INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
IFPRI

POLICY NOTE APRIL 2023

Data Issues in Analyzing Agri-Food Trade in BIMSTEC

Challenges and Recommendations

Sunil Saroj, Devesh Roy, Abul Kamar, and Mamata Pradhan

Introduction

The focus of research on international trade has recently shifted from industries and countries to firms. Firm heterogeneity is shown to be a determinant of trade at both the intensive margin (increase exports per firm/product) and extensive margins (the number of firms exporting – new products, new partners, new varieties, and new prices). It is now widely accepted that exporting firms are larger, comparatively productive, more skilled, and capital-intensive, and pay higher wages than non-exporting firms.

The innovations in international trade literature that explains both the emergence as well as levels and the nature of trade flows through value chain integration necessitates examining trade-based exchanges at the highest possible levels of product disaggregation. Developments in trade theory emphasize that it is individual firms not countries that trade and analysis needs to incorporate firm characteristics in decisions and ability for exporting and importing. Firms are the appropriate unit of analysis for trade flows. It helps several paradoxes once the import of firm heterogeneity is understood.

Despite the substantive importance of granular level data and the significant level of disaggregated product-level bilateral trade flow data and enhanced computing power that are becoming available, most studies have tended to rely on analysis with high level of aggregation. Recent research on firm heterogeneity in international trade highlights the importance of extensive margins i.e., new products, new partners, new varieties, and cumulative of these i.e., new prices in trade patterns and firms' responses to trade liberalization and other policy changes. However, the high dimensionality of the data and the large number of responses to changes can easily overwhelm researchers. Additionally, bigger data sets may contain more noise, which can mask important systematic patterns. In analysis of trade flows, notwithstanding the rising incidence of differentiated products (varieties) and value chains that transcend national boundaries, methods in agri-food trade analysis in particular have not kept pace in terms of empirical methods and suitable data.

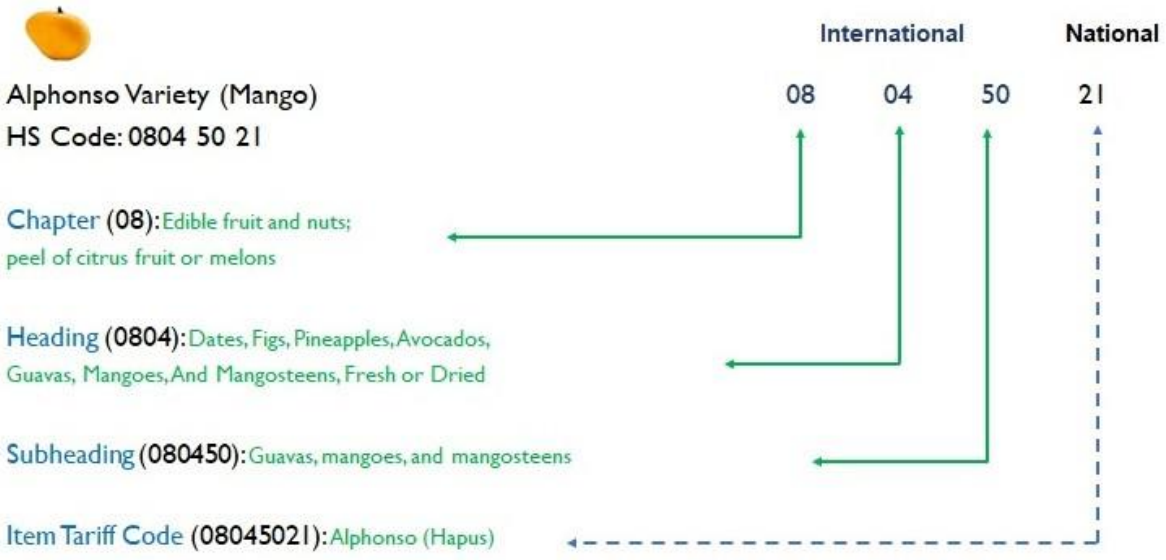
1. Trade flows data: Levels of (dis)aggregation

Harmonized System (HS) codes are commonly used in analysis of exports and imports of goods. The HS is a standardized numerical method of classifying traded products, and it is used by customs authorities around the world to identify products when assessing duties, taxes and for gathering statistics. The HS serves as the foundation for the import and export classification systems used by the trading partners and is administrated by the World Customs Organization (WCO). The system was first introduced in 1988 and is updated every five years. It is used by more than 200 countries and covers over 98% of goods in international trade. The HS comprises over 5,000 groups of goods which are grouped into 98 Chapters and 21 Sections. Each chapter is divided into headings and, where appropriate, sub-headings. Goods are generally arranged in order of degree of manufacture, with raw materials first, followed by unworked products, semi-finished products, and finished products. HS 2022 is the latest classification that has been approved by the WCO, and trade data is now available at this level.

