Related Party Trade and Gravity: Revisiting the Distance Effects

Saera Oh *

*Link to the latest manuscript

28 October 2025

Abstract

This paper examines how internal firm distance costs shape the geography of foreign direct investment (FDI) and the export behavior of multinational affiliates. Using shipment-level export data from India, I document how the patterns and composition of related-party trade have evolved over the past decade. The evidence reveals that foreign affiliates engage in complex export behavior that is not fully captured in existing FDI frameworks and that this behavior has shifted markedly in recent years. To account for these patterns, I develop a partial-equilibrium model with heterogeneous distance elasticities across related-party and arm's-length exports and yields a firm-level gravity equation. The model is estimated using Poisson pseudo-maximum likelihood. The results show that within-firm exports are less sensitive to distance than arm's-length exports, but that this difference has diminished over time, consistent with the growing importance of inter-firm trade among multinational affiliates.

Keywords: Multinational Firms, Related Party Trade, Firm-level Gravity, Distance

*PhD Candidate, Michigan State University, E-mail: ohsaera@msu.edu. This work is supported by the Glenn and Sandy Johnson Fellowship (2024-2025) from the Department of Agricultural, Food, and Resource Economics, and by Dr. Oren Ziv and Dr. Justin Kirkpatrick from the Department of Economics.

1 Introduction

Foreign direct investment (FDI) plays a central role in the global economy, serving as a key channel through which multinational enterprises (MNEs) expand their operations across borders.[1] FDI and multinational activity shape labor markets, productivity, and innovation,[2] underscoring the importance of understanding how multinationals organize their activities across space. The geography of FDI is commonly analyzed through two lenses: horizontal FDI, which substitutes for exports by replicating production abroad, and vertical FDI, which fragments production across countries and complements trade. Empirical and theoretical work has suggested that intra-firm supply chains matter for the geography of multinational activity, and that multinational firms appear to face within-firm distance costs (Irarrazabal et al., 2013; Atalay et al., 2019). However, no estimates of within-firm overseas trade costs currently exist to substantiate these hypotheses. Furthermore, despite extensive documentation of MNEs' FDI decisions, the structure and nature of MNE trade within this global production geography remain opaque.

This paper studies the evolving role that internal firm distance costs play in determining the geography of FDI and the export behavior of foreign affiliates of MNEs. I use shipment-level export data from India to examine the patterns and evolution of related-party trade over the past decade.[3] The data show that foreign affiliates exhibit complex export behavior that is not well captured by the existing FDI literature and that this behavior has changed rapidly in recent years. I develop a partial-equilibrium model that matches these patterns and yields an estimable firm-level gravity equation allowing for heterogeneous distance elasticities between related-party and arm's-length exports. The estimation results indicate that related-party exports are less elastic to distance than arm's-length exports, but that this difference has eroded over time, helping to explain a recent surge in inter-firm trade among the foreign affiliates of multinational firms.

This study presents a new method for identifying within-firm trade, granting the ability to better observe affiliate-to-affiliate transactions. Most firm-level datasets widely used in the literature on FDI and MNEs limited in coverage.[4] Many of these confidential, administrative datasets identify related-party trade using ownership information, includ-

[1] According to the IMF, the global stock of inward FDI exceeded USD 41 trillion in 2023—roughly two-fifths of world GDP. MNEs account for about one-third of global output (Delpeuch et al., 2025).

[2] FDI and multinational activity influence employment (Lipsey and Sjöholm, 2004), wages (Setzler and Tintelnot, 2021; Alfaro-Urena et al., 2021), worker mobility (Balsvik, 2011), productivity spillovers (Smarzynska Javorcik, 2004; Alfaro and Chen, 2018), product quality and price premiums (Ge et al., 2015), and drive innovation and knowledge diffusion (Keller, 2010). See also Moran et al. (2005); Moran (2011).

[3] Throughout the paper, I use the terms related-party and within-firm interchangeably, as well as arm's-length, between-firm, across-firm. The term intrafirm specifically refers to trade between a parent firm and its affiliate (or vice versa), unless otherwise indicated by the context.

[4] The majority of existing studies rely on data from the United States, while others use datasets from selected European countries (e.g., Belgium (Conconi et al., 2025), France(Davies et al., 2018)), Latin America (e.g., Costa Rica (Alfaro-Ureña et al., 2022)), and Asia (e.g., China (Wang, 2021)).

ing the presence and share of foreign ownership. In this study, I use Bill of Lading data which is publicly accessible to match shipping and receiving firms and classify transactions as either within-firm or across-firm. Using information on each exporting firm's ultimate owner and the number of countries in which that owner operates, I first determine whether the ultimate owner qualifies as an MNE and then identify the corresponding foreign affiliates by matching exporter and importer names. Each shipment is classified as either an arm's-length or a related-party transaction, including both affiliate-to-parent and affiliate-to-affiliate shipments. Crucially, my matching procedure to identify sister establishments both in home and foreign countries does not rely on ownership relations and is therefore better suited to documenting affiliate-to-affiliate transactions.[5]

Using the data to take a detailed look at the shipment-level trade activity of foreign MNE affiliates, I document four stylized facts about their export behavior. First, as documented in Ramondo et al. (2016), I confirm foreign MNEs sell through both channels—related-party (within) and arm's-length (across)—to foreign markets via their affiliates in India. Second, I further document that they both sell intermediate inputs but typically export distinct sets of products across the two modes. These facts imply that MNE foreign affiliates are engaged in both within-firm and across-firm supply chains. Third, the probability of engaging in related-party exports is positively correlated with distance. Fourth, related-party exports were disproportionately directed toward distant markets through 2020, but this pattern has weakened in the past few years. Taken together, these facts suggest that the distance elasticity of exports may differ between internal trade within MNE networks and arm's-length trade, and further that this gap may have narrowed in recent years.

Guided by these facts, I develop a partial-equilibrium framework grounded in the Ricardian model of Eaton and Kortum (2002). A large set of suppliers produce and sell a continuum of goods and can serve foreign markets either through related-party or arm's-length exports, with trade costs that differ across these two modes. A representative buyer sources inputs to minimize production costs. The model's comparative statics are consistent with the observed patterns: the higher share of related-party exports to distant destinations—and the subsequent attenuation of this pattern—can be interpreted as changes in the underlying distance elasticity. The equilibrium conditions yield estimable firm-level trade equations that capture both within- and across-firm distance elasticities.

I show that the difference in distance elasticity can be estimated using a Poisson pseudo maximum likelihood estimator. I implement the estimation using Bill of Lading data.[6] The results indicate that related-party exports are less sensitive to distance than

[5] Alfaro et al. (2025) point out that the Longitudinal Firm Trade Transaction Database (LFTTD), the most widely used source on MNE activity in the United States, flags related-party trade based on ownership shares. Further details are provided in Footnote 7

[6] To align the empirical implementation with the theoretical model, I combine India's Bill of Lading data with Mexico's, which allows for the inclusion of destination-country fixed effects.

arm's-length exports, but that this difference has indeed diminished over time. Within the context of the model, this erosion of distance elasticities helps explain the recent expansion of cross-firm trade among MNEs: as the relative advantage of within-firm trade at a distance declines, MNE affiliates increasingly engage in arm's-length transactions abroad.

The contributions of this study are threefold. First, I contribute to the empirical literature on the behavior of multinational firms by presenting a new method to identify related-party exports by foreign MNE affiliates in readily-accessible data, including and especially for affiliate-to-affiliate flows. Second, I provide new descriptive and empirical evidence on their export behavior in India, shedding light on nature of the within- vs across-firm export behavior of MNE foreign affiliates and in particular the heterogeneous response of this behavior to distance. Third, and most importantly, I contribute to the literature on the geography of FDI by estimating heterogeneous distance elasticities separately for within- and across-firm trade, and by documenting rapid recent changes in these differences that correlate with shifts in the trade behavior of foreign MNE affiliates.

First, this study contributes to the literature on identifying within-firm trade among MNEs. A key challenge in the empirical analysis of multinational firm activity has been the difficulty of observing both within- and across-party trade. Earlier seminal work concluded that within-firm trade is relatively limited or sparse among U.S. MNEs (Ramondo et al., 2016). However, Alfaro et al. (2025) raises concerns about potential undercounting in U.S. MNE data, proposes a methodological remedy, and finds that within-firm trade between parents and affiliates is, in fact, prevalent. Alfaro et al. (2025) also notes that other studies—particularly those using U.S. Customs data—document parent–affiliate trade, but a persistently unaccounted-for aspect of within-firm trade is affiliate-to-affiliate transactions.[7] Although recent papers, such as Antràs et al. (2024), combine multiple data sources, important limitations remain. In this study, I use publicly accessible transaction-level data and match exporting and importing firm names to identify related-party trade. By doing so, I am able to observe affiliate-to-affiliate transactions and incorporate them into the broader measure of related-party trade. A related study, Hong (2023), also begins to address this gap by developing a new proxy and methodology to identify related-party trade more broadly, using web-scraped financial statements of firms in South Korea.

[7] The most widely used source of data on MNE activity in the United States is the Longitudinal Firm Trade Transaction Database (LFTTD), which is based on U.S. Customs data. In this dataset, for U.S. exports, a related-party transaction is defined as "a transaction involving trade between a U.S. principal party in interest and an ultimate consignee where either party owns directly or indirectly 10 percent or more of the other party." Another dataset, the Related Party Trade data published by the U.S. Census Bureau, uses the same definition (Ruhl, 2015). This definition narrowly captures related-party trade and essentially implies a parent–affiliate relationship. An exception to this may be studies using Bureau of Economic Analysis (BEA) survey data. The BEA provides information on affiliates' sales by transactor and destination; however, the BEA data consolidate information across all foreign affiliates, and shipment-level information on affiliate-to-affiliate trade is not available. As a result, sector- or product-level details are lost. Moreover, as discussed by Alfaro et al. (2025), the BEA data may undercount affiliate-to-affiliate transactions.

Second, this study connects to the literature documenting the trade behavior of foreign affiliates of MNEs. Cadestin et al. (2018) document that foreign affiliates are often highly export-oriented, yet affiliate trade beyond transactions with the parent firm remains underexplored. A growing body of work has examined affiliate-level export patterns using firm-level data. For example, Garetto et al. (2024) show that foreign affiliates of U.S. MNEs typically begin with host-market sales before expanding into exports, and that their market entry patterns are largely unaffected by sibling affiliates in other countries.[8] Similarly, Antràs et al. (2024) and Wang (2021) highlight the importance of geography, showing that affiliates tend to trade more with countries geographically closer to their foreign affiliate's locations or headquarters. Building on this line of research, my study focuses on the transaction channel— within or across firm—rather than on the timing or geographic scope of exports. This study is closely related to analyses of how trade is organized within multinational firms. Ramondo et al. (2016) show that foreign affiliates of U.S. MNEs sell largely outside the firm, a result also documented in Antràs et al. (2024). Similarly, Davies et al. (2018) find limited within-firm trade among MNEs in France. I confirm these results using data on the universe of MNE affiliate transactions in India and capturing all affiliate-to-affiliate exports including those with no direct ownership relations. I further extend these results by leveraging the product-level detail in my data, showing that affiliates export distinct sets of products within and across firms. Moreover, I show how over time, selling outside the firm has become increasingly dominant.

Finally, I contribute to a growing literature exploring the role of distance in multinational production. Irarrazabal et al. (2013); Alviarez and Saad (2023); Bombarda and Marcassa (2020), highlighting the distance effect on MNE's sales in the host country due to the existence of intra-firm trade in input. Alfaro and Charlton (2009) multinationals tend to own the stages of production proximate to their final production giving rise to a class of high-skill intra-industry vertical FDI Garg et al. (2023) find a strong preference for sourcing from within-firm suppliers in India over distance due to the investment specificity. Finally, Atalay et al. (2019) show that firms are more likely to conduct internal shipments at any given distance, suggesting the presence of a distance premium in the United States. However, they focus on domestic trade between MNE establishments. Together, this literature strongly suggests that distance plays a large role in the geography of trade within the firm, and that distance may have differential effects within and across firms. I contribute to this literature in two ways. First, I confirm that the within- vs across-firm sourcing behavior of MNEs are impacted by distance from my data and show that these patterns have changed dramatically in recent years. Second, I provide direct estimates of within- and across-firm distance elasticities using Bill of Lading data from India and Mexico, key locations for FDI, thereby confirming indirect evidence in the literature on differential distance costs.

[8] They use a long panel of U.S. MNE data from the BEA spanning 1987–2011, focusing on affiliate activity in the top ten U.S. FDI destinations.

The rest of the paper is structured as follows. Section 2 describes the data and explains how I identify foreign MNE affiliates and related-party trade. Section 3 presents the stylized facts on the export behavior of foreign MNE affiliates. Section 4 outlines the theoretical framework and derives the estimable firm-level trade equation. Section 5 details the empirical implementation and discusses the results. Section 6 concludes.

2 Data

In this section, I introduce the Bill of Lading data, outline the data processing procedures, and describe the additional datasets used in this study.

2.1 Bill of Lading and S&P Firm Data

Bills of Lading (BoL) are essential legal instruments in global shipping and logistics that provide detailed information on physical trade transactions between exporting and importing firms. BoL data have been widely used to study trade shock and recovery (Flaaen et al., 2023), trade routes and indirect shipping costs (Ganapati et al., 2024), transshipment (Do et al., 2025), and the links between financial globalization and international trade (Bruno and Shin, 2023) among others.

This study uses BoL data for India obtained from the S&P Global Panjiva Supply Chain Intelligence database (hereafter S&P Panjiva), complemented with data on firms' corporate structures.[9] The BoL records provide detailed information on exporting and importing firms—including firm names and unique identifiers—traded products (identified by Harmonized System codes), shipment destinations, and the value of goods, covering the period from 2016 to 2024.

Unlike in many other countries, India's BoL records are reported at the item level rather than solely at the shipment level.[10] This item-level structure enables precise product identification and allows the construction of firm–product pairs as the unit of analysis. To identify destination-country fixed effects in the estimation, I additionally incorporate item-level BoL data from Mexico, as described in Section 5.

In the following sections, I describe the data processing procedures and provide a summary of the dataset.

[9]Sources: S&P Capital IQ Base and S&P Ultimate Parent.

[10]A single shipment may contain multiple items or product lines; thus, using only shipment-level data can obscure product-level variation.

2.1.1 Identifying MNEs and Foreign MNE Affiliates

First, I link firms in the BoL data to S&P's firm-level data on firm's corporate structure. S&P's firm-level data on firm's corporate structure provide information on Ultimate Parent of firms in BoL data.[11] I count the number of countries in which each Ultimate Parent has affiliated entities. An Ultimate Parent is classified as MNE if it has affiliates in more than one country. Based on the Ultimate Parent firm's country of origin, I then identify foreign MNE affiliates and home MNE affiliates in the BoL data. An Ultimate Parent with affiliates located in only one country is classified as non-MNE, meaning it is itself a domestic firm that appears in the BoL data. The overall matching and classification process is illustrated in Figure A1.

