m:n$). Treating these clusters as integral groups is crucial: the full value of every code in the group must be carried through the concordance to preserve consistency when translating trade values from one vintage to another. Omitting even one code could break the internal accounting identity and distort the conversion weights, as it will be clear below. The routine only considers groups containing unknown conversion weights, as groups with only predetermined weights ($1:1$ or $m:1$) require no estimation. A group of products is a subset of the WCO's correlation table, providing sufficient information to represent the entire mapping of all products within a group. Panel C of Figure 2 illustrates a subset of a product group since not all products needed to map all represented products are present.

For a handful of product codes, the WCO correlation tables provide no links (especially in the HS1996-HS1992 conversion). For these codes, we simply connect the 6-digit code to every other 6-digit code that shares its 4-digit root. The weighting procedure then drives irrelevant

links toward zero. If no alternative 6-digit code exists under the same 4-digit heading, we manually assign the closest match based on product descriptions, which was done for five products (synthetic fibres for SITC1 6516 to SITC2 6514, gold jewlery for SITC1 897 to SITC2 9710, finishing agents & dye carriers for HS1992 380999 to SITC3 5989, binapacryl for HS2007 291636 to HS2012 291616, and petroleum oils for SITC3 334 parent code mapped to HS1992 parent code 2710). Importantly, these cases affect only a few products and have a negligible impact on the overall routine.

Step 2: Build trade matrices for timely country reporters by product groups

Second, we prepare the data input for the optimization routine (the third step below). We use the total imports V of product k by a country i. To capture product hierarchies within groups and control for group-specific trends, we scale each product-level import by the total trade within that group

$$v_{i,k} \equiv \frac{V_{i,k}}{\sum_{\hat{i}} \sum_{\hat{k}} V_{\hat{i},\hat{k}}}.$$

This scaling is computed separately for each group and year, and the analysis focuses on countries that were timely reporters and switched to the latest available vintage in the years where there was a vintage update. This is a minor difference from the main specification in the LT method, which uses country-pair-level trade flows. However, the conversion yields very similar results when using importer-level data.[2] We adopt the latter for simplicity.

Step 3: Compute conversion weights

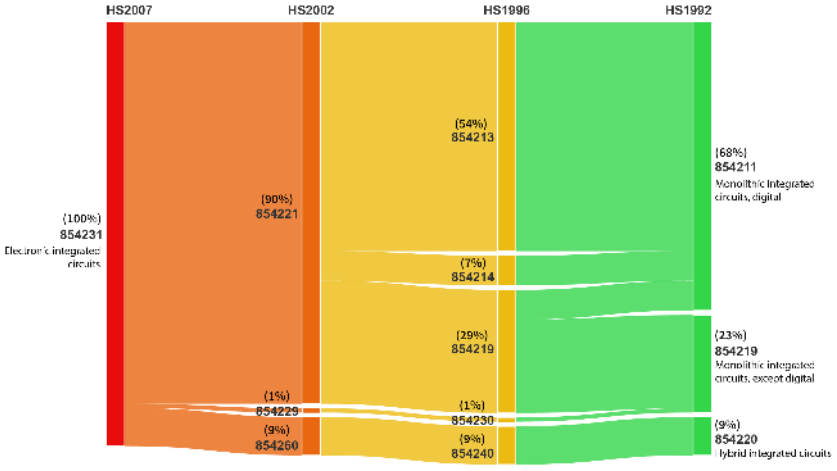
The third component is the core of the LT method. We fit trade flows reported in a target vintage (denoted by subscript s) using flows from an initial vintage (denoted by subscript k) through a constrained least-squares procedure. By construction, the trade flows reported in the different vintages are from different reporting years. We denote the year of the initial (target) vintage by t_0 (t_1). For instance, when calculating the weights to convert HS2007 into HS2002, we will have $t_0 = 2007$ (the year where HS2007 was implemented) and $t_1 = 2006$ (the last year in which HS2002 was the newest vintage). The optimization problem minimizes the

squared deviations between observed and predicted imports

$$\min_{\{\beta_{k,s}\}} \sum_s \sum_i \left(v_{i,s}^{t_1} - \sum_k v_{i,k}^{t_0} \beta_{k,s} \right)^2$$
$$\text{s.t.} \quad \beta_{k,s} \geq 0 \; \forall \; k, s \quad \text{and} \quad \sum_s \beta_{k,s} = 1 \; \forall \; k.$$