In terms of the classification system, HS assigns specific six-digit codes (aggregate) to classify commodities into various categories. Countries can add longer codes to the first six digits for further classification at the national tariff lines level, resulting in codes with eight to ten digits (disaggregated). In the United States and Japan, a 10-digit code is used to classify products for export and import, with the first six digits being the HS number. Each product is identified by a six-digit code in the HS, arranged in a legal and logical manner according to specific rules. However, the China Customs Commodity HS Code has been changed from the original 10-digit HS code to the new 13-digit HS code in 2018. The 9^{th} and 10^{th} digits are customs supervisory additional numbers, and 11-13 are additional numbers for inspection and quarantine in China.

However, the six digits of an HS code can be broken down into three parts: a) the first two digits identify the chapter the goods fall in, for example, “08” (Edible fruit and nuts; peel of citrus fruit or melons); b) the next two digits identify the specific heading within that chapter, for example: “0804” (Dates, Figs, Pineapples, Avocados, Guavas, Mangoes, And Mangosteens, Fresh or Dried); and c) the last two digits denote a sub-heading that makes the classification even more specific (Figure 1), for example: “080450” (Guavas, mangoes, and mangosteens). In the absence of a subheading, the last two digits will be indicated by zeroes. However, in the national tariff line (disaggregated), we can further identify the variety of the products, for example, in Mango product at eight-digit level in India, “08045021” is Alphonso (mango variety), “08045022” is Banganapalli (mango variety), “08045023” is Chausa (mango variety), “08045024” is Dasheri (mango variety), “08045025” is Langda (mango variety), “08045026” is Kesar (mango variety), “08045027” is Totapuri (mango variety), “08045028” is Mallika (mango variety), “08045029” is other varieties of Mango, “08045030” is sliced/dried mango, and “08045040” is Mango Pulp. Hence, information at the disaggregated level is indispensable for the analysis of trade at the differentiated variety level and required for policies formation and analysis representing national tariff lines and as pathway for value chain integration including Globalized Value Chains (GVC). However, the importance of long period data at monthly level allows the analysis of emerging and urgent issues within a framework, captures seasonality, and trade responses to shocks and gives a wide geographical scope to understand the increase export per firm/product and new varieties, new products, new partners, and new prices.

Figure 1: Classification of HS Codes



Source: Central Board of indirect Taxes and Customs, Department of Revenue, Ministry of Finance, Government of India.

2. Challenges in Trade Data

The national tariff line data has another advantage in providing better information to monitor a product at a disaggregated level. This can help us understand which variety of the product is driving the trade growth and aid in understanding the new product and new partner margins. Focusing on such information and resources will improve the understanding of the dynamics in trade and the pathways for trade responses to policies. Yet, there are some interlinked challenges with tariff lines data that need highlighting from point of view of operationality, a case of exiguous disaggregation.

A. Currently, some important varieties within each chapter/ heading/subheading still get clubbed under “other” category, making it difficult to identify the trade of certain product varieties. Other options are meant to capture products that cannot be classified into any product groups. While traditionally the task of classifying a particular data point was put upon the exporting country (reporter), a current trend is to request information on a much more granular level and leave the classification to the authorities.