Figure 1 presents the aggregated export by multinational status of firm that I identified above. As shown in the figure, there are firms for which multinational status (of its own or parent) cannot be identified. The matched firms account for approximately 50–80% of total trade flows each year. The total aggregated export value from S&P Panjiva is considered representative, as it closely aligns with the aggregate trade statistics reported in UN Comtrade indicated in red dotted line. [12]

In terms of export shares, the participation of MNEs has expanded in line with the overall growth of India's exports. The sharp increase between 2020 and 2021 likely reflects the rebound in export activity following the COVID-19 pandemic. Sector-level details are provided in Appendix Figure A2 and Figure A3. Summarizing the sectoral patterns, the role of MNEs has become particularly important in the Mineral Products and Base Metals sectors. Over the past three years, India's exports of Mineral Products have ranged between USD 60 and 80 billion, with MNEs accounting for more than 80 percent of total trade. MNEs also continue to play a substantial role in the Chemical Products, Plastics and Rubber, Machinery and Mechanical Appliances, and Vehicles and Aircraft sectors.

2.1.2 Identifying Related Party and Arm's Length Export

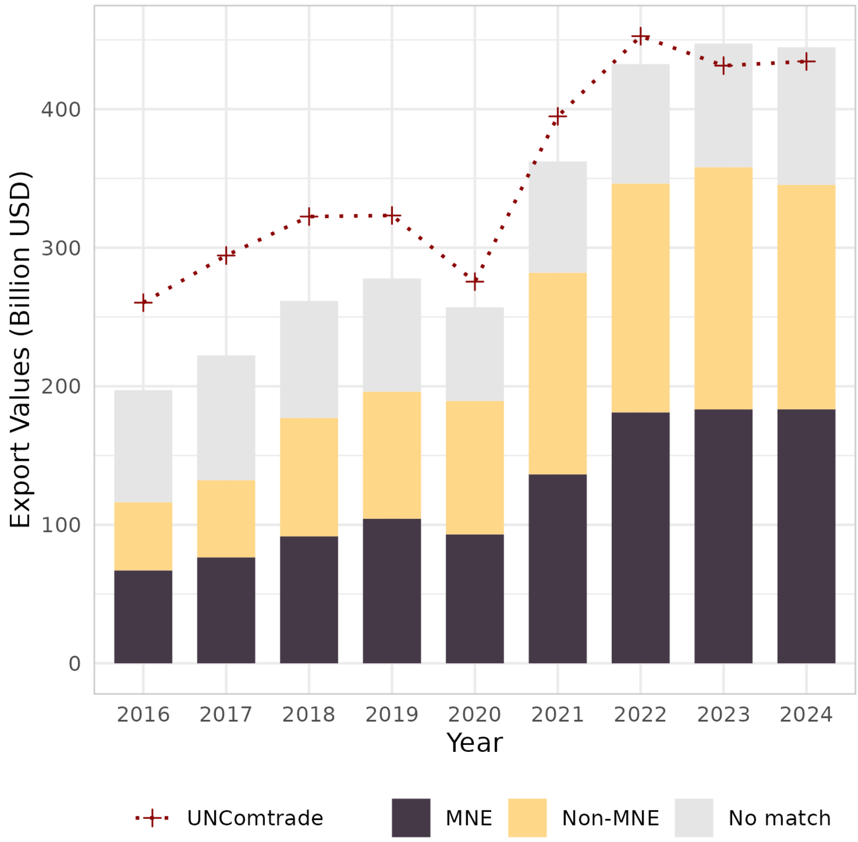
To identify the related party and arm's length trade, I employ the two type of string match algorithm that R package provide. [13] Detailed matching procedure is provided in

[11] According to S&P's technical documentation, the Ultimate Parent is defined as "the company at the top of a corporate structure or the legal organization that is ultimately responsible for all associated entities below it" (S&P Global, 2024).

[12] Nonetheless, some discrepancies remain. These differences are likely attributable to the exclusion of observations with missing destination information during the aggregation process. Although misreporting may also contribute to these discrepancies, investigating such issues lies beyond the scope of this study.

[13] One is the Jaccard distance which measures how different two sets are by comparing the number of shared elements to the total number of unique elements between them. In the case of strings, it typically compares sets of overlapping pieces (called q-grams) taken from each string. The larger number means more similar strings. Another is the Jaro-Winkler. It is a way to measure how similar two strings are, especially useful for short texts like names. It looks at how many characters match and whether they appear in the same order.

Figure 1: Export from India by Firm Type



Notes: The y-axis reports export values in billions of U.S. dollars. Multinational enterprise (MNE) status is defined according to the classification of the parent firm. 'No match' refers to firms that are either not identified in the S&P firm database or lack parent information. The dark red dashed line represents UN Comtrade export values, which are reported on a free on board (FOB) basis. India's Bill of Lading export values are also recorded on an FOB basis. MNE category includes home MNE establishment and foreign MNEs affiliates.

Appendix A.3. During the matching process, I aim to match firms that either have exactly the same name or share key elements of their names (e.g., Nestlé India and Nestlé U.S.).[14]

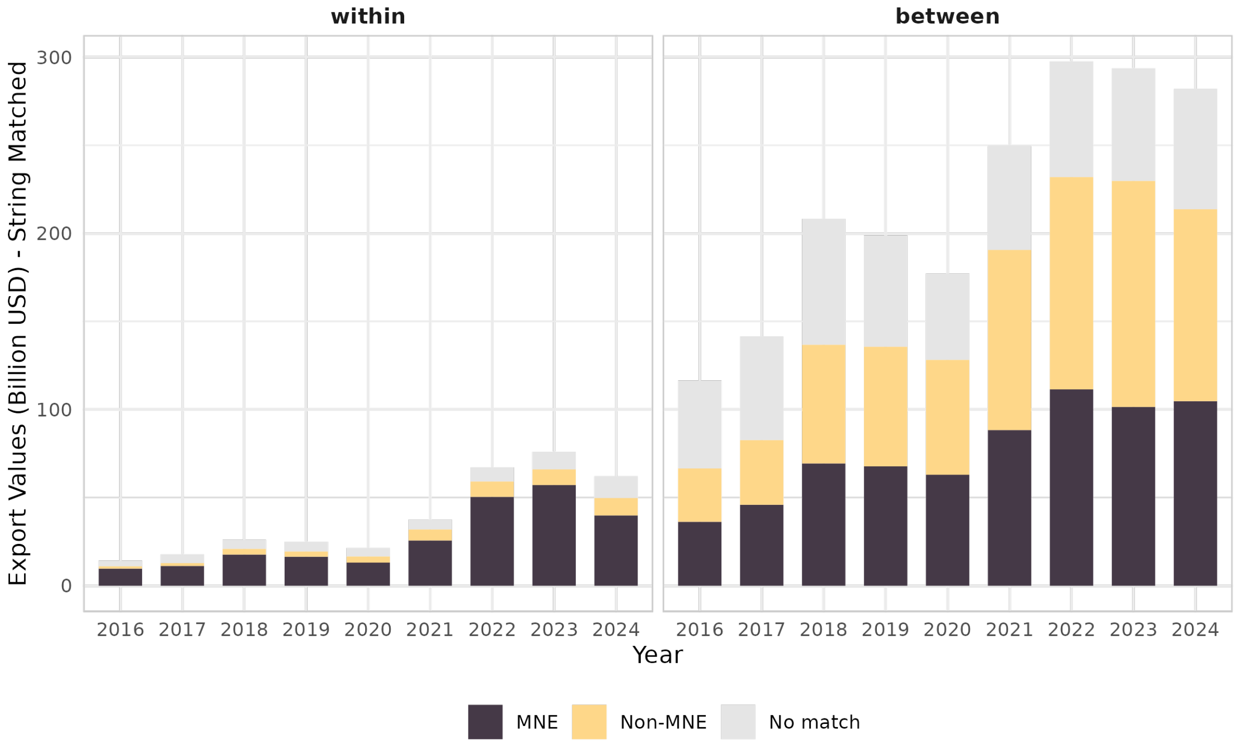
Technically, if the string-matching process performs well, MNEs should exhibit both related-party (within-firm) and arm's-length (between-firm) trade. In contrast, non-MNEs should display only between-firm export. In this study, non-MNEs are defined as firms whose ultimate parent operates solely within a single country. Consequently, it is unlikely that such firms have a parent or subsidiary abroad. Although unmatched cases cannot be definitively classified due to the absence of identifiable firm structure, I assume they resemble non-MNEs. Accordingly, I expect this group to display a higher incidence of between-firm transactions than within-firm transactions.

If the beginning of the two strings is the same, it gives a small bonus to the score. The result is a number between 0 and 1, where 1 means the strings are exactly the same. The Jaro-Winkler distance is just 1 minus this score, so smaller numbers mean more similar strings. The threshold used to determine related-party trade has been manually tested through trial and error, and it can be further refined.

[14] It is possible that related firms do not share any common name but in the current version of this study and it might underestimate the related party flows. Further revision will address this concerns.

The matching results in Figure 2 show that the majority of within-firm flows are concentrated among MNEs category, which is the expected and desirable outcome. However, I also observe some within-firm transactions among non-MNEs and unmatched firms. This likely reflects imperfections in identifying firm relationships based on names and firm structure. Nonetheless, given the clear patterns observed among MNEs, I remain reasonably confident in the overall performance of the matching procedure.

Figure 2: Related Party and Arm's length Export Matching

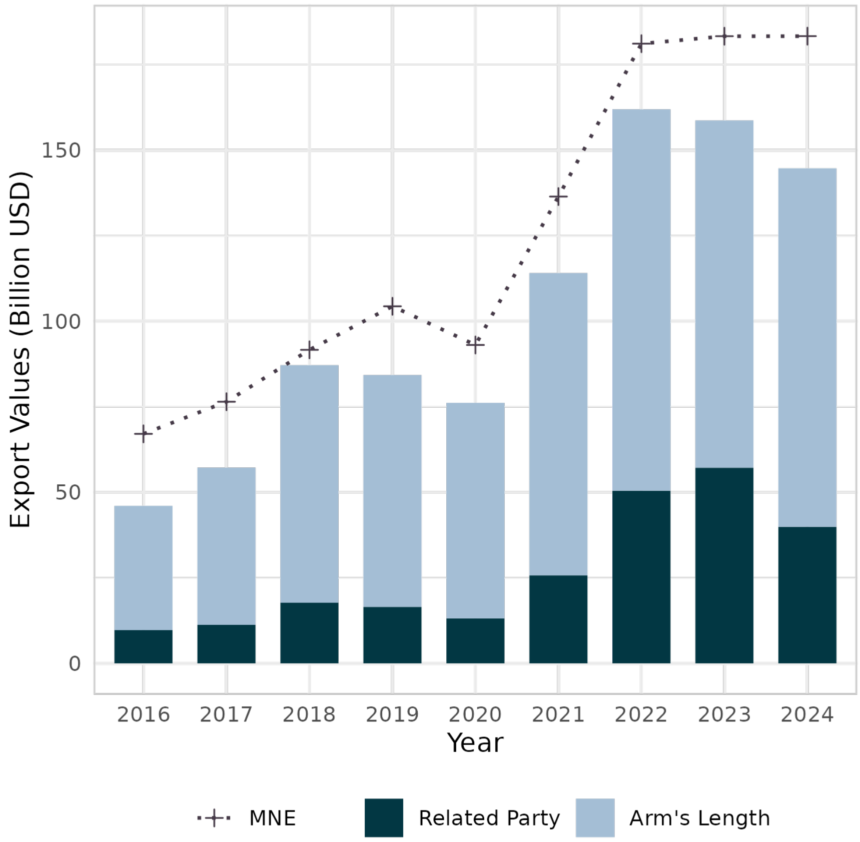


Notes: As explained in the main text MNE status is determined by the parent firm's classification. "No match" refers to firms that are either not found in the S&P firm database or lack parent firm information. Related party trade is identified through string matching between exporter and importer names. Transactions with missing exporter or importer firm names are also removed. As a result, the total export value reported in this figure (sum of within and between) is lower than that in Figure 1. In the labels, "within" denotes related-party export and "between" denotes arm's-length export.

2.1.3 Export by MNE Establishments

Finally, I describe related-party and arm's-length exports of MNE establishments in India in Figure 3. The related-party category includes affiliate-to-parent, parent-to-affiliate, and affiliate-to-affiliate shipments. The matching results for India are robust, covering approximately 80–90% of total exports by MNE establishments. While arm's-length exports continue to account for the larger share of MNEs' total exports, related-party exports have expanded rapidly, roughly doubling since 2022. Further I find that this increase was driven primarily by the sharp rise in home MNEs' related-party exports (parent-to-affiliate, intra-firm trade) (Figure A6).

Figure 3: Related Party and Arm's Length Export from MNE Establishment in India



Notes: Y-axis is in billions of dollars. Multinational status is determined by the parent firm's classification. Related party trade is identified through string matching between the exporter and importer names. Dashed line indicate the total export of MNEs from India illustrated in Figure 1.

2.2 Other Data

Distance and Other Gravity Controls Distance is obtained from the Centre d'Études Prospectives et d'Informations Internationales (CEPII) Gravity Database, where it is measured as the great-circle distance between the largest cities of each country pair. The database also provides additional variables that capture factors influencing bilateral trade costs, such as contiguity, common official language, colonial ties, free trade agreements, and common legal origins and unilateral information on if country currently is a WTO member or not.

GDP and Population GDP and population data are sourced from the World Bank's World Development Indicators (WDI). When data for the most recent year are unavailable, values from the preceding year are used instead.

2.3 Summary

The BoL data used in this study consist of export transactions from India. These data include all firms exporting from India, regardless of whether they are domestic or foreign-owned. Each export record provides identifier and names on (1) the exporting firm that actually conducts the shipment and (2) the importing firm that receives it. Using supplementary datasets and name-matching procedures, I further identify (1) the parent company of each exporting firm, (2) the multinational status of the parent firm, and (3) whether the exporting and importing firms are related parties.

I take careful steps to minimize potential measurement error in the data processing. First, I exclude Bill of Lading records that lack importer names or destination information, which may lead to a conservative estimate of total trade flows. Second, when matching exporter and importer names, I link firms with identical or closely aligned names, while recognizing that some related entities may not share common keywords in their names. Ongoing work will further refine these matching procedures to enhance data accuracy.

The rest of paper focus on MNEs' export.

3 Facts

This section presents several stylized facts on multinational firms' exports conducted through related-party and arm's-length transactions, with particular attention to how these patterns vary with distance. Unless otherwise specified, the analysis includes both foreign MNE affiliates operating in India and home MNEs headquartered in India. First, I look at the export patterns of foreign MNEs in Section 3.1, and then I explore the relationship between distance and related-party exports in Section 3.2.