Here, $\beta_{k,s}$ represents the conversion weight from initial product k to target product s, indicating what fraction of trade in product k should be allocated to product s when converting between vintages. The constraints ensure that the weights are between zero and one, and that they sum to one for an initial product across target products. To apply the conversion weights, one has to multiply the trade flow reported in the initial HS vintage by the respective weight, and then sum up across initial codes. That is, $\widehat{v}_{i,s}^{t_0} = \sum_k v_{i,k}^{t_0} \hat{\beta}_{k,s}$, where hats denote estimates. For non-adjacent vintage conversions (e.g., HS2017 to HS2007 via HS2012), we multiply weights across intermediate conversions. Importantly, the algorithm estimates conversion weights separately for each group to reduce the computational burden and ensure that cross-group weights remain zero by construction.

Figure 4: An Illustration of Our Product Conversion Weights



Notes. This figure illustrates the allocation of trade value using the weights obtained by the LT method[2] when concording from HS2007 to HS1992, using code 854231 as an example. It shows how each unit value reported under the 2007 vintage is sequentially allocated across intermediate vintages using the weights obtained by the LT method, until it reaches the target classification in HS1992. The figure highlights the differences in outcomes compared to those shown in Figure 2 (Panels C and D).

To illustrate the use of the calculated weights and how they compare to other mapping

approaches, we revisit in Figure 4 the example of product 854231, "electronic processors and controllers with integrated circuits," previously shown in Panels C and D of Figure 2. The figure shows a sankey diagram of how the trade value reported under HS2007 (left) is distributed across earlier vintages using our method, ultimately reaching the target classification of HS1992 (right). The LT method preserves the official relationships specified in the WCO's correlation tables: product 854231 is mapped consistently across vintages, following the WCO-defined structure. In the sankey diagram, the width of each arrow reflects the magnitude of the calculated conversion weights. This figure conveys two key takeaways. First, the method does not assign positive weights to all possible links, as product 854800, for instance, receives none. Second, the conversion outcome diverges sharply from Comtrade's practice. Comtrade allocates the entire value (100%) to code 854219, which is "Monolithic integrated circuits, except digital," (see Panel D of Figure 2), whereas the LT method assigns only 23% to 854219 and directs the largest share (68%) to code 854211—"Monolithic integrated circuits, digital,", which Comtrade omits entirely. Given that the integrated circuits are mostly digital currently, this code assignment change is consequential. Overall, this is far from an isolated case: across the conversion from HS2007 to HS1992, we estimate that approximately 16% of world trade in the conversion is weighted differently under our method compared to Comtrade's default approach. This figure includes the 3% of world trade associated with codes that Comtrade's 1:1 mapping omits entirely as target codes in HS1992.

3 Data Records

The resulting conversion weights and bilateral trade datasets applying the methdology discussed in our methods section are available via Harvard's Dataverse. This dataset is updated regularly, at least twice per year, to incorporate the updated data.

The data provides conversion weights for all SITC and HS classification vintages along with bilateral trade data for SITC Rev. 2 and HS1992 classification vintages, as the former allows for six decades of product-level data, and HS1992 is the longest panel dataset following the current classification system.