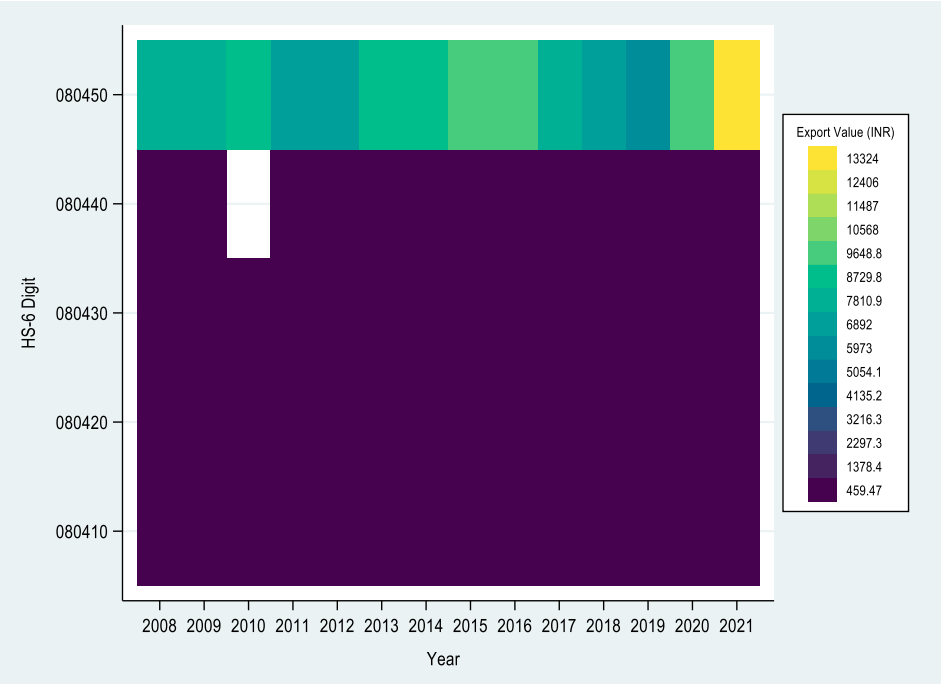
   However, the “Other products” category still forms a significant share of overall chapter/heading/sub-heading, resulting in a significant loss of information at the disaggregated level. Over time, HS codes keep updating and information on new varieties get revised after realization of the need for and importance of variety of specific product. In Figure 2, a product code “080450 - Guavas, Mangoes, and Mangosteens” is being traded in large volume, but it is not clear which variety is driving the trade of the product code “080450”. This can only be identified at the disaggregated level of product code “080450” where more than 9 varieties were introduced in HS2017, and only the sub-product/variety as driving trade can be identified. Extensive margins in effect at intra-firm level comprise new products/new varieties and new partners.

B. Using disaggregated data, policymakers/researchers can identify an extensive margin by looking partners that had never traded and/or a product/variety that had never previously been exported or imported. As the switch or shifts occur in varieties in trade, it captures the dynamics on the

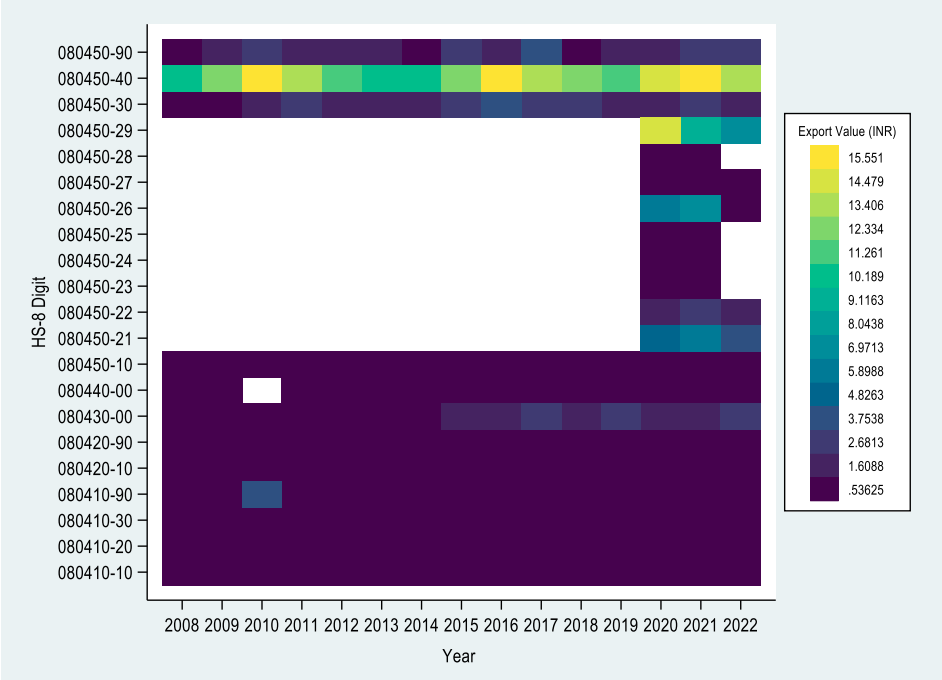
extensive margin and similarly the shift in partners. It is natural to expect these changes to be better captured with disaggregated trade data at comparatively high frequency. The disaggregated data also provides information on intra-industry trade and can also help understand processing trade as an element of GVC. It is going to be helpful to the extent that "other" category compresses and new HS codes are assigned except for structural changes in technology or products becoming obsolete from the market due to exogenous factors.