3.1 Characteristics of Export Pattern

In our dataset, approximately 2,000 multinational enterprises (parent firms) were operating in India in 2024. The total number of MNEs has steadily increased since 2016 (about 1,700),[15] and this increase is attributed to the rise in foreign MNEs. The number of foreign MNEs in India was 1,028 in 2016 and increased to 1,190 in 2024. This trend reflects India's growing integration into global production networks and the sustained inflow of foreign direct investment. According to the World Bank, India's net FDI inflows remained consistently positive over the past decade, averaging around 2–3 percent of GDP between

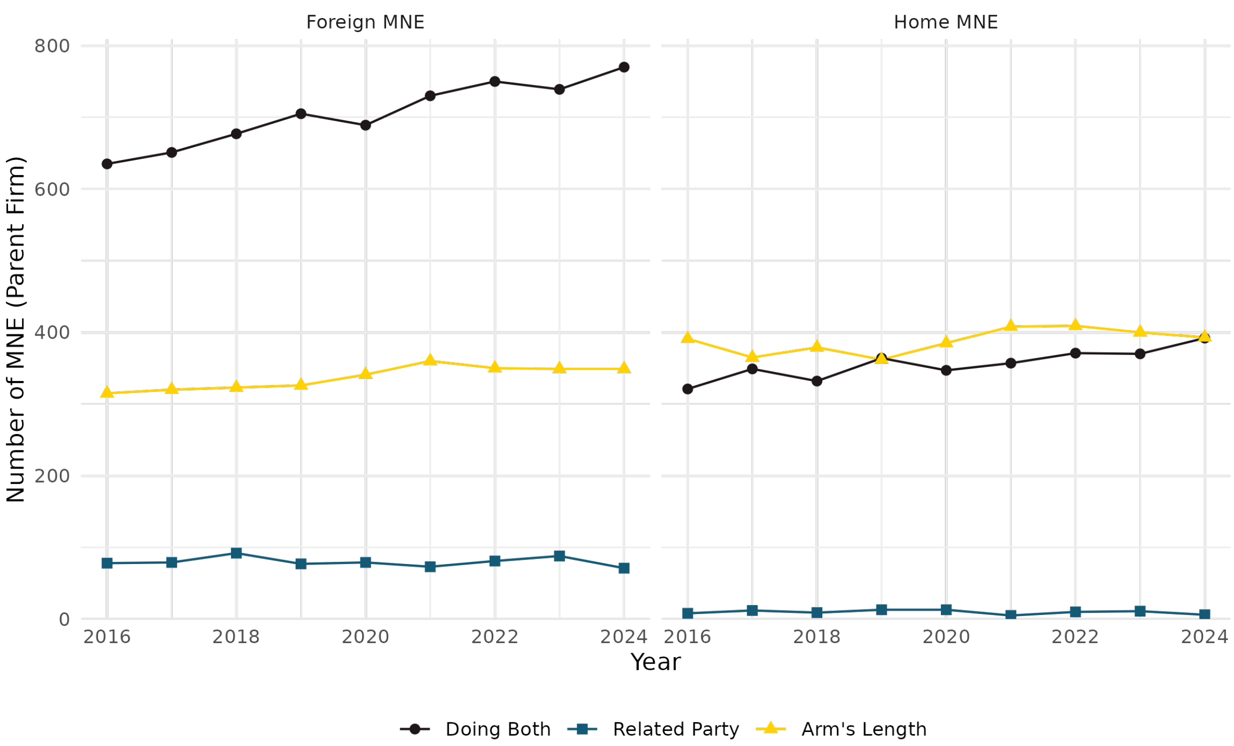
[15]Note that this figure refers to parent firms, not affiliates.

2016 and 2019, surging to about 5.4 percent in 2020, and stabilizing thereafter at around 0.7 percent of GDP (roughly USD 27 billion) in 2024.[16]

MNE Trading Partner Composition First, I examine the nature of MNE exports. In particular, I examine whether these MNEs engage primarily in related-party or arm's-length export. Figure 4 presents the distribution of export modes by MNE type (foreign versus home). About 61–64% of foreign MNEs engage in both related-party and arm's-length exports, indicating that most firms simultaneously serve their own corporate networks and independent buyers. A subset of firms specialize exclusively in one mode, and among these, arm's-length exporters outnumber those trading solely within related parties. This pattern suggests that even within multinational networks, firms actively pursue external market opportunities beyond their corporate boundaries—a feature that holds for both foreign and home MNEs, although the share of home MNEs engaging in both types of export modes is lower than that of foreign MNEs (44–59%). I summarize these findings below in Fact 1:

Fact 1. *MNEs affiliate establishments engaged in both related party and arm's length export.*

Figure 4: To Whom MNEs Sell



Notes: The y-axis reports the number of parent firms, not affiliate firms, which are the actual exporting entities.

[16]World Bank, Foreign Direct Investment, Net Inflows (BoP, Current USD), accessed October 2025.

Export Composition Next, I explore the composition of goods foreign MNEs exports via two channel. In Figure 5, I plot the export values of foreign MNE affiliates by end-use classification defined by the United Nations System of National Accounts (SNA). Exports by firms that engage in both related-party and arm's-length transactions exhibit a similar composition, with both modes dominated by intermediate and capital goods rather than final goods. However, exports by home MNE affiliates show a markedly different pattern (see Figure A4). Among home MNEs that export to both related parties and independent buyers, the majority of intermediate goods are sold to independent buyers, while exports to related parties are concentrated in fuels (motor spirits) and passenger motor vehicles. This pattern has been apparent in the data since 2022.

Next, I compare the export structures of the two transaction modes by calculating an export similarity index that quantifies the overlap in export commodity baskets between related-party and arm's-length trade. This allows us to assess the extent to which the goods traded under the two modes are differentiated. The index, originally proposed by Finger and Kreinin (1979) for cross-country comparisons of export composition, is adapted here to the parent-firm level to compare a firm's export structures across transaction modes. Conceptually, the index captures the shared portion of export shares across products—the area of overlap between two normalized export distributions. I select firms that engage in both related-party and arm's-length exports. For each firm f and destination j, it measures how similar the export composition is, in terms of product mix, between arm's-length and related-party exports. Formally, the measure is defined as follows:

$$\mathrm{ESI}_{fj}(\mathrm{Arm}, \mathrm{Related}) = \sum_k \min\left(\frac{\mathrm{Arm}^k_{fj}}{\sum_k \mathrm{Arm}^k_{fj}}, \frac{\mathrm{Related}^k_{fj}}{\sum_k \mathrm{Related}^k_{fj}} \right) \tag{3.1}$$

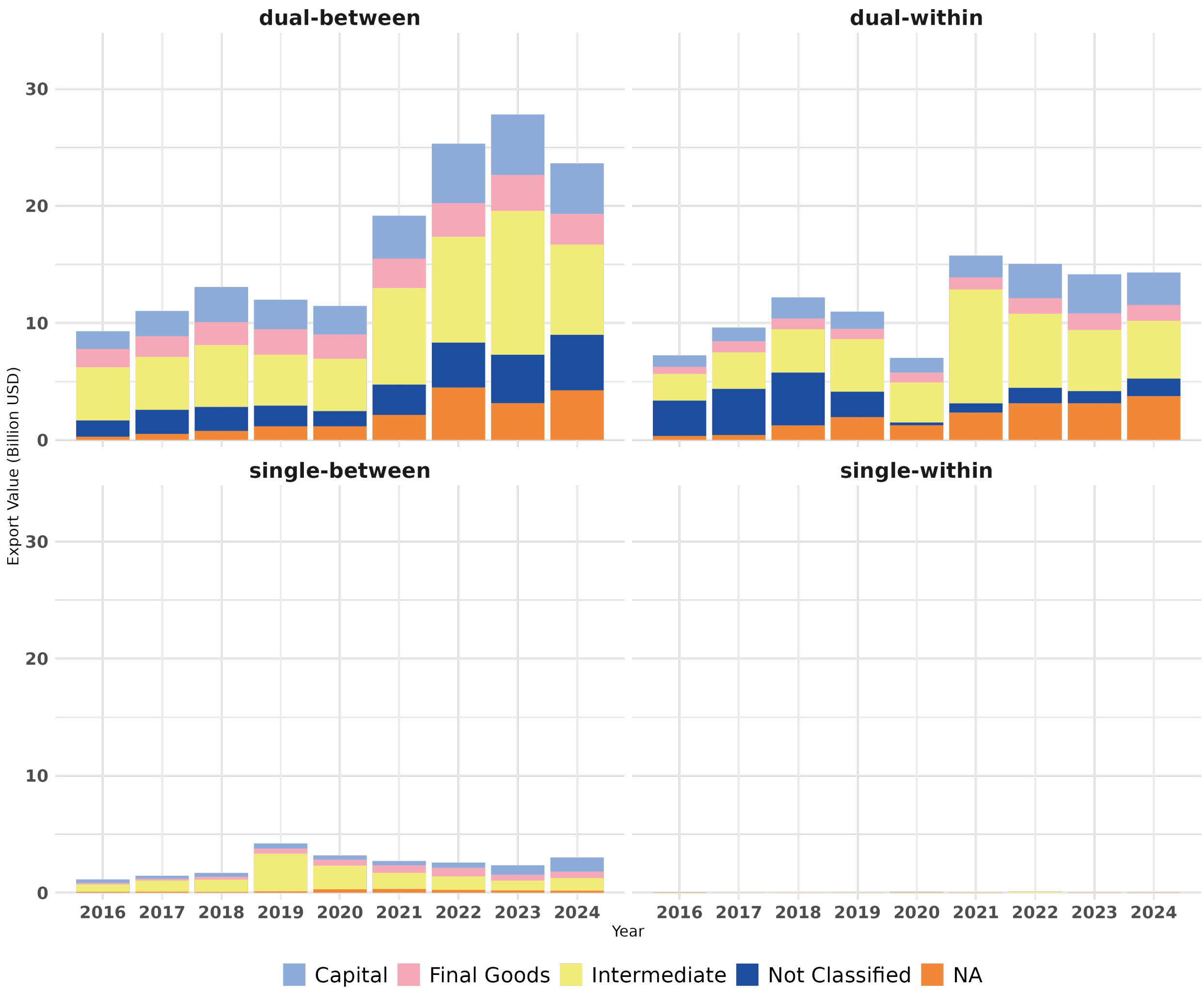
where k denotes products at the HS 6-digit level. Based on this index, I compute the average percentage of firms with a value of zero for a given destination. On average, between 94% and 96% of firms exporting to a destination exhibit no overlap in their exported products across the two transaction modes (see Table 1). This finding indicates that most firms export distinct sets of goods under related-party and arm's-length trade to the same destination.

Table 1: Average Percentage (%) of Firms with Zero Export Similarity Between Arm's-Length and Related-Party Exports, by Foreign and Home MNE

	2016	2017	2018	2019	2020	2021	2022	2023	2024
Foreign MNEs	95.47	94.23	94.60	94.90	95.66	95.18	94.54	95.23	95.54
Home MNEs	97.33	97.24	97.45	97.08	97.06	97.16	94.14	93.15	93.69

Note: Author's calculation. Foreign MNEs refer to affiliates of multinational firms headquartered outside India, while Home MNEs refer to Indian multinationals with foreign affiliates

Figure 5: Export Volume (Value) By Foreign MNE affiliates



Notes: The y-axis is measured in billions of U.S. dollars. Classification is based on the concordance between HS codes and the Broad Economic Categories (BEC Rev. 4), which are then aggregated into three basic end-use categories defined by the United Nations System of National Accounts (SNA): capital goods, intermediate goods, and final (consumption) goods. Capital goods include machinery, equipment, dwellings, and other durable assets that are less disposable. "Not Classified" refers to items such as motor spirits (gasoline) and passenger motor vehicles, while "NA" indicates unmatched observations. In the labels, "within" denotes related-party export and "between" denotes arm's-length export. The category labeled "dual" indicates firms engaged in both types of exports (as shown in left panel (Foregin MNEs) in Figure 4), whereas "single" indicates firms operating in only one mode either related-party or arm's-length exports. The category labeled "single-within" contains non-zero values that are not visible in the figure due to the axis scale.

I summarize these findings below in Fact 2:

Fact 2. *Foreign MNE affiliates mainly sell intermediate goods to both related parties and independent buyers abroad. They typically export distinct sets of products through related-party and arm's-length channels.*

3.2 Related Party Export and Distance

In the following subsection, I examine the relationship between foreign MNE affiliates related-party exporting and geographic distance. I first analyze the binary decision to engage in related-party exports as a function of distance, and then explore how the share of related-party exports varies with distance over time.

The Impact of Distance on Probability of Related-Party Export I estimate a linear probability model (LPM) to statistically examine the relationship between distance and the decision to engage in related-party exports[17]. The dependent variable, Y_{fhijt}, equals 1 if parent firm f, headquartered in h, conducts any *related-party* export from India (i) to destination j in year t, and 0 otherwise. Although the dataset contains all affiliate-level export records, they are aggregated to the parent-firm level for this analysis. Therefore, the unit of analysis is the parent firm rather than the affiliate. Recall that all affiliates are located in India, while their parent firms headquartered other than India. Our main regressor is the log of the distance between India and destination j, $\log(dist_{ij})$. Following the "headquarters-gravity" literature (e.g., Head and Mayer, 2019; Wang, 2021), I also include the log of the distance between the parent firm's headquarters h and destination j, $\log(dist_{hj})$. I control for lagged firm-level characteristics measured in India (e.g., the number of destination markets and products exported, and the number of Indian affiliates) to proxy for the scale of production and operations. The model is estimated using ordinary least squares (OLS), and heteroskedasticity-robust standard errors are clustered at the parent-firm level (the unit of regression).

$$
\begin{aligned}
Pr(Y_{fhijt} = 1 | X_{fhijt}) = & \beta_1 \log(dist_{ij}) + \beta_2 \log(dist_{hj}) + \beta_3 fta_{ijt} \\
& + \beta_4 \log(gdp_{jt}) + \beta_5 \log(pop_{jt}) \\
& + \gamma_F \text{FirmChar}_{fi(t-2)} + \\
& + I(h = j) + \alpha_{ht} + \varepsilon_{fhijt}
\end{aligned}
\tag{3.2}
$$

[17]It is well known that the LMP is straightforward to estimate and easy to interpret, though it may not be the most appropriate specification for modeling binary outcomes because of its linearity and the potential for predicted probabilities to fall outside the [0,1] range. Nevertheless, when the primary interest lies in estimating the average partial effect, the LPM can yield meaningful results under certain conditions, as discussed in Chen et al. (2023)

On average, I find a positive association between geographic distance and the probability of exporting to a related party, and this relationship is consistent across all model specifications. The effect of distance between the headquarters and destination appears to be negative, although its statistical significance is weaker. Related-party exports to the headquarters country are more likely, which is consistent with previous findings in Wang (2021). I summarize these findings below in Fact 3:

Fact 3. *On average, there is positive association between geographic distance and the probability of exporting to a related party.*

Table 2: Probability of Related Party Trade by Foreign MNE's in Inida - OLS estimates