Bilateral Trade Datasets

- **SITC Revision 2 (Standard International Trade Classification)**: Features four-digit product-level granularity covering 1976 –2023. Includes the reporting exporter, the reporting importer and their associated reported exporter and imports respectively, along with our imputed trade value in USD aggregated by year.
- **HS1992 (Harmonized System, 1992 Vintage)**: Features six-digit product-level granularity covering 1995–2023. Includes the reporting exporter, the reporting importer and their associated reported exporter and imports respectively, along with our imputed trade value in USD aggregated by year. Conversion Weights
- **Conversion weights for adjacent classification vintages:** provides the source classification's product code, target classification's product code, and the associated conversion weights.

4 Technical Validation

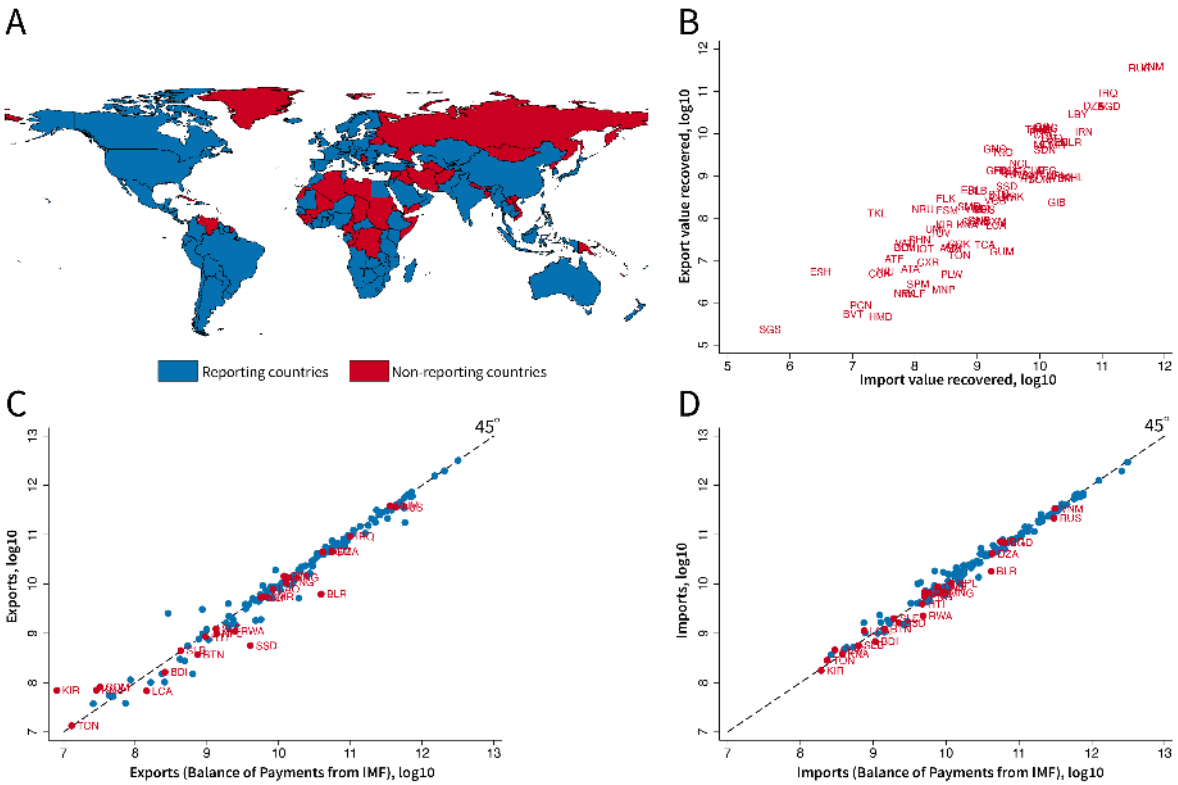
To assess the quality of our data, we perform three complementary validation checks. First, we compare our bilateral trade dataset with the International Monetary Fund's (IMF) Balance of Payments (BoP) data and show that the two series are highly correlated. Second, we evaluate the sector-level impact of mirroring and product conversion. Finally, we show how our mirroring and product-conversion procedures substantially increase the number of positive origin-destination-product observations relative to the raw UN Comtrade data.