Figure 2: Example of aggregated and disaggregated trade data

Aggregated



Disaggregated



Source: Directorate General of Commercial Intelligence and Statistics (DGCIS), Kolkata, Ministry of Commerce and Industry, Government of India.

C. Some countries in BIMSTEC region lack data at the disaggregated level (8 digits), obviating rigorous analysis on the importance of extensive margins in the trade growth. This creates challenges for policymakers and decision-makers in advocating for the expansion of specific products and for trade expansion or adjustment strategies.

D. The Directorate General of Commercial Intelligence and Statistics (DGCIS) database in India provides port-wise data at a disaggregated level with the monthly frequency. It includes information on trade at the port level, covering different types of ports. such as Container Freight Station (CFS), Inland Container Depot (ICD), Special Economic Zone (SEZs), Export Processing Zone (EPZs), Air Cargo Complex (ACC), Free Trade Zone (FTZ), and Land Custom Stations (LCS) and modes of trading such as road, rail, air, and sea-ports. This detailed level of trade data offers insights into various aspects of trading, such as time-to-trade, cost-to-trade, effects of infrastructure bottlenecks, and product/trader centric port. However, the port-wise data lacks information on the origin of products and the firms involved in trade.

   Without this information, it is challenging to analyze the preferences and behavior of traders and undertake a firm-level analysis of trade. The costs of transporting commodities vary widely across countries, routes, directions, and commodities. Small traders and new start-ups often face higher transport and insurance costs for their exports and imports. Factors such as distance and geographical location, market size, scale effects, market structures, and infrastructure availability and quality can contribute to these cost variations. Obtaining firm level data is costly, and it is typically provided by private agencies. Therefore, it can be challenging to undertake firm-level analysis in the context of new trade analysis regimes.

E. In the face of these challenges, BIMSTEC member countries should provide 8-digit/national tariff line data for enabling effective, efficient, and analytically amenable systems and learn from China, Japan, and US. More granularities also imply very large datasets, creating computational opportunity to work on machine learning and big data techniques using the trade data. However, at later stage researcher and policymakers can showcase how new trade theory and advancement in data science and machine learning can provide a data-driven approach to designing and understanding trade patterns and policies in best possible manner.

F. In the case of the BIMSTEC region, there is an even more glaring absence of studies on export and import behavior. While BIMSTEC has the potential to become an influential exporter of raw and processed products and participate in value chain integration, due to its sheer size, there is a tendency to focus on exports and imports of agri-food products. If BIMSTEC region were to break free from states of under-trading i.e., trading below the potential, it requires understanding trade in terms of changing composition and organization. It would require the sheer focus on exports over imports to alter. In the changing landscape of type and nature of trade flows imports are important also for exports. However, it is important to also examine import behavior to gain a comprehensive understanding of the region's trade dynamics.

G. The two main hurdles to quantifying trade policies are conversion problems (such as tariffs, import quotas, border, and domestic policies) as how the gamut of policies across products and countries be expressed in a common unit and aggregation problems (tariff- equivalent measures, subsidy-

equivalent measures, quota-equivalent measures). The non-tariff barriers must be converted into tariff equivalents, and it becomes very difficult and subjective to stick to some random numbers.

H. Four out of the seven BIMSTEC member countries require using mirror data from partner countries for analysis which may not be frequently updated. Obtaining monthly-level data for Bangladesh, Bhutan, Myanmar, and Sri Lanka countries can be challenging, and researchers often must rely on the mirror data provided by ITC Trade Map or World Integrated Trade Solution (WITS). While this approach may be suitable for analyzing bilateral trade flows but becomes difficult to get the complete picture of the value chain process and multilateral trade.