	Pr(Related party export to j in t = 1)					
	(1)	(2)	(3)	(4)	(5)	(6)
$log(dist_{ij})$	0.073***	0.062***	0.058***	0.063***	0.061***	0.056***
	(0.005)	(0.006)	(0.007)	(0.005)	(0.006)	(0.007)
$log(dist_{hj})$	-0.011*	-0.007	-0.007	-0.007*	-0.006	-0.006
	(0.004)	(0.005)	(0.006)	(0.004)	(0.004)	(0.005)
fta_{ijt}	0.095***	0.102***	0.098***	0.071***	0.072***	0.068***
	(0.009)	(0.009)	(0.009)	(0.008)	(0.009)	(0.009)
$log(gdp_{jt})$	0.087***	0.095***	0.096***	0.089***	0.091***	0.091***
	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
$log(pop_{jt})$	-0.031***	-0.031***	-0.029***	-0.029***	-0.028***	-0.026***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
wto_{jt}	0.012	0.014	0.011	0.022*	0.021*	0.017+
	(0.010)	(0.011)	(0.011)	(0.009)	(0.010)	(0.010)
$I(h = j)$	0.170***	0.196***	0.191***	0.192***	0.198***	0.185***
	(0.018)	(0.019)	(0.023)	(0.015)	(0.017)	(0.021)
number of market (t-2)		0.000	0.000			
		(0.000)	(0.000)			
number of goods (t-2)		0.001***	0.001***			
		(0.000)	(0.000)			
number of affiliates (t-2)		0.006	0.006			
		(0.006)	(0.006)			
R^2	0.164	0.185	0.185	0.403	0.396	0.397
Observations	134,429	97,525	97,177	134,429	97,525	97,177
Other Gravity	N	N	Y	N	N	Y
Headquarter Country - Year FE	✓	✓	✓			
Parent Firm (unit) - Year FE				✓	✓	✓

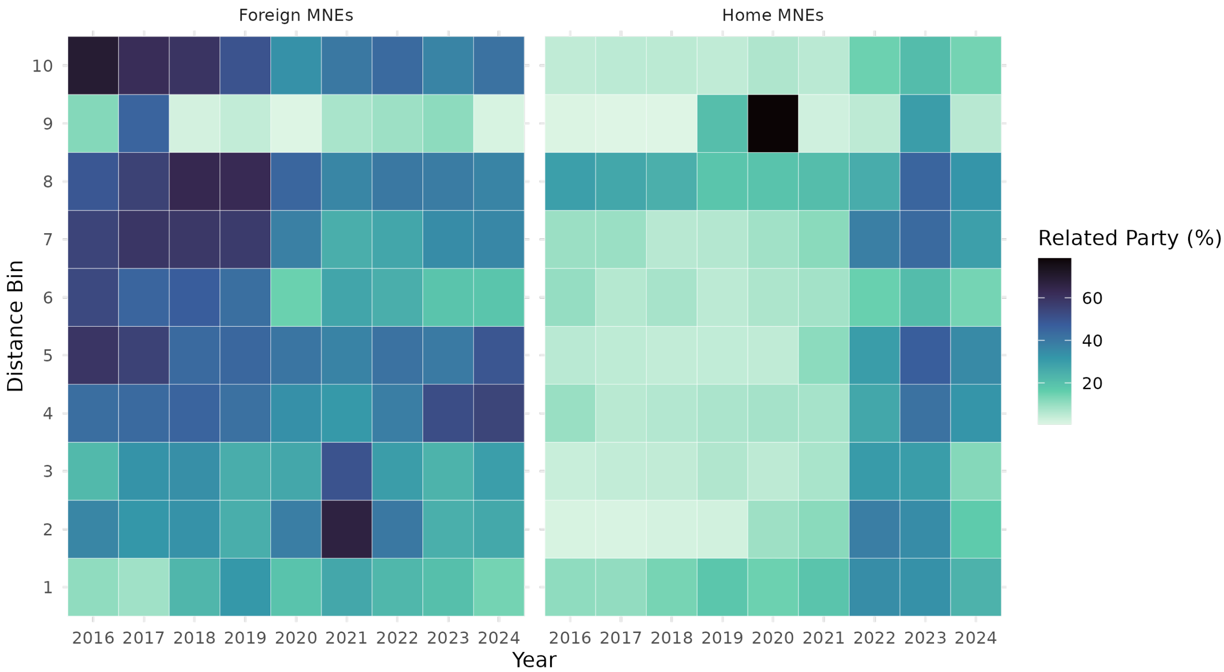
Signif. Codes: ***: 0.001, **: 0.01, *: 0.05. The unit of regression is the parent firm (headquarters), and standard errors are clustered at this level. $\log(dist_{hj})$ denotes the distance between the headquarters of a foreign MNE and destination j. $\log(dist_{ij})$ represents the distance between India and destination j. fta_{ijt} is a binary variable indicating whether a Free Trade Agreement exists between India and j at time t. $\log(gdp_{jt})$ and $\log(pop_{jt})$ denote the GDP and population of destination j, respectively. wto_{jt} equals 1 if destination j is a WTO member, and 0 otherwise. $I(h = j)$ is an indicator variable that equals 1 if the destination of export is the same as the headquarters country of the foreign MNE. *Other Gravity* variables include bilateral characteristics defined between India and destination j, as well as between the headquarters country h and destination j. These include colonial relationship, common language, contiguity, and shared legal origins.

The Evolution of Related-Party Export Figure 6 illustrates the share of related-party trade across distance bins. Distance refers to the great circle distance between the most populated cities of each country, as provided by the CEPII gravity database. The distances are sorted from shortest to longest and divided into ten equal-count bins, so that each bin contains roughly the same number of countries. Note that these bins do not represent equal distance intervals but still capture the ordering of distances. Distance bins correspond to deciles of the distribution, with detailed country lists provided in the Table A2. Export values are aggregated into ten distance bins, and the share of related-party exports is calculated for each bin.

Foreign MNE affiliates exhibited a relatively clear pattern in their related-party exports: a larger share of such trade occurred with geographically distant destinations, suggesting a positive association between distance and the intensity of related-party transactions. This pattern persisted until 2020 but has gradually disappeared since then. The data suggest that this change is driven by an increase in arm's-length exports rather than a decline in related-party exports (Figure A6). The same pattern holds when the related-party share is computed at the affiliate level with weight based on their total export to each bin (Figure A5). I summarize these findings below in Fact 4:

Fact 4. *Until 2020, arm's-length exports were more common at shorter distances, whereas related-party exports dominated at longer distances. However, this pattern has weakened in recent years.*

Figure 6: Related Party Trade Share by Distance Bin and Year - India



Notes: This figure illustrates the related-party trade share by distance for MNE firms headquartered outside of India versus those headquartered in India. The ratio is calculated at each bin using aggregated trade values. This analysis includes multinational firms only. For India, the seventh bin includes West and Southern African countries such as South Africa and Nigeria. The eighth bin covers Canada and United States, Oceania and some West African countries.

4 Theoretical Framework

In this section, I develop a partial-equilibrium model of related-party and arm's-length exports that captures the stylized facts and provides the foundation for the corresponding estimating equations. The framework is a direct application of Eaton and Kortum (2002) at the firm level, featuring a representative downstream buyer and multiple suppliers across countries. In each destination country $j \in J$, a representative buyer produces and sells a fixed quantity of a final good using imported inputs. The buyer minimizes total input costs by sourcing varieties from multiple origin countries $n \in N$ and from a large set of potential suppliers $i \in S_n$ within each origin. Each supplier produces a continuum of differentiated varieties ω and can serve the buyer through one of two channels (modes): *related-party* (within-firm) or *arm's-length* (between-firm) exports. For a given supplier, the sets of varieties exported under each mode are distinct, reflecting mode-specific productivity draws and trade costs.

4.1 Technology

Each supplier i in source country n draws idiosyncratic productivity terms $z^m_{nij}(\omega)$ independently across varieties ω and modes $m \in \{\mathrm{W}, \mathrm{B}\}$ from a Fréchet distribution:

$$\Pr(z^m_{ni}(\omega) \le z) = \exp[-T_{ni} z^{-\theta}], \qquad T_{ni} > 0,\ \theta > 1.$$

This implies that suppliers' relative efficiency of each mode is random across goods. A supplier may be relatively more efficient at producing certain varieties for related-party exports and other varieties for arm's-length exports even though both come from the same underlying technology. The unit production cost in source country n is denoted by c_n. The delivered price of variety ω from supplier i in n to destination j via mode m is therefore

$$p^m_{nij}(\omega) = \frac{c_n}{z^m_{ni}(\omega)}\, \tau^m_{nj},$$

where τ^m_{nj} represents the iceberg trade cost associated with mode m.

4.2 Demand

Final-good producers in j combine a continuum of varieties $\omega \in \Omega$ via

$$Q_j \;=\; \left(\int_{\Omega} q_j(\omega)^{\frac{\sigma-1}{\sigma}}\, d\omega \right)^{\frac{\sigma}{\sigma-1}}. \tag{4.1}$$

Final good producer at j is selling an exogenous quantity of final product and cost-minimizing the a continuum of intermediate inputs it's sourcing from multiple sources either from arms length suppliers or from affiliate.

4.3 Equilibrium

The equilibrium is characterized by the equilibrium price, sourcing probabilities of suppliers that correspond to the expenditure shares of destinations, and the firm-to-destination-level gravity equation. Detailed derivations are provided in Appendix C.3.

4.3.1 Price

Supplier side price offer. Denote the probability that supplier i in n is able to offer country j good ω via export channel (mode) m for a price less than price p as $G^m_{nij}(p)$.

$$\begin{aligned} G^m_{nij}(p) &\equiv \Pr\Big(p^m_{nij}(\omega) \le p\Big) = \Pr\left(\frac{c_n}{z^m_{ni}(\omega)}\tau^m_{nj} \le p\right) = \Pr\left(z^m_{ni}(\omega) \ge \frac{c_n\tau^m_{nj}}{p}\right) \\ &= 1 - \exp\Big\{- T_{ni}\,(c_n\tau^m_{nj})^{-\theta}\,p^\theta\Big\}. \end{aligned}$$

Cheapest price distribution that is accepted by buyer The lowest price is the realization of p_j is the minimum delivered price among all supplier-mode pairs available to buyers in j which is expressed as $p_j(\omega) \equiv \min_{ni,m}\{p^m_{nij}(\omega)\}$. In other words, buyer chooses the cheapest across modes (channels) and source countries. Assuming all prices are independent and all buyers in j are identical, I can derive the distribution of prices across goods for buyers in country j.

$$\begin{aligned} G_j(p) &= \Pr\big(p_j(\omega) \le p\big) \\ &= \Pr\left(\min_{ni,m}\{p^m_{nij}(\omega)\} \le p\right) \\ &= 1 - \exp\Big\{-\Phi_j\,p^\theta\Big\} \quad \text{where } \Phi_j \;\equiv\; \sum_{n\in N}\sum_{i\in S_n}\sum_{m\in\{W,B\}} T_{ni}\,(c_n\,\tau^m_{nj})^{-\theta} \end{aligned}$$

4.3.2 Total expenditure in j on a supplier's channel

Under CES demand system, we derive Marshallian input demand (derivation in Appendix C.1) with price index $P_j = \left(\int_\Omega p_j(\omega)^{1-\sigma}\,d\omega\right)^{\frac{1}{1-\sigma}}$

$$q_j(\omega) \;=\; \frac{E_j}{P_j}\left(\frac{p_j(\omega)}{P_j}\right)^{-\sigma} \tag{4.2}$$

Integrating the demand in equation (4.2) over the set of varieties sourced from supplier i in country n via channel (mode) m yields the total expenditure

$$X_{nij}^{m} = \int_{\omega \in \Omega_{nij}^{m}} p_j(\omega)\, q_j(\omega)\, d\omega = E_j \cdot \underbrace{\Pr\Big(p_{nij}^{m}(\omega) = p_j(\omega)\Big)}_{\text{sourcing probability}}.$$

4.3.3 Sourcing probability

Next, I derive the probability that supplier i in source country n, using mode m, is the lowest-cost provider of variety ω in destination j. This probability, which I refer to as the *sourcing probability*, represents the likelihood that supplier i in source country n using mode m offers the minimum delivered price among all potential supplier-channel pair. Recall that firm i has two channels through which it can export goods, denoted by modes $m \in \{W, B\}$, where W represents exports to related parties (*W*ithin-firm) and B represents exports to independent buyers (*B*etween-firm, or arm's-length, trade).

Because there is a continuum of goods and each good receives an independent and identically distributed mode-specific productivity draw, the law of large numbers ensures that the share of varieties for which supplier i in source country n using mode m is the lowest-cost provider equals the probability of being the lowest-cost supplier for any given variety.

The fraction of goods that firm i exports to destination j using mode $m = W$ corresponds to the share of its exports conducted through related-party transactions.

$$\begin{aligned} \Pi_{nij}^{W} &\equiv \Pr\Big(p_{nij}^{W}(\omega) \leq \min_{(nk,m) \neq (ni,W)} p_{nkj}^{m}(\omega) \Big) \\ &= \frac{T_{ni}(c_n \tau_{nj}^{W})^{-\theta}}{\sum_{n \in N} \sum_{i \in S_n} \sum_{m \in \{W,B\}} T_{ni}\,(c_n\, \tau_{nj}^{m})^{-\theta}} \end{aligned} \tag{4.3}$$

Same applies for the case m = B, hence the fraction of goods that firm i in n export to j using mode (channel) $m = B$:

$$\Pi_{nij}^{B} = \frac{T_{ni}(c_n \tau_{nj}^{B})^{-\theta}}{\sum_{n \in N} \sum_{i \in S_n} \sum_{m \in \{W,B\}} T_{ni}\,(c_n\, \tau_{nj}^{m})^{-\theta}} \tag{4.4}$$

4.3.4 Trade Flow by Channel (Firm-to-Destination)

Finally, we obtain the expenditure share of country j on supplier i via channel (mode) m equals its sourcing probability:

$$\frac{X_{nij}^m}{E_j} = \Pi_{nij}^m \tag{4.5}$$

where $m \in \{W, B\}$ and E_j is total expenditure by buyer j.

This yields the the expenditure in j on goods sold by firm i in source country n:

$$X_{nij}^m = E_j \, \Pi_{nij}^m = E_j \, \frac{T_{ni}(c_n \tau_{nj}^m)^{-\theta}}{\Phi_j}.$$

Substituting $\Phi_j = P_j^{-\theta} C^{\theta}$ gives

$$X_{nij}^m = E_j \, T_{ni} \, \left(P_j^{-1} \, c_n \, \tau_{nj}^m\right)^{-\theta} C^{-\theta} \tag{4.6}$$

where P_j is price index in the destination market derived in the Eaton and Kortum (2002), $C = \Gamma\left(\frac{\theta+1-\sigma}{\theta}\right)^{\frac{1}{1-\sigma}}$ is constant.