Comparison with IMF BoP data

Figure 5 provides an overview of our recovered trade values and assesses their external validity. Specifically, we aggregate the recovered trade data at the country level and compare it to the IMF data.[43] The IMF uses total exports and imports as key variables in its assessments of the sustainability of a country's balance of payments. Although the IMF is not specifically concerned with the sectoral composition of trade and focuses primarily on aggregate val-

ues, it places strong emphasis on data accuracy. Reliable balance of payments trade figures—including exports and imports statistics—are essential for evaluating a country's ability to access foreign currency and meet external debt obligations.[44] Panel A displays a world map that highlights in red the countries for which we recover product-level trade statistics in 2023. The map shows that non-reporting countries are most prevalent in regions such as Africa; without the mirroring procedure, no product-level data would be available for these economies.

Figure 5: Data Recovered Through Our Mirroring Procedure



Notes. This figure presents the trade data recovered using our method for the year 2023 and compares it to other sources. Panel A displays a map highlighting the geographic distribution of recovered data, with countries shown in red where product-level trade information is available only through the mirroring process. Panel B presents a scatterplot of exports and imports (in logs) for these recovered countries. Panels C and D compare export and import values to those reported by the IMF in their Balance of Payments assessments (shown in blue), and highlight in red the countries for which product-level data are recovered through the mirroring process. The dashed lines in Panels C and D are 45-degree lines.

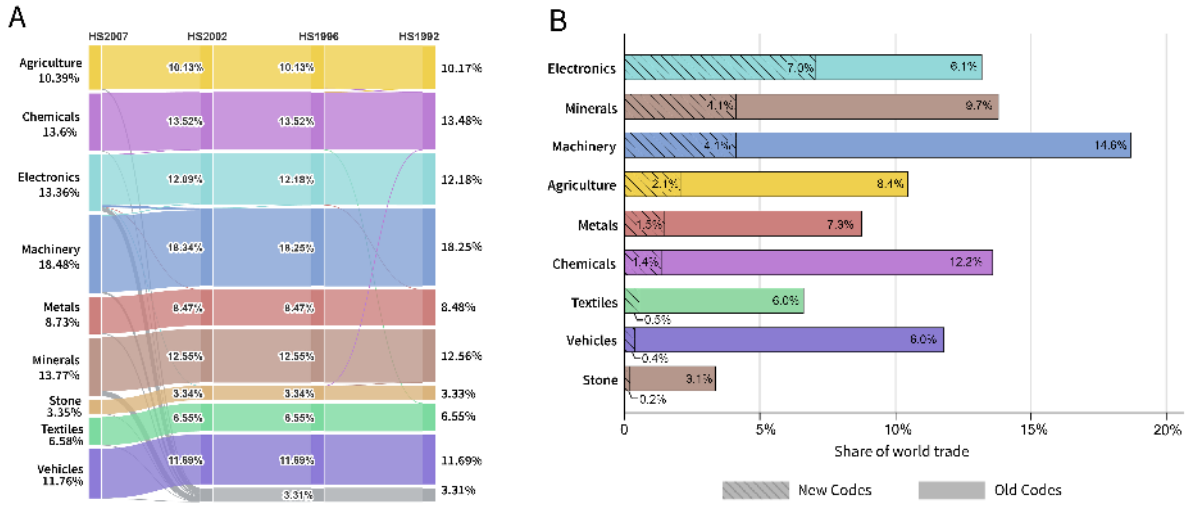
Panel B shows a scatterplot of log exports and imports for countries that did not report trade statistics to the Comtrade data repository. All of these values are recovered exclusively by mirroring the trade data reported by their trading partners. Our recovered export values (Panel C) and import values (Panel D) are highly correlated with the IMF Balance of Payments

(BoP) statistics across countries, with both sets of values being of similar magnitude. In both panels, reporting countries are shown in blue, while non-reporting countries—whose trade values are recovered through mirroring—are highlighted in red, consistent with the previous panels. Each panel includes a 45-degree line to facilitate visual comparison. The close alignment between our estimates and the IMF's BoP data—compiled independently by national statistical agencies—provides strong validation that our mirroring approach successfully recovers trade values.