I. Another challenge is related to the introduction of new HS codes every five years. It is good to have new HS codes every five years, but it gives another issue related to concordance. Roughly about 10 percent of the trade data break in time series continuity where HS is amended, revise (add, split and merge) of a particular commodity code. In addition to that, some countries report their data in old HS code, and it becomes very difficult to concordance or map the 10 percent of the products which can be a potential product to understand the new variety/products.

3. How IFPRI can help in the BIMSTEC region

IFPRI has taken an important step towards modernizing how it disseminates and visualizes its 8-digit data, by launching the BIMSTEC Agricultural Trade Monitor Plus (BATM+). This dashboard provides a wealth of information on trade statistics such as trade patterns, trade policies, and port-wise disaggregated data for BIMSTEC member countries. This BATM+ presents a high frequency trade patterns and policy monitor at high frequency. It is available in an easily accessible form and provides a coherent overview of trends and novel summary analyses of countries, regions, sectors, value chains, extensive and intensive margins. Additionally, the BATM+ identifies contributing factors, including low product diversification in exports and a high dependency on raw materials compared with other regions of the world. This tool will support all BIMSTEC member countries in developing the trade statistics needed to monitor and manage their performance in the global economy. Internationally comparable and timely statistics, along with cross-country analyses, can strengthen the foundation for designing and implementing national economic policies. The BATM+ covers a variety of trade policies including tariffs, quotas, antidumping duties, technical regulations which affect different aspects of the agri-food supply chain (primary producers, food processors, traders, consumers) and provide exact source of information.

4. Recommendations

A. Trade data is increasingly being shared globally at the HS-6-digit level. However, there is a need to move towards more open data at the HS-8-digit level, even moving to the firm level and for agri-food products there should be a proper a demarcation of products by food and fodder products such that identification of products can be easily determined at HS-8-digit level.

B. Port-wise data on a granular level allows us to pinpoint the source of goods and firms, as well as the location of the trade. This helps us better understand both the extensive and intensive margins of trade. Also, the whole matrix of taste costs including domestic cost becomes important in assessing trade.

C. The assignment of legal mandates to different agencies can make sharing data difficult, depending on a country's institutional framework. In BIMSTEC member countries, the availability of trade data is very sparse. National partners or international organizations can help build the capacity of the data governance team. Given the often-complex governance, it is no wonder that reporting requirements are a patchwork of overlapping and underlapping components. Additionally, sharing data among agencies is often difficult because existing laws prohibit sharing confidential data. Unlike DGCIS in India, other BIMSTEC member countries may provide data in a proprietary mode with detailed information.

D. The costs of transporting and time required to trade commodities vary widely across countries in BIMSTEC region, depending on the routes, directions, and commodities. Often, developing, landlocked developing countries, and least developed countries spend more than average on the international transport and insurance of their merchandize exports and imports. A wide range of factors cause variation in transport cost and time in trade, and broadly depend on distance and geographical location, the size of economies and markets, scale effects, market structures, connectivity to transport networks, infrastructure availability and product quality. Therefore, a trade dataset that includes information on time and cost of trade is an important practical need for more granular, global, and up-to-date information to understand the detailed dynamics.

E. In firm and port level data, the information related to firms can be anonymized and available at a disaggregated level with high frequency like DGCIS publishes 8-digit trade data at monthly frequency. DGCIS or any other authority from the BIMSTEC region can charge for the data and make it available for public use, so that policy makers and researchers do not have to rely on mirror data which has several drawbacks. LDCs should update the data regularly to stay competitive in the market and understand the dynamics of the trade patterns. Another important dimension is to introduce the gender information in the trade data. While collecting the firm level data, there should be an option to include the number of male and female workers engaged in the firms, as well as whether the firm is owned by a male or female.

ABOUT THE AUTHORS

Sunil Saroj is Senior Research Analyst**, Devesh Roy** is Senior Research Fellow, **Abul Kamar** is Senior Research Analyst, and **Mamata Pradhan** is Research Coordinator in the South Asia Regional Office of the International Food Policy Research Institute (IFPRI), New Delhi, India.

ACKNOWLEDGMENTS

Funding for this work was provided by the Bill & Melinda Gates Foundation (BMGF). This publication has been prepared as an output of Regional Trade and Agricultural Transformation in the Bay of Bengal countries and has not been independently peer reviewed. Any opinions expressed here belong to the author and are not necessarily representative of or endorsed by IFPRI.