4.4 Trade Costs

I parameterize iceberg costs by mode $m \in \{\mathrm{W}, \mathrm{B}\}$ as

$$\tau_{nj}^m = d_{nj}^{\delta_m} \, \exp\{\, z_{nj}' \Gamma_m \,\}, \tag{4.7}$$

where d_{nj} is the bilateral distance between source n and destination j, and δ_m is the mode-specific distance elasticity of trade costs. The vector z_{nj} collects other bilateral factors (e.g., contiguity, common language, or FTA), and Γ_m is a conformable vector of mode-specific loadings. This specification ensures $\tau_{nj}^m > 0$ and allows dummy or zero-valued covariates to enter linearly within the exponential.

Comparative statistics. Substituting (4.7) into the sourcing probabilities (4.4) and (4.5), the ratio of within- to between-firm exports becomes

$$\frac{X_{nij}^W}{X_{nij}^B} \;\equiv\; R_{nij} \;=\; \frac{\Pi_{nij}^W}{\Pi_{nij}^B} \;=\; \frac{T_{ni}(c_n \tau_{nj}^W)^{-\theta}}{T_{ni}(c_n \tau_{nj}^B)^{-\theta}} \;=\; \left(\frac{\tau_{nj}^W}{\tau_{nj}^B}\right)^{-\theta}.$$

Using (4.7),

$$R_{nij} = \left(\frac{d_{nj}^{\delta_W} \exp\{z_{nj}' \Gamma_W\}}{d_{nj}^{\delta_B} \exp\{z_{nj}' \Gamma_B\}} \right)^{-\theta} = \exp\Big\{ -\theta \big[(\delta_W - \delta_B) \ln d_{nj} + z_{nj}' (\Gamma_W - \Gamma_B) \big] \Big\}. \tag{4.8}$$

Taking logarithms,

$$\ln R_{nij} = -\theta(\delta_W - \delta_B) \ln d_{nj} \; - \; \theta\, z_{nj}' (\Gamma_W - \Gamma_B).$$

Differentiating with respect to $\ln d_{nj}$ yields

$$\frac{\partial \ln R_{nij}}{\partial \ln d_{nj}} = -\theta(\delta_W - \delta_B).$$

Implication. If related-party (within-firm) exports are *less* sensitive to distance than arm's-length exports, so that $\delta_W < \delta_B$, then $\frac{\partial \ln R_{nij}}{\partial \ln d_{nj}} > 0$. In this case, the ratio $\frac{X_{nij}^W}{X_{nij}^B}$ increases with distance, consistent with the stylized fact that related-party exports tended to dominate at longer distances during 2016–2020. If the difference in distance elasticities between the two channels ($\delta_W - \delta_B$) has declined over time, this could help explain the attenuation of that pattern observed after 2020.

5 Empirical Implementation

5.1 Empirical specification

Deriving estimable equation Plugging (4.7) to (4.6) gives:

$$X_{nij}^m = E_j \frac{T_{ni} (c_n \tau_{nj}^m)^{-\theta}}{\Phi_j} = E_j \, T_{ni} \, \left(P_j^{-1} \, c_n \, d_{nj}^{\delta_m} \, \exp\{ \, Z_{nj}' \Gamma_m \, \} \right)^{-\theta} C^{-\theta} \tag{5.1}$$

This structural equation holds for each observation exactly, which may not be empirically satisfied. For empirical studies, it is common to interpret (5.1) in a conditional mean sense as follows: Let L be the collection of random variables appear on right hand side of (5.1).

$$E[X_{nij}^m | L] = E_j \, T_{ni} \, \left(P_j^{-1} \, c_n \, d_{nj}^{\delta_m} \, \exp\{ \, Z_{nj}' \Gamma_m \, \} \right)^{-\theta} C^{-\theta} \tag{5.2}$$

The conditional mean model (5.2) relaxes the restriction in (5.1) and implies that the structural relationship holds only in expectation. Let ε_{nij}^m denote a stochastic error term such that $E[\varepsilon^m_{nij} \mid L] = 1$ for both m types. Then, by construction, we have[18]

$$X^m_{nij} = E_j\, T_{ni}\, \left(P_j^{-1}\, c_n\, d_{nj}^{\delta_m}\, \exp\{\, Z'_{nj}\Gamma_m \,\} \right)^{-\theta} C^{-\theta} \varepsilon^m_{nij}, \quad E[\varepsilon^m_{nij}|L] = 1 \tag{5.3}$$

To estimate the parameters in (5.3), a common practice is to log-linearize it first and estimate the linear model by OLS. However, as pointed out by Santos Silva and Tenreyro (2006), the OLS estimation can be inconsistent and biased. To see that, we note that the error term that appears in the log-linear model is in the form of $\ln \varepsilon$:

$$\begin{aligned} \ln X^m_{nij} = {} & \underbrace{\ln E_j + \theta \ln P_j}_{\text{destination FE },\, \varsigma_j} - \underbrace{\theta \ln C}_{\textit{constant}} + \underbrace{\ln T_{ni}}_{\text{supplier FE },\, \psi_i} - \underbrace{\theta \ln c_n}_{\text{source FE },\ \ln \eta_n} \\ & - \theta\, \delta^m\, \ln d_{nj} - \theta \sum_{z \in \mathcal{Z}} \gamma^m_z\, z_{nj,z} + \ln \varepsilon^m_{nij} \end{aligned} \tag{5.4}$$

For OLS to be consistent, it necessitates the uncorrelatedness between $\ln \varepsilon$ and L, which is not guaranteed by $E[\varepsilon^m_{ijt}|L] = 1$. A sufficient condition is independence between ε and L but this is challenged by common heteroskedasticity of the error term found in practice. Therefore, to proceed, we rewrite (5.3) by taking the natural exponential on both sides of (5.4):

$$X^m_{nij} = \exp\Big\{ \varsigma_j + \psi_i + \eta_n - \theta\, \delta_m\, \ln d_{nj} - \theta \sum_{z \in \mathcal{Z}} \gamma_{m,z}\, z_{nj,z} - \theta \ln C \Big\} \varepsilon^m_{nij}. \tag{5.5}$$

Single pooled equation with a mode indicator. Equation (5.5) implies a conditional mean model as a exponential function of a linear index. Note that the parameters in this model vary by type $m = W, B$. To test their difference empirically, one way is to pool the model across the two types. For the pooled model to be identified, we need to impose an extra exogeneity condition that ε^B_{nij} and ε^W_{nij} are independent conditional on L. This condition implies that the extra randomness introduced in (5.3) for two types are independent conditional on the observed characteristics. This is sensible in this setting because our structural model in (5.1) has already implied that the trade outcomes by types are completely captured by random variables in the collection L. However, we need to be aware that this restriction implies that random deviations of empirical observations from the structural model should not arise from common shocks that simultaneously affect both types once L is controlled for. In other words, the unexplained parts of trade for the two types shouldn't be caused by the common shock. To prove why this exogeneity condition is sufficient, I illustrate it using a stylized model for the sake of notation simplicity as follows:

$$X_m = L^{\alpha_m} \varepsilon_m, \quad E[\varepsilon_m|L] = 1 \text{ for } m = W, B \tag{5.6}$$

[18] This follows from rewriting the additive error specification in multiplicative form.

Define the binary mapping between mode and indicator

$$\ell = \mathbf{1}\{m = W\} \;=\; \begin{cases} 1, & \text{if } m = W, \\ 0, & \text{if } m = B. \end{cases}$$

We note that this is a stylized model corresponding to (5.3). We can rewrite the model two types as follows

$$X = X_W^{\ell} X_B^{1-\ell} = (L^{\alpha_W} \varepsilon_W)^{\ell} (L^{\alpha_B} \varepsilon_B)^{1-\ell} \tag{5.7}$$

. By the conditional independence condition, we have

$$E[\varepsilon_W^{\ell} \varepsilon_B^{1-\ell} | L] = E[\varepsilon_W^{\ell} | L] \cdot E[\varepsilon_B^{\ell} | L] = 1.$$

Therefore, we can denote $\epsilon = \varepsilon_W^{\ell} \varepsilon_B^{1-\ell}$.

Take log-transform,

$$lnX = \ell(\alpha_W lnL) + (1-\ell)(\alpha_B lnL) = \ell(\alpha_W - \alpha_B) lnL + \alpha_B lnL$$

and then taking exponential as above and substitute ℓ with indicator notation, which gives

$$X = \exp((\alpha_W - \alpha_B) \ln L \cdot \mathbf{1}\{m = W\} + \alpha_B \ln L)\epsilon, \;\; E[\epsilon | L] \; = 1 \tag{5.8}$$

which allows us to test if $\alpha_W - \alpha_B$ is 0 empirically.

Using full notation, the single equation can be defined as follows with $E[\epsilon_{nijm} | L] = 1$:

$$\begin{aligned} X_{nijm} = \exp\Bigg\{ & \varsigma_j + \psi_i + \eta_n + \underbrace{\beta_d^{\Delta} \left(\ln d_{nj}\right) \cdot 1\{m = W\}}_{\text{W–B difference}} + \underbrace{\beta_d^{B} \ln d_{nj}}_{\text{baseline (B)}} \\ & + \sum_{z \in \mathcal{Z}} \left[\beta_z^{\Delta} \left(z_{nj,z} \cdot 1\{m = W\} \right) + \beta_z^{B} z_{nj,z} \right] + \beta_0 1\{m = W\} \Bigg\} \epsilon_{nijm}. \end{aligned}$$

Conditional mean:

$$\begin{aligned} E[X_{nijm}] = \exp\Bigg\{ & \varsigma_j + \psi_i + \eta_n + \underbrace{\beta_d^{\Delta} \left(\ln d_{nj}\right) \cdot 1\{m = W\}}_{\text{W–B difference}} + \underbrace{\beta_d^{B} \ln d_{nj}}_{\text{baseline (B)}} \\ & + \sum_{z \in \mathcal{Z}} \left[\beta_z^{\Delta} \left(z_{nj,z} \cdot 1\{m = W\} \right) + \beta_z^{B} z_{nj,z} \right] + \beta_0 1\{m = W\} \Bigg\} \end{aligned} \tag{5.9}$$

The structural mappings implied by the arm's length as baseline (β_d^B) and the difference in distance elasticity β_d^Δ:

$$\beta_d^B = -\theta\,\delta_B, \quad \beta_d^\Delta = -\theta\,(\delta_W - \delta_B), \quad \beta_z^B = -\theta\,\gamma_{B,z}, \quad \beta_z^\Delta = -\theta\,\left(\gamma_{W,z} - \gamma_{B,z}\right). \tag{5.10}$$

5.2 Estimation

Following Gourieroux et al. (1984b,a); Wooldridge (1999); Santos Silva and Tenreyro (2006, 2011), we estimate the β family parameters in (5.10) using Poisson pseudo–maximum likelihood (PPML), interpreting the structural relationship as a restriction on the conditional mean. Let $\mu(x;\beta) \equiv \mathbb{E}[X \mid L]$ given by (5.9). In the pseudo or quasi–maximum likelihood framework Gourieroux et al. (1984b), the estimator chooses β so that the sample moment conditions

$$\sum_{n,i,j,m} x_{nijm}\left[X_{nijm} - \mu(x_{nijm};\beta)\right] = 0 \tag{5.11}$$

hold.[19] We report robust standard errors, which are valid under general forms of heteroskedasticity and PPML naturally accommodates zero outcomes (Santos Silva and Tenreyro, 2006; Wooldridge, 2023) even if when the zero outcome observations are large (Santos Silva and Tenreyro, 2011).

The regression dataset includes the top 60 destination countries for India, and Mexico from 2016 to 2024. Destination countries are selected annually, and all MNE affiliates that export to these destinations are included in the data. Restricting the sample to major destination countries helps address the lumpiness of the data. The dataset consists of unique firm–product pairs across all selected destinations. If a firm–product pair (i.e., a firm exporting a specific product of HS 6-digit) is observed for some destinations, zero-export observations are included for other destinations. Due to computational constraints associated with the non-linear estimation, regressions are estimated separately for each year. Therefore, any shocks common to all observations within a year are constant and do not affect identification. The number of observations included in the regression per year ranges from approximately 6.1 million to 8.5 million. [20]

[19] Poisson regression is a class of models that employs the likelihood function of the Poisson distribution but does not impose any specific restriction on the variance. The Poisson Pseudo–Maximum Likelihood (PPML) estimator is a maximum likelihood estimator based on this model. Since the first-order condition of the PPML estimator coincides with the sample moment condition given by (5.11), the estimator that solves this condition is referred to as the PPML estimator(Jeff's citation needed)

[20] The number of observations used in the PPML regressions by year is as follows: 2016 – 6,129,173; 2017 – 6,406,351; 2018 – 7,326,832; 2019 – 7,264,076; 2020 – 6,774,926; 2021 – 7,378,666; 2022 – 7,430,890; 2023 – 7,680,547; and 2024 – 8,516,207.

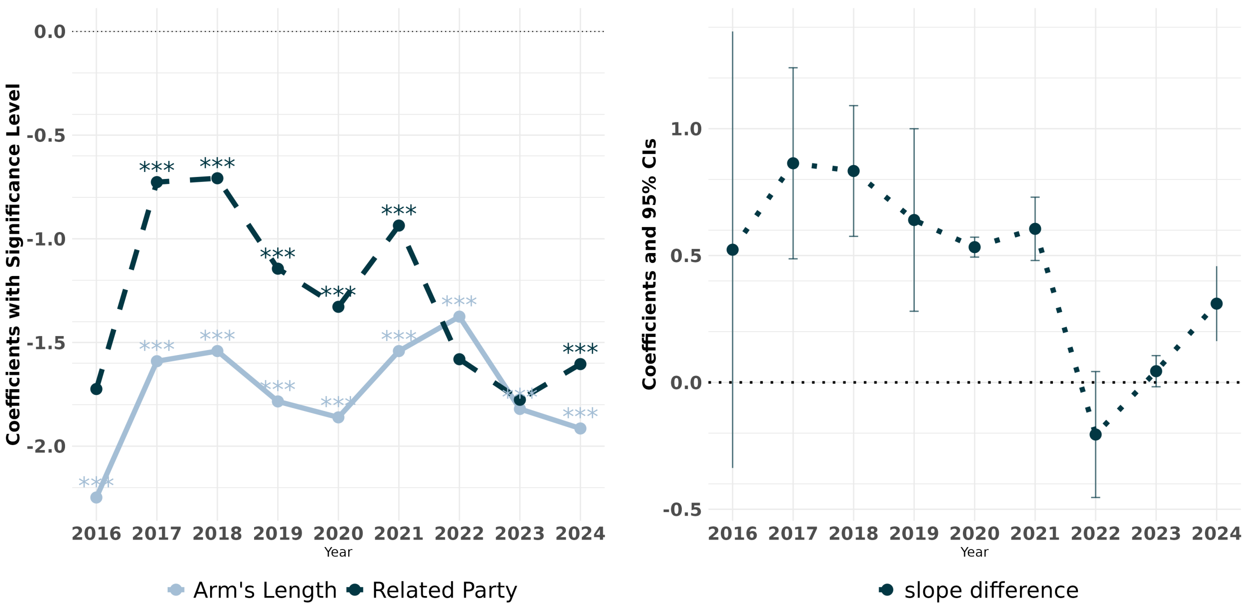
5.3 Results

The PPML estimates indicate that related-party exports are less elastic to distance than arm's-length exports, but this difference has eroded in recent years. Figure 7 presents the results. The left panel reports the coefficients β_d^B and $\beta_d^B + \beta_d^\Delta$, while the right panel shows β_d^Δ, which captures the difference in distance (semi-)elasticities between the two export modes. Because distance elasticities are negative, a positive value of β_d^Δ implies that related-party exports are less elastic (or less sensitive) to distance than arm's-length exports. In other words, at a given distance and holding all other factors constant, related-party exports exhibit higher trade volumes relative to the baseline. The results suggest that related-party exports were significantly less sensitive to distance during 2017–2021, whereas no significant difference is found in 2016. However, this difference has weakened over the past three years (2022–2024). The estimates of β_d^Δ from 2017 to 2021 are 0.864, 0.833, 0.640, 0.533, and 0.605, respectively, indicating that, at a given distance, related-party exports were about 137%, 130%, 90%, 70%, and 83% higher than the baseline arm's-length exports.[21] The magnitude of the gap between related-party and arm's-length exports has gradually converged toward zero since 2022. Consistent with this pattern, Figure 6 shows that the share of related-party exports to distant destinations has declined in recent years. Overall, these findings suggest that the advantage of exporting to distant markets through related-party channels had been gradually eroding but experienced an abrupt drop between 2021 and 2022, with a modest rebound emerging in 2024.