Expected patterns at the sector level

Changes in the Harmonized System (HS) classification over time pose significant challenges for conducting consistent sector-level analyses of international trade. Product codes are frequently revised—split, merged, or reassigned—when new HS vintages are introduced, complicating the task of harmonizing trade data across years. We quantify the extent of these reclassifications and highlight their implications using two complementary visualizations in Figure 6. Together, they show that some sectors, particularly electronics, undergo substantial reallocation, while others are largely unaffected.

Figure 6: Extent of Sectoral Reallocation across Classification Vintages



Notes. Panel A shows how trade is redistributed across HS chapters when converting from HS2007 to HS1992, highlighting both cross-chapter shifts and misallocated trade under Comtrade's 1:1 concordance. Electronics experience the largest reallocation, dropping from 13.4% to 12.2% of world trade. Panel B quantifies the extent of code revisions by sector, with electronics again most affected: 7% of world trade in that chapter involves revised codes. In contrast, stones and minerals show minimal reclassification.

Panel A of Figure 6 presents a Sankey diagram that tracks the share of world trade in each HS chapter and how it is reallocated across chapters due to changes in classification. While most of the trade value in the initial classification ultimately maps back to the same chapter, the diagram reveals that a non-negligible share is redirected to different chapters. Notably, the electronics chapter undergoes the most extensive reallocation: under HS2007, it accounts for 13.4% of world trade in the year 2007, but this share falls to 12.2% after concordance to HS1992, as several codes are reassigned to other chapters, such as machinery. The visualization also includes a "misallocated trade" category, highlighting that approximately 3% of world trade in 2007 is not properly redistributed under Comtrade's 1:1 concordance method. This trade, according to our method, should be assigned to valid HS1992 codes that are omitted as target codes by Comtrade's 1:1 mapping practice, with the extent of trade affected calculated using our approach. As shown in the figure, most of this misallocation arises during the conversion from HS2007 to HS2002.

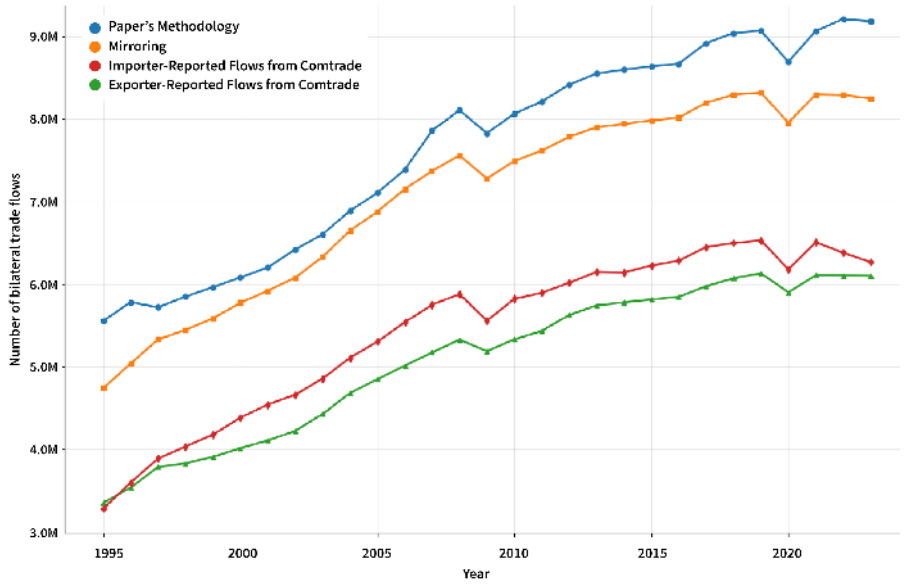
Panel B of Figure 6 illustrates the extent of code revisions by sector when converting from HS2007 to HS1992. The figure highlights the heterogeneity in the extent revisions, with chapters undergoing extensive changes, while others are minimally affected. For example, the electronics chapter, which accounts for approximately 12% of world trade in 2007, includes revised codes—defined as codes that were merged, split, or entirely redefined—that represent a value equivalent to 7% of world trade. One such case is product 854231, previously discussed, whose trade value is split across four distinct codes in HS1992. At the other end of the spectrum, the stones and minerals chapter is minimally affected by reclassification: only 0.2% of world trade within this sector is affected by concordance revisions, of a total sectoral share of 3% of world trade.