Figure A10 presents the estimated effects of additional proxies for trade costs. We find that the presence of a Free Trade Agreement (FTA) between countries has a positive effect on related-party trade, although the statistical significance varies and is stronger before 2020. This pattern may reflect the fact that relatively few new FTAs have been established in recent years. Interestingly, the effect of FTAs on arm's-length trade is not statistically different from zero across all years. A shared official language is positively associated with related-party trade, while historical colonial ties appear to be more relevant for arm's-length trade, though these effects are statistically significant only in the most recent two years. Finally, while geographic contiguity has a negative effect on arm's-length trade, it is positively associated with related-party trade. The evolution of statistical significance over time follows a similar trend to that observed for the coefficients on (log) distance. Overall, gravity variables other than distance exert limited influence on arm's-length exports, whereas several continue to play a meaningful role in shaping related-party exports.

[21]The estimates can be translated into semi-elasticities, calculated as $(e^{\hat{\beta}} - 1) \times 100$.

Figure 7: Heterogeneous distance (semi-)elasticity: related-party vs. arm's-length export



(a) Distance Elasticity (b) Difference in Distance Elasticity

Notes: Robust standard errors are multi-way clustered by MNE affiliate, source country, product, and destination to account for correlated errors across multiple dimensions.

6 Conclusion

Using shipment-level Bill of Lading data from India, I document four stylized facts about the export behavior of foreign MNE affiliates. First, foreign MNEs sell through both channels—related-party and arm's-length—to foreign markets via their affiliates in India. Second, they typically export distinct sets of products across the two modes. Third, the probability of engaging in related-party exports is positively correlated with distance. Fourth, related-party exports were disproportionately directed toward distant markets through 2020, but this pattern has weakened in recent years. Taken together, these facts suggest that the distance elasticity of exports differs between internal trade within MNE networks and arm's-length trade, but that this gap has narrowed over time.

Guided by these facts, I develop a partial-equilibrium model that rationalizes these patterns and yields an estimable firm-level gravity equation allowing for heterogeneous distance elasticities between related-party and arm's-length exports. Estimating this equation using Poisson Pseudo-Maximum Likelihood, I find that related-party exports are less elastic to distance than arm's-length exports, but that this difference has eroded over time—consistent with the recent surge in inter-firm trade by foreign affiliates of multinational firms.

An additional pattern emerging from the data is a marked increase in the trade volume of home-country MNEs in recent years, both within and across firms. This trend may reflect broader supply-chain reorganization and shifting demand patterns surrounding India. Investigating the underlying economic forces behind this trend remains an important direction for future research.

References

Alfaro, Laura and Andrew Charlton, "Intra-Industry Foreign Direct Investment," *American Economic Review*, December 2009, *99* (5), 2096–2119.

— **and Maggie X. Chen**, "Selection and Market Reallocation: Productivity Gains from Multinational Production," *American Economic Journal: Economic Policy*, May 2018, *10* (2), 1–38.

— **, Paola Conconi, Fariha Kamal, and Zachary Kroff**, "Trade within Multinational Boundaries," *National Bureau of Economic Research*, June 2025.

Alfaro-Ureña, Alonso, Isabela Manelici, and Jose P Vasquez, "The Effects of Joining Multinational Supply Chains: New Evidence from Firm-to-Firm Linkages," *The Quarterly Journal of Economics*, July 2022, *137* (3), 1495–1552.

Alfaro-Urena, Alonso, Isabela Manelici, and Jose P. Vasquez, "The Effects of Multinationals on Workers: Evidence from Costa Rican Microdata," *Working Papers*, April 2021, (285).

Alviarez, Vanessa and Ayhab F. Saad, "Multinational Production and Intra-Firm Trade," *Review of International Economics*, 2023, *31* (3), 826–853.

Antràs, Pol, Evgenii Fadeev, Teresa C. Fort, and Felix Tintelnot, "Exporting, Global Sourcing, and Multinational Activity: Theory and Evidence from the United States," *The Review of Economics and Statistics*, April 2024, pp. 1–48.

Atalay, Enghin, Ali Hortaçsu, Mary Jialin Li, and Chad Syverson, "How Wide Is the Firm Border?*," *The Quarterly Journal of Economics*, November 2019, *134* (4), 1845–1882.

Balsvik, Ragnhild, "Is Labor Mobility a Channel for Spillovers from Multinationals? Evidence from Norwegian Manufacturing," *The Review of Economics and Statistics*, 2011, *93* (1), 285–297.

Bombarda, Pamela and Stefania Marcassa, "Welfare and Trade Elasticity with Multinational Production," *The World Economy*, 2020, *43* (2), 388–411.

Bruno, Valentina and Hyun Song Shin, "Dollar and Exports," *The Review of Financial Studies*, August 2023, *36* (8), 2963–2996.

Cadestin, Charles, Koen De Backer, Isabelle Desnoyers-James, Sébastien Miroudot, Ming Ye, and Davide Rigo, "Multinational Enterprises and Global Value Chains: New Insights on the Trade-Investment Nexus," OECD Science, Technology and Industry Working Papers 2018/05, OECD March 2018.

Chen, Kaicheng, Robert S. Martin, and Jeffrey M. Wooldridge, "Another Look at the Linear Probability Model and Nonlinear Index Models," October 2023.

Conconi, Paola, Fabrizio Leone, Glenn Magerman, and Catherine Thomas, "Multinational Networks and Trade Participation," *Working Paper*, 2025, (456).

Davies, Ronald B., Julien Martin, Mathieu Parenti, and Farid Toubal, "Knocking on Tax Haven's Door: Multinational Firms and Transfer Pricing," *The Review of Economics and Statistics*, March 2018, *100* (1), 120–134.

Delpeuch, Samuel, Ana Cinta González Cabral, Felix Hugger, and Pierce O'Reilly, "MNE Business Functions and Corporate Taxation," OECD Taxation Working Papers January 2025.

Dixit, Avinash K. and Joseph E. Stiglitz, "Monopolistic Competition and Optimum Product Diversity," *The American Economic Review*, 1977, *67* (3), 297–308.

Do, Anh, Sharat Ganapati, Woan Foong Wong, and Oren Ziv, "Transshipment Hubs, Trade, and Supply Chains," *Working Paper*, 2025.

Eaton, Jonathan and Samuel Kortum, "Technology, Geography, and Trade," *Econometrica*, 2002, *70* (5), 1741–1779.

Finger, J. M. and M. E. Kreinin, "A Measure of 'Export Similarity' and Its Possible Uses," *The Economic Journal*, December 1979, *89* (356), 905–912.

Flaaen, Aaron, Flora Haberkorn, Logan Lewis, Anderson Monken, Justin Pierce, Rosemary Rhodes, and Madeleine Yi, "Bill of Lading Data in International Trade Research with an Application to the COVID-19 Pandemic," *Review of International Economics*, 2023, *31* (3), 1146–1172.

Ganapati, Sharat, Woan Foong Wong, and Oren Ziv, "Entrepôt: Hubs, Scale, and Trade Costs," *American Economic Journal: Macroeconomics*, October 2024, *16* (4), 239–278.

Garetto, Stefania, Xiao Ma, Lindsay Oldenski, and Natalia Ramondo, "Multinational Expansion in Time and Space," *Working Paper*, 2024.

Garg, Shresth, Pulak Ghosh, and Brandon Joel Tan, "Within Firm Supply Chains: Evidence from India," *Journal of International Economics*, September 2023, *144*, 103793.

Ge, Ying, Huiwen Lai, and Susan Chun Zhu, "Multinational Price Premium," *Journal of Development Economics*, July 2015, *115*, 181–199.

Gourieroux, C., A. Monfort, and A. Trognon, "Pseudo Maximum Likelihood Methods: Applications to Poisson Models," *Econometrica*, 1984, *52* (3), 701–720.

__ , __ , and __ , "Pseudo Maximum Likelihood Methods: Theory," *Econometrica*, 1984, *52* (3), 681–700.

Head, Keith and Thierry Mayer, "Brands in Motion: How Frictions Shape Multinational Production," *American Economic Review*, September 2019, *109* (9), 3073–3124.

Hong, YongKi, "Related-Party Trades in Vertical Integration," *Working Paper*, 2023.

Irarrazabal, Alfonso, Andreas Moxnes, and Luca David Opromolla, "The Margins of Multinational Production and the Role of Intrafirm Trade," *Journal of Political Economy*, February 2013, *121* (1), 74–126.

Javorcik, Beata Smarzynska, "Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers Through Backward Linkages," *American Economic Review*, June 2004, *94* (3), 605–627.

Keller, Wolfgang, "Chapter 19 - International Trade, Foreign Direct Investment, and Technology Spillovers," in Bronwyn H. Hall and Nathan Rosenberg, eds., *Handbook of the Economics of Innovation*, Vol. 2 of *Handbook of the Economics of Innovation, Volume 2*, North-Holland, January 2010, pp. 793–829.

Krugman, Paul, "Scale Economies, Product Differentiation, and the Pattern of Trade," *The American Economic Review*, 1980, *70* (5), 950–959.

Lipsey, Robert E and Fredrik Sjöholm, "Foreign Direct Investment, Education and Wages in Indonesian Manufacturing," *Journal of Development Economics*, February 2004, *73* (1), 415–422.

Moran, Theodore H., "Foreign Direct Investment and Development: Launching a Second Generation of Policy Research: Avoiding the Mistakes of the First, Reevaluating Policies for Developed and Developing Countries," *Peterson Institute Press: All Books*, 2011.

_ , Edward M. Graham, and Magnus Blomstrom, "Does Foreign Direct Investment Promote Development?," *Peterson Institute Press: All Books*, 2005.

Ramondo, Natalia, Veronica Rappoport, and Kim J. Ruhl, "Intrafirm Trade and Vertical Fragmentation in U.S. Multinational Corporations," *Journal of International Economics*, January 2016, *98*, 51–59.

Ruhl, Kim J., "How Well Is US Intrafirm Trade Measured?," *American Economic Review*, May 2015, *105* (5), 524–529.

Setzler, Bradley and Felix Tintelnot, "The Effects of Foreign Multinationals on Workers and Firms in the United States*," *The Quarterly Journal of Economics*, August 2021, *136* (3), 1943–1991.

Silva, J. M. C. Santos and Silvana Tenreyro, "The Log of Gravity," *The Review of Economics and Statistics*, November 2006, *88* (4), 641–658.

_ and _ , "Further Simulation Evidence on the Performance of the Poisson Pseudo-Maximum Likelihood Estimator," *Economics Letters*, August 2011, *112* (2), 220–222.

S&P Global, "Panjiva Supply Chain Intelligence User Guide," Technical Report 2024.

Wang, Zi, "Headquarters Gravity: How Multinationals Shape International Trade," *Journal of International Economics*, July 2021, *131*, 103477.

Wooldridge, Jeffrey M., "Quasi-Likelihood Methods for Count Data," in "Handbook of Applied Econometrics Volume 2: Microeconomics," John Wiley & Sons, Ltd, 1999, chapter 8, pp. 321–368.

Wooldridge, Jeffrey M, "Simple Approaches to Nonlinear Difference-in-Differences with Panel Data," *The Econometrics Journal*, September 2023, *26* (3), C31–C66.

Appendix

A Appendix - For Section 2

A.1 Process of Identifying Multinational

Figure A1: Data Manipulation and Database Merge Process

PanjivaID	V_1	...	V_N	CompanyID	Ultimate Parent ID	Multinational UP
p1013				c5678	189093	1
p1078				c4200	9998980	0
p2600				*NA*	*NA*	No match

m:1

PanjivaID	CompanyID	Ultimate Parent ID	N of CompanyID country	Multinational UP
....		189093	3	1
		9998980	1	0

m:1

PanjivaID	V_1	...	V_N
p1013			
p1078			
p2600			
...			

BoL (Trade Data) – Country Files

PanjivaD	CompanyID	CompanyID Country
p1013	c5678	US
p1024	c5678	US
p1058	c7100	Canada
p1078	c4200	Mexico
...		

S&P Capital IQ

CompanyID	Ultimate Parent ID
c5678	189093
c5678	189093
c7100	189093
c4200	9998980
...	

S&P Ultimate Parent

Notes:

A.2 Harmonized System (HS) 2 digit codes and Section group

This subsection provide further details on MNEs' exports by product group. Product groups are classified according to the Harmonized System (HS) Sections, as defined by the World Customs Organization (WCO). The last three sections in the table (Table A1) covering arms, miscellaneous manufactured goods, works of art, and unclassified items are excluded.

Table A1: HS 2-digit Codes and Corresponding Sections

HS 2-digit Codes	**Section Description**
01–05	Live Animals & Animal Products
06–14	Vegetable Products
15	Animal or Vegetable Fats and Oils
16–24	Prepared Foodstuffs; Beverages, Spirits & Tobacco
25–27	Mineral Products
28–38	Products of the Chemical or Allied Industries
39–40	Plastics & Articles Thereof; Rubber & Articles Thereof
41–43	Raw Hides, Skins, Leather, Furskins & Articles Thereof
44–46	Wood & Articles of Wood; Cork & Articles of Cork
47–49	Pulp of Wood or Other Fibrous Material; Paper & Paperboard
50–63	Textiles & Textile Articles
64–67	Footwear, Headgear, Umbrellas & Related Articles
68–70	Articles of Stone, Plaster, Cement, Ceramic & Glass
71	Natural or Cultured Pearls, Precious Stones & Metals
72–83	Base Metals & Articles Thereof
84–85	Machinery & Mechanical Appliances; Electrical Equipment
86–89	Vehicles, Aircraft, Vessels & Transport Equipment
90–92	Optical, Measuring, Medical Instruments
93	Arms & Ammunition; Parts & Accessories
94–96	Miscellaneous Manufactured Articles
97	Works of Art, Collectors' Pieces & Antiques

Note: HS 2-digit codes are grouped according to the Harmonized System (HS) Sections, as defined by the World Customs Organization (WCO). Section descriptions have been slightly abbreviated for clarity.

Figure A2: Export from India by Firm Type and By Product Group

Notes: Y-axis is export value in million USD. Each product group is based on the Harmonized System (HS) Sections, as defined by the World Customs Organization (WCO). Details on the group classification can be found in Table A1.

Figure A3: Export from India by Firm Type and By Product Group (Continued)

Notes: Y-axis is export value in million USD. Each product group is based on the Harmonized System (HS) Sections, as defined by the World Customs Organization (WCO). Details on the group classification can be found in Table A1.

A.3 Process of Identifying Related Party Trade

In this section, I outline how I match shipper and consignee names to categorize whether a shipment is related-party trade.

Entity cleaning. I harmonize shipper (exporter, sender) and consignee (importer, receiver) names before matching. Using case–insensitive regex, I remove country–specific legal forms and common suffixes, then strip trailing digits, stray punctuation/whitespace, and orphan trailing letters. Examples of removed tokens include *S.A.S., S.A.C., C.V., S.A., R.L., DE*, as well as *LTD/Ltda, LLC/Llc, Inc/INC, Pvt./PVT, Co./Corporation, International/Internacional, Trading*.[22]

Observations with empty cleaned names are dropped and I use cleaned name to compute the similarity between two name as follows:

Similarity metrics. For each shipment, I compute two complementary similarity measures:

- **Jaro–Winkler similarity** ($JW \in [0, 1]$), capturing character-level closeness between full cleaned strings (implemented with parameter $p = 0.1$).
- **Jaccard token similarity** ($J \in [0, 1]$), computed on sets of word tokens extracted from the cleaned strings.
  - For each pair of cleaned names a and b, I lowercase both strings and tokenize on non–word boundaries, remove empty tokens, and then form sets A and B of unique tokens. The Jaccard index is defined as

    $$J(a, b) = \frac{|A \cap B|}{|A \cup B|},$$

    This token–based Jaccard index is robust to reordered or duplicated words (e.g., "*Cute Orange*" vs. "*Orange Cute*") and this index is computed for all shipment pairs. [23]

Decision rule. A shipment is labeled **within-firm** ("related-party") if any of the following holds:

$$JW \geq 0.80, \quad \text{or}$$
$$0.70 \leq JW < 0.80 \text{ and } J > 0.10, \quad \text{or}$$
$$J \geq 0.50.$$

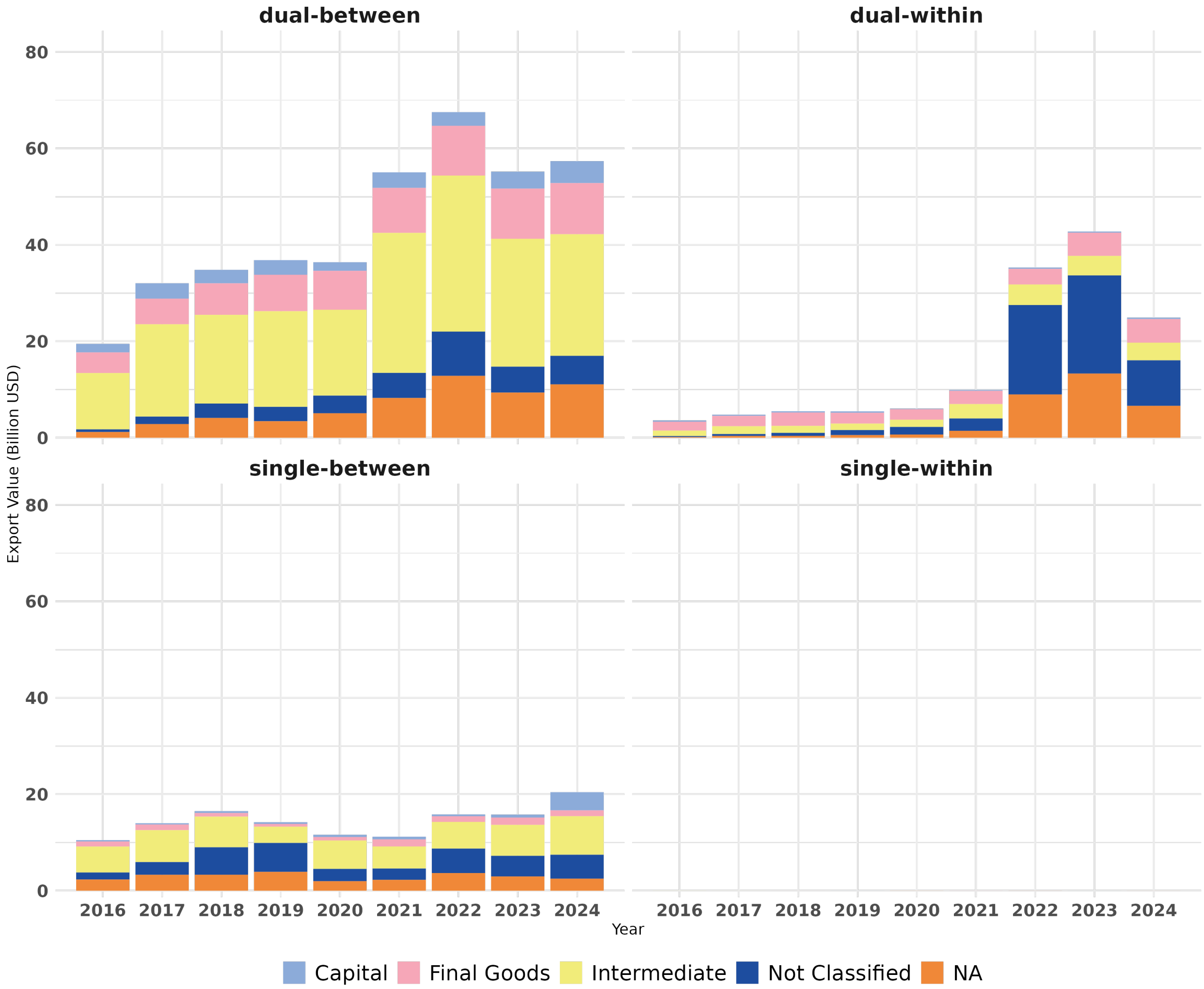
[22]Token lists differ by country (Mexico vs. India) to reflect local conventions.
[23]The baseline uses an unweighted set Jaccard for transparency and speed.

If none of the conditions hold, the shipment is labeled **between-firm** ("Arm's Length", or "across-firm").

Discussion. The rule first prioritizes high character-level similarity (JW), then allows medium JW when modest token overlap exists ($J > 0.10$), and finally captures cases with strong token overlap ($J \geq 0.50$) even when character-level similarity is lower—such as reordered, abbreviated, or partially truncated firm names. Because Jaro–Winkler and Jaccard capture different linguistic dimensions—JW measures fine-grained spelling similarity, while J focuses on shared word content—they complement each other: the former performs better for short or highly similar strings, whereas the latter is more robust to reordering and structural variation. Combining these metrics therefore allows robust detection of within-firm shipments across common spelling variants, token reorderings, and language differences. Thresholds (0.80, 0.70, 0.10, 0.50) were chosen conservatively through manual inspection and iterative trial-and-error to balance precision (avoiding false positives) and recall (capturing genuine within-firm matches). Future work may explore hybrid or machine-learning approaches, including natural language–based name matching. However, developing such models requires a sufficiently comprehensive and accurately labeled set of multinational firms and their subsidiaries—data that are only partially available through existing commercial sources such as Orbis or Panjiva. Building and validating such a dataset remains an important avenue for future research.

B Appendix - For Section 3

Figure A4: Export Value By Home MNEs affiliate establishments

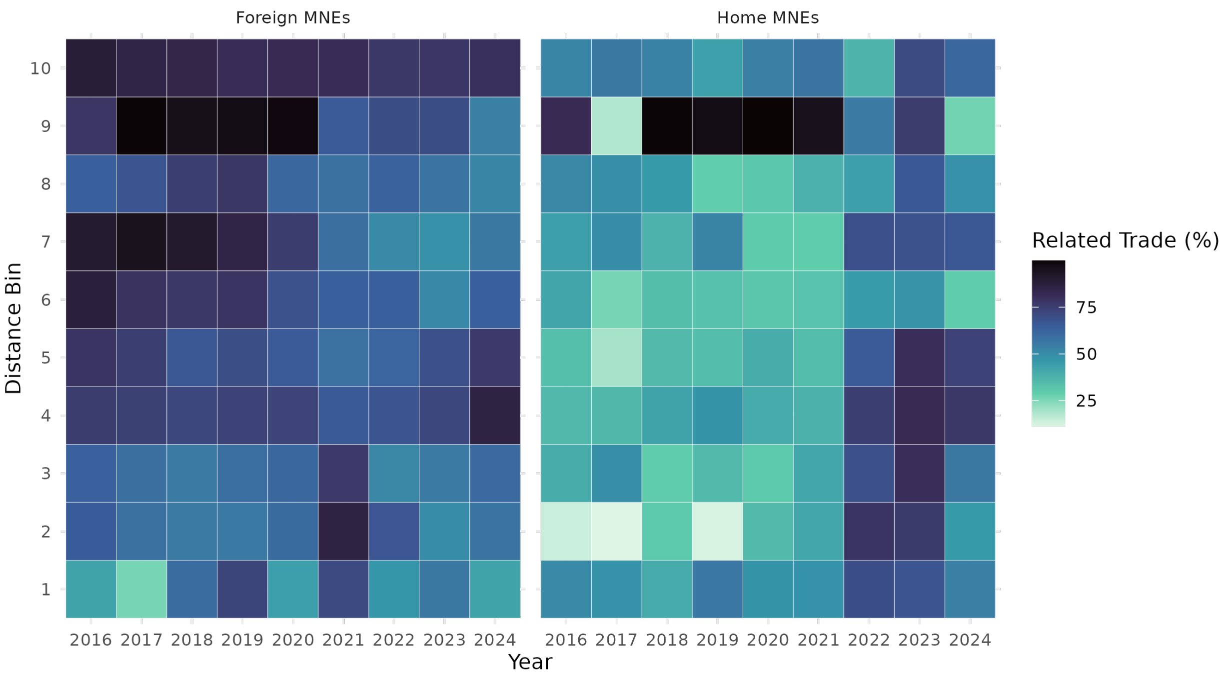


Notes: The y-axis is measured in billions of U.S. dollars. Classification is based on the concordance between HS codes and the Broad Economic Categories (BEC Rev. 4), which are then aggregated into three basic end-use categories defined by the United Nations System of National Accounts (SNA): capital goods, intermediate goods, and final (consumption) goods. Capital goods include machinery, equipment, dwellings, and other durable assets that are less disposable. "Not Classified" refers to items such as motor spirits (gasoline) and passenger motor vehicles, while "NA" indicates unmatched observations. In the labels, "within" denotes related-party trade and "between" denotes arm's-length trade. The category labeled "dual" indicates firms engaged in both types of exports (as shown in right panel (Home MNEs) in Figure 4), whereas "single" indicates firms operating in only one mode—either related-party or arm's-length exports. The category labeled "single-within" contains non-zero values that are not visible in the figure due to the axis scale.

Table A2: Countries by Distance Bin (India)

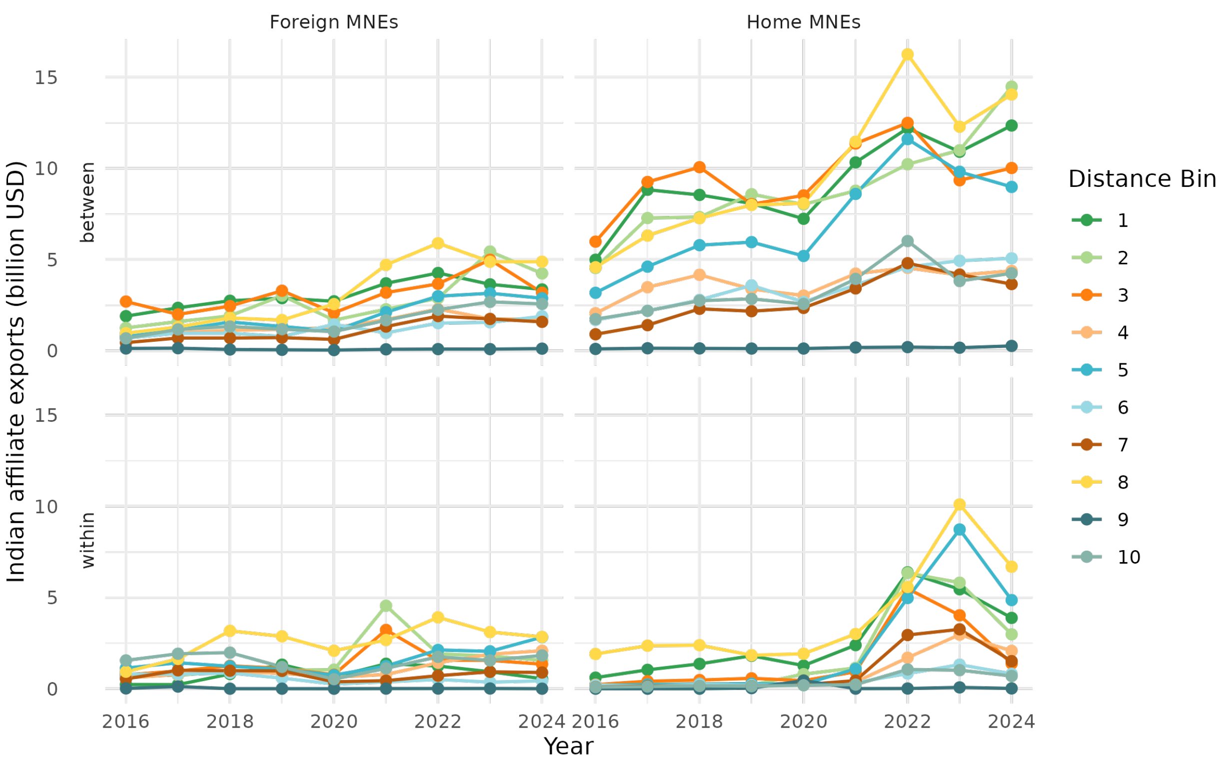
Distance Bin	Countries
1	United Arab Emirates, Nepal, Sri Lanka, Bhutan, Kuwait, Bangladesh, Bahrain, Qatar, Pakistan, Azerbaijan, Oman, Iran, Maldives, Myanmar, Turkmenistan, Uzbekistan, Laos, Tajikistan, Kazakhstan, Kyrgyzstan
2	Djibouti, Thailand, Israel, Hong Kong, Saudi Arabia, Singapore, Malaysia, Yemen, Vietnam, Iraq, Jordan, Cambodia, Lebanon, Mongolia, Georgia, Cyprus, Armenia, Macau, Eritrea, Syria
3	Indonesia, Somalia, Romania, Russia, Bulgaria, Ukraine, Egypt, Turkey, China, Lithuania, South Korea, Greece, Taiwan, Belarus, Philippines, Ethiopia, Seychelles, Latvia, Sudan, North Korea
4	Mauritius, Kenya, Tanzania, Germany, Slovenia, Serbia, Albania, Poland, Uganda, Croatia, Austria, Czech Republic, Sweden, Finland, Slovakia, Comoros, Estonia, Hungary, North Macedonia, Montenegro
5	Netherlands, Belgium, France, Japan, Italy, Denmark, Switzerland, Libya, Liechtenstein, Rwanda, Norway, Malta, Madagascar, Tunisia, Burundi, Réunion, Luxembourg, Monaco, San Marino
6	United Kingdom, Zambia, Portugal, Spain, Mozambique, Cameroon, Ireland, Zimbabwe, Chad, Algeria, Equatorial Guinea, Malawi, Gabon, Iceland, Central African Republic, Niger, Eswatini, Gibraltar, Andorra
7	South Africa, Nigeria, Mali, Sierra Leone, Guinea, Togo, Benin, Botswana, Burkina Faso, Ghana, Papua New Guinea, Angola, Lesotho, Morocco, Liberia, Namibia, Mauritania, Guinea-Bissau, Greenland
8	Canada, United States, New Zealand, Australia, Senegal, Fiji, Gambia, New Caledonia, Tonga, Bermuda, Vanuatu, Solomon Islands, Cape Verde, Anguilla, Montserrat, Marshall Islands, Kiribati, Tuvalu, Nauru
9	Suriname, Dominican Republic, Guyana, Jamaica, Venezuela, Cayman Islands, Haiti, Cuba, Bahamas, Saint Lucia, Netherlands Antilles, Guadeloupe, Barbados, Aruba, Grenada, Dominica, Martinique, Cook Islands, French Guiana
10	Falkland Islands, Mexico, Costa Rica, Brazil, Chile, Colombia, Ecuador, Guatemala, Panama, Paraguay, Peru, Argentina, Uruguay, Nicaragua, Honduras, El Salvador, Bolivia, Belize, French Polynesia