Expected patterns of positive trade flows by methodology

Finally, an important consequence of our mirroring and product-conversion method is that the number of positive trade flows increases. Mirroring recovers trade from non-reporting countries and the product conversion maintains a consistent set of products over time, avoiding the product diversity loss we identified with Comtrade. This means that our dataset contains

more observations in any given year compared to raw data from Comtrade. An observation is a country pair with a positive product-level trade flow.

Figure 7: Number of Observations over Time by Methodology



Notes. The figure compares the total number of positive trade flows (at the origin-destination-product level) based on four different approaches to processing bilateral trade data over the period 1995–2023. The mirroring and LT method[2] (blue line) and, hence, our dataset, recovers the most data points, identifying up to 9M positive trade flows compared to only around 6M flows captured by Comtrade's standard HS1992 datasets (green and red lines). This represents a 50% increase in data coverage through our approach. The mirroring-only approach (orange line) provides an intermediate gain, recovering approximately 2M additional data points beyond the Comtrade data per year.

This is precisely what we find in Figure 7, which tracks the number of observations over time. In recent years, our methodology results in around 9 million observations. There are 235 total possible reporters and 234 reciprocal trade partners (excluding the "Not Elsewhere Specified" partners). In the HS1992 vintage, there are about 5,000 possible products in any given year. As a result, if all countries traded all products with all other countries, we would see around 275 million observations. It is well-known that not all countries trade with every other country, and a country's export basket is typically a subset of all the available product codes. Thus, the roughly 9 million observed trade flows in recent years correspond to approximately 3.0% of the potential maximum number of observations. Furthermore, the consistent order of the four lines validates each component of our approach. Mirroring alone (orange line) systematically outperforms both single-reporter datasets, while the addition of Lukaszuk-Torun

harmonization (blue line) provides further substantial gains. This hierarchy demonstrates that each methodological component contributes meaningful value to data recovery without introducing spurious trade flows. The smooth temporal pattern of all methodologies, particularly our combined approach, indicates that we recover reasonable economic patterns.

5 Code Availability

The code used to acquire data from Comtrade, generate conversion weights, and mirror the bilateral trade data, respectively, is available for public use via GitHub in the following repositories:

- https://github.com/harvard-growth-lab/comtrade-downloader
- https://github.com/harvard-growth-lab/comtrade-conversion-weights
- https://github.com/harvard-growth-lab/comtrade-mirroring

Each of these packages is currently at their v1.0 release. These packages are written primarily in Python, with some functionality using R and Matlab. When used with a UN Comtrade API subscription, these packages enable anyone to reproduce these datasets with the most recent data available from Comtrade in SITC and HS classification vintages.

Acknowledgements

We would like to thank Timothy P. Cheston for help with the data validation and contributions to the Atlas of Economic Complexity. We would like to thank Mali Akmanalp and Romain Vuillemot for working on earlier versions of the Atlas data and architecture. We would like to thank the current and former members of the Harvard Growth Lab for continuous feedback on the data.

Author Contributions

Conceptualization: S.B., M.A.Y., R.H. *Methodology*: E.J., S.B., M.A.Y., D.T., P.L. *Formal Analysis*: S.B., E.J., D.T., M.A.Y. *Investigation*: S.B., E.J., D.T. *Data Curation*: S.B., E.J., D.T., B.L. *Software*:

S.B., E.J., D.T., B.L. *Validation*: S.B., E.J., B.L. *Visualization* S.B., E.J., N.T., B.L. *Writing - Original Draft*: S.B., E.J., D.T., M.A.Y. *Writing - Review & Editing*: S.B., E.J., D.T., B.L., M.A.Y., A.W. *Project Administration*: E.J. *Supervision*: M.A.Y., A.W., R.H. *Funding Acquisition*: R.H.

Competing interests

The authors declare no competing interests.

Correspondence

Correspondence should be addressed to Ricardo Hausmann (ricardo_hausmann@hks.harvard.edu) and Muhammed A. Yildirim (muhammed_yildirim@hks.harvard.edu).

References

1. UN Department of Economic and Social Affairs. *Correlation and conversion tables used in UN Comtrade* 2022. https://unstats.un.org/unsd/classifications/Econ/corr-notes/HS2022%20conversion%20to%20earlier%20HS%20versions%20and%20other%20classifications%20%20-%20v.1.0.pdf.
2. Lukaszuk, P. & Torun, D. *Harmonizing the Harmonized System* SEPS Discussion Paper 2022-12 (2022).
3. Bustos, S. & Yildirim, M. A. Uncovering trade flows. *Unpublished Mimeo* (2024).
4. Hausmann, R., Stock, D. P. & Yildirim, M. A. Implied comparative advantage. *Research Policy* **51,** 104143 (2022).
5. O'Clery, N., Yildirim, M. A. & Hausmann, R. Productive ecosystems and the arrow of development. *Nature Communications* **12,** 1479 (2021).
6. Bustos, S. & Yildirim, M. A. Production ability and economic growth. *Research Policy* **51,** 104153 (2022).
7. Bustos, S. & Morales-Arilla, J. Gains from globalization and economic nationalism: AMLO versus NAFTA in the 2006 Mexican elections. *Economics & Politics* **36,** 202–244 (2024).

8. Hausmann, R., Schetter, U. & Yildirim, M. A. On the design of effective sanctions: The case of bans on exports to Russia. *Economic Policy* **39,** 109–153 (2024).

9. Egger, P., Foellmi, R., Schetter, U. & Torun, D. Gravity with History: On Incumbency Effects in International Trade. *Journal of the European Economic Association* **jvae052** (2024).

10. Torun, D. *Quantifying the Extensive Margins of Trade and Production* Mimeo, Harvard Kennedy School. 2025.

11. Head, K. & Mayer, T. in *Handbook of International Economics* 131–195 (Elsevier, 2014).

12. Anderson, J. E. & Van Wincoop, E. Gravity with gravitas: A solution to the border puzzle. *American Economic Review* **93,** 170–192 (2003).

13. Yotov, Y. V. *Gravity at sixty: the workhorse model of trade* CESifo Working Paper No. 9584. 2022.

14. untradestats. *comtradeapicall* pipinstallcomtradeapicall==1.0.1. python package to access UN Comtrade's API. 2024.

15. Mayer, T. & Zignago, S. *Notes on CEPII's distances measures: The GeoDist database* Working Papers 2011-25 (CEPII, Dec. 2011). https://www.cepii.fr/CEPII/en/publications/wp/abstract.asp?NoDoc=3877.

16. International Monetary Fund. *IMF Data Mapper API* https://www.imf.org/external/datamapper/api/v1. REST API providing access to IMF economic indicators and datasets. Version 1.0. 2025.

17. World Bank. *World Bank Indicators API V2* https://api.worldbank.org/v2/country/all/indicator. REST API providing access to World Bank country indicators. 2025.

18. Federal Reserve Bank of St. Louis. *FRED© API* https://api.stlouisfed.org/fred/series/search. REST API providing access to FRED data. 2025.

19. UN Comtrade. *UN Comtrade Database* https://unstats.un.org/unsd/classifications/Econ. Correlation tables provided by United Nations Statistics Division. 2025.

20. Allen, R. L. & Berliner, J. S. *Soviet Economic Warfare* 1961.

21. Bhagwati, J. On the underinvoicing of imports. *Oxford Bulletin of Economics and Statistics* **27,** 389–397 (1964).

22. Naya, S. & Morgan, T. The accuracy of international trade data: the case of Southeast Asian countries. *Journal of the American Statistical Association* **64,** 452–467 (1969).

23. Sheikh, M. A. Smuggling, production and welfare. *Journal of International Economics* **4,** 355–364 (1974).

24. Yeats, A. J. On the accuracy of partner country trade statistics. *Oxford Bulletin of Economics and Statistics* **40,** 341–361 (1978).

25. McDonald, D. C. Trade data discrepancies and the incentive to smuggle: An empirical analysis. *Staff Papers* **32,** 668–692 (1985).

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

27. Rozanski, J. & Yeats, A. On the (in) accuracy of economic observations: An assessment of trends in the reliability of international trade statistics. *Journal of Development Economics* **44,** 103–130 (1994).

28. Gehlhar, M. *Reconciling bilateral trade data for use in GTAP* tech. rep. (GTAP Technical Papers, 1996).

29. Makhoul, B. & Otterstrom, S. M. Exploring the accuracy of international trade statistics. *Applied Economics* **30,** 1603–1616 (1998).

30. Pohit, S. & Taneja, N. India's Informal Trade with Bangladesh: A Qualitative Assessment. *The World Economy* **26,** 1187–1214 (2003).

31. Beja, E. L. Estimating Trade Mis-invoicing from China: 2000–2005. *China & World Economy* **16,** 82–92 (2008).

32. Ferrantino, M. J. & Zhi, W. Accounting for discrepancies in bilateral trade: The case of China, Hong Kong, and the United States. *China Economic Review* **19,** 502–520 (2008).

33. Barbieri, K., Keshk, O. M. & Pollins, B. M. Trading data: Evaluating our assumptions and coding rules. *Conflict Management and Peace Science* **26,** 471–491 (2009).

34. Gaulier, G. & Zignago, S. *BACI: International Trade Database at the Product-level The 1994-2007 Version* CEPII Working Paper, No: 2010-23. 2010.

35. Dong, G. *Mirror statistics of international trade in manufacturing goods: The case of China* (United Nations Industrial Development Organization, 2010).

36. Hamanaka, S., Tafgar, A., *et al.* Usable Data for Economic Policymaking and Research? The Case of Lao PDR's Trade Statistics. *Asia-Pacific Research and Training Network on Trade (ARTNeT)* (2010).

37. Hamanaka, S. Whose trade statistics are correct? Multiple mirror comparison techniques: A test case of Cambodia. *Journal of Economic Policy Reform* **15,** 33–56 (2012).

38. Ferrantino, M. J., Liu, X. & Wang, Z. Evasion behaviors of exporters and importers: Evidence from the US–China trade data discrepancy. *Journal of International Economics* **86,** 141–157 (2012).

39. Pierce, J. R. & Schott, P. K. Concording U.S. Harmonized System Codes Over Time. *Journal of Official Statistics* **28,** 53–68 (2012).

40. Cebeci, T. A Concordance among Harmonized System 1996, 2002 and 2007 Classifications. *World Bank Working Papers, No. 74576* (2012).

41. Diodato, D. A Network-based Method to Harmonize Data Classifications. *Papers in Evolutionary Economic Geography*. No 18.43 (2018).

42. Bellert, N. & Fauceglia, D. A Practical Routine to Harmonize Product Classifications over Time. *International Economics* **160,** 84–89 (2019).

43. International Monetary Fund. *World Economic Outlook* https://www.imf.org/en/Publications/SPROLLs/world-economic-outlook-databases. 2025.

44. International Monetary Fund. Statistics Dept. *External debt statistics: Guide for compilers and users* (International Monetary Fund, 2014).