Figure A5: Related Party Trade Share by Distance Bin and Year - India, firm level weighted average



Notes: This figure illustrates the related-party trade share by distance for MNE firms headquartered outside of India versus those headquartered in India. The ratio is first calculated at the firm level. To obtain the average for each bin, I take a weighted mean where the weight is the firm's trade volume with each destination. In this way, destinations that account for more trade have a greater influence on the bin-level average.

Figure A6: Related Party Trade Volume by Distance Bin - India



Notes: ..

C Appendix - For Section 4

C.1 Demand and Price Index

Final-good producers in j combine a continuum of varieties $\omega \in \Omega$ with CES elasticity $\sigma > 1$.

$$Q_j = \left(\int_\Omega q_j(\omega)^{\frac{\sigma-1}{\sigma}} \, d\omega \right)^{\frac{\sigma}{\sigma-1}}. \tag{C.1}$$

Given prices $p_j(\omega)$, the producer chooses $\{q_j(\omega)\}_{\omega\in\Omega}$ to minimize cost for a target Q_j:

$$\min_{\{q_j(\cdot)\geq 0\}} \int_\Omega p_j(\omega) \, q_j(\omega) \, d\omega \quad \text{s.t.} \quad \left(\int_\Omega q_j(\omega)^{\frac{\sigma-1}{\sigma}} \, d\omega \right)^{\frac{\sigma}{\sigma-1}} \geq Q_j. \tag{C.2}$$

Solving Buyer's Problem Let λ_j denote the multiplier on the output constraint. The Lagrangian is

$$\mathcal{L} = \int_\Omega p_j(\omega) \, q_j(\omega) \, d\omega + \lambda_j \left[Q_j - \left(\int_\Omega q_j(\omega)^{\frac{\sigma-1}{\sigma}} \, d\omega \right)^{\frac{\sigma}{\sigma-1}} \right].$$

The first-order condition for any ω with $q_j(\omega) > 0$ is

$$p_j(\omega) = \lambda_j \left(\int_\Omega q_j(\tilde{\omega})^{\frac{\sigma-1}{\sigma}} \, d\tilde{\omega} \right)^{\frac{1}{\sigma-1}} \cdot q_j(\omega)^{-\frac{1}{\sigma}}.$$

From the CES definition in (C.1),

$$\int_\Omega q_j(\omega)^{\frac{\sigma-1}{\sigma}} \, d\omega = Q_j^{\frac{\sigma-1}{\sigma}}.$$

Substituting into the FOC gives

$$p_j(\omega) = \lambda_j \, Q_j^{\frac{1}{\sigma}} \, q_j(\omega)^{-\frac{1}{\sigma}}.$$

Rearranging yields

$$q_j(\omega) = \lambda_j^\sigma \, Q_j \, p_j(\omega)^{-\sigma}. \tag{C.3}$$

Plug (C.3) into aggregation constraint (C.1):

$$Q_j^{\frac{\sigma-1}{\sigma}} = \int_\Omega q_j(\omega)^{\frac{\sigma-1}{\sigma}} \, d\omega = \lambda_j^{\sigma-1} \, Q_j^{\frac{\sigma-1}{\sigma}} \int_\Omega p_j(\omega)^{1-\sigma} \, d\omega.$$

Hence,

$$\lambda_j^{\sigma-1} = \left(\int_\Omega p_j(\omega)^{1-\sigma} \, d\omega \right)^{-1}.$$

We can express $\lambda_j = P_j^{-1}$ by introducing Dixit-Stiglitz price index [24]:

Rewrite the input demand function by substituting $\lambda_j = P_j^{-1}$ back into (C.3) then we have Hicksian input demand function:

$$q_j(\omega) = Q_j \left(\frac{p_j(\omega)}{P_j} \right)^{-\sigma}. \tag{C.4}$$

Minimum cost and expenditure. Plugging (C.4) into total cost gives

$$C_j(Q_j) = \int_\Omega p_j(\omega)\, q_j(\omega)\, d\omega = Q_j P_j^\sigma \int_\Omega p_j(\omega)^{1-\sigma}\, d\omega = Q_j P_j^\sigma \cdot P_j^{1-\sigma} = Q_j P_j$$

Thus P_j is the unit expenditure function.

If total expenditure on the CES composite is E_j then $E_j = C_j(Q_j) = P_j Q_j$ and $Q_j = E_j / P_j$. Substituting into (C.4) gives the Marshallian demands:

$$q_j(\omega) = \frac{E_j}{P_j} \left(\frac{p_j(\omega)}{P_j} \right)^{-\sigma}$$

C.2 Equilibrium Price Index

This derivation is directly from Eaton and Kortum (2002). In equilibrium, price index can be rewritten as follows:

$$P_j^{1-\sigma} = \int_0^\infty p^{1-\sigma}\, dG_j(p) = \theta\, \Phi_j \int_0^\infty p^{\theta-\sigma} \exp\{-\Phi_j p^\theta\}\, dp,$$

where

$$\Phi_j \equiv \sum_{n \in N} \sum_{i \in S_n} \sum_{m \in \{W,B\}} T_{ni}\, (c_n\, \tau_{nj}^m)^{-\theta}.$$

Letting $x \equiv p^\theta \Phi_j$, and applying the change of variables $dx = \theta\, \Phi_j\, p^{\theta-1}\, dp$, I obtain

$$P_j^{1-\sigma} = \Phi_j^{\frac{1-\sigma}{\theta}} \int_0^\infty x^{\frac{1-\sigma}{\theta}} e^{-x}\, dx = \Phi_j^{\frac{1-\sigma}{\theta}} \Gamma\left(\frac{\theta + 1 - \sigma}{\theta} \right).$$

[24]The CES price index

$$P_j = \left(\int_\Omega p_j(\omega)^{1-\sigma}\, d\omega \right)^{\frac{1}{1-\sigma}}$$

is the unit expenditure function associated with the CES aggregator, representing the minimum cost of assembling one unit of the composite good from a continuum of varieties (Dixit and Stiglitz, 1977). This formulation is widely adopted in both monopolistic-competition and Ricardian trade models and seminal examples include Krugman (1980) and Eaton and Kortum (2002). Intuitively, the Lagrange multiplier λ_j in the cost-minimization problem equals the inverse of the unit price, $\lambda_j = 1/P_j$, as it represents the marginal cost of producing one additional unit of the CES composite.

Therefore,

$$P_j = \Phi_j^{-\frac{1}{\theta}} \left[\Gamma\left(\frac{\theta + 1 - \sigma}{\theta}\right)\right]^{\frac{1}{1-\sigma}}.$$

C.3 Equilibrium Derivation

C.3.1 Price

Supplier side price offer.

$$\begin{aligned}
G_{nij}^m(p) &\equiv \Pr\Big(p_{nij}^m(\omega) \leq p\Big) = \Pr\left(\frac{c_n}{z_{ni}^m(\omega)}\tau_{nj}^m \leq p\right) = \Pr\left(z_{ni}^m(\omega) \geq \frac{c_n \tau_{nj}^m}{p}\right) \\
&= 1 - \Pr\left(z_{ni}^m(\omega) < \frac{c_n \tau_{nj}^m}{p}\right) = 1 - \exp\left\{-T_{ni}\left(\frac{c_n \tau_{nj}^m}{p}\right)^{-\theta}\right\} \\
&= 1 - \exp\Big\{-T_{ni}\,(c_n \tau_{nj}^m)^{-\theta}\, p^\theta\Big\}.
\end{aligned}$$

Cheapest price distribution that is accepted by buyer

$$\begin{aligned}
G_j(p) &= \Pr\big(p_j(\omega) \leq p\big) \\
&= \Pr\left(\min_{ni,m}\{p_{nij}^m(\omega)\} \leq p\right) \\
&= 1 - \Pr\left(\min_{ni,m}\{p_{nij}^m(\omega)\} \geq p\right) \\
&= 1 - Pr\{p_{n1j}^W(\omega) \geq p,\ p_{n1j}^B(\omega) \geq p, \ldots,\ p_{NS_Nj}^W(\omega) \geq p,\ p_{NS_Nj}^B(\omega) \geq p\} \\
&= 1 - \prod_{n \in N} \prod_{i \in S_n} \prod_{m \in \{\mathrm{W,B}\}} \Big(1 - G_{nij}^m(p)\Big) \\
&= 1 - \exp\Big\{-\Phi_j\, p^\theta\Big\} \quad \text{where } \Phi_j \ \equiv \sum_{n \in N} \sum_{i \in S_n} \sum_{m \in \{W,B\}} T_{ni}\,(c_n\, \tau_{nj}^m)^{-\theta}
\end{aligned}$$

C.3.2 Sourcing probability

$$
\begin{aligned}
\Pi_{nij}^{W} &\equiv \Pr\Big(p_{nij}^{W}(\omega) \le \min_{(nk,m)\neq(ni,W)} p_{nkj}^{m}(\omega) \Big) \\
&= \int_0^\infty \Pr\Big(\min_{(nk,m)\neq(ni,W)} p_{nkj}^{m}(\omega) \ge p \,\Big|\, p_{nij}^{W}(\omega) = p \Big)\, dG_{nij}^{W}(p) \\
&= \int_0^\infty \prod_{(nk,m)\neq(ni,W)} \big(1 - G_{nkj}^{m}(p)\big)\, dG_{nij}^{W}(p) \\
&= \int_0^\infty \exp\left(-p^{\theta} \sum_{(nk,m)\neq(ni,W)} \kappa_{nkj}^{m} \right) \frac{d}{dp}\Big[1 - \exp\Big(-p^{\theta}\kappa_{nij}^{W}\Big)\Big]\, dp \\
&= \int_0^\infty \exp\left(-p^{\theta} \sum_{(nk,m)\neq(ni,W)} \kappa_{nkj}^{m} \right) \Big[\theta p^{\theta-1}\kappa_{nij}^{W} \exp\Big(-p^{\theta}\kappa_{nij}^{W}\Big)\Big] dp \\
&= \int_0^\infty \theta p^{\theta-1}\kappa_{nij}^{W} \exp\left[-p^{\theta}\Big(\sum_{(nk,m)\neq(ni,W)} \kappa_{nkj}^{m} + \kappa_{nij}^{W} \Big) \right] dp \\
&= \kappa_{nij}^{W} \int_0^\infty \theta p^{\theta-1} \exp(-p^{\theta}\Phi_j)\, dp \\
&= \kappa_{nij}^{W} \Big[-\frac{1}{\Phi_j} \exp\Big(- p^{\theta}\Phi_j \Big) \Big]_0^\infty \\
&= \frac{\kappa_{nij}^{W}}{\Phi_j} \quad \text{where } \Phi_j \equiv \sum_{n\in N} \sum_{i\in S_n} \sum_{m\in\{\mathrm{W,B}\}} \kappa_{nij}^{m} = \sum_{n\in N} \sum_{i\in S_n} \sum_{m\in\{W,B\}} T_{ni}\,(c_n\,\tau_{nj}^{m})^{-\theta} \qquad \text{(C.5)}
\end{aligned}
$$

Apply same derivation for the case $m = B$ for firm i and use full notation, I can define two probability as follows:

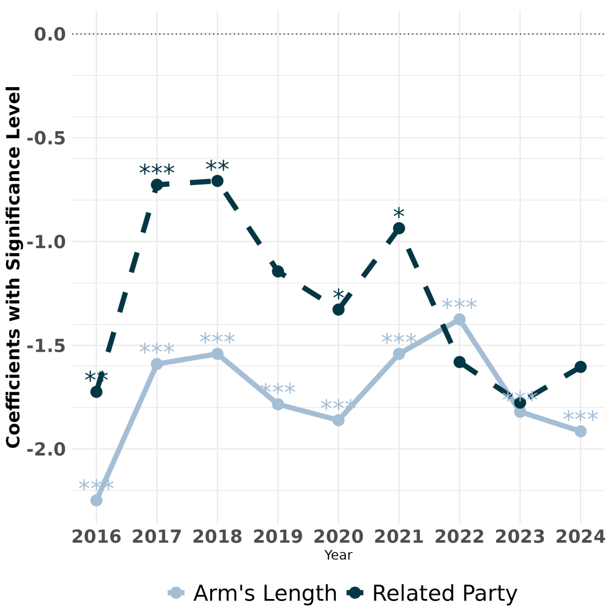
$$
\Pi_{nij}^{W} = \frac{\kappa_{nij}^{W}}{\Phi_j} = \frac{T_{ni}(c_n \tau_{nj}^{W})^{-\theta}}{\sum_{n\in N} \sum_{i\in S_n} \sum_{m\in\{W,B\}} T_{ni}\,(c_n\,\tau_{nj}^{m})^{-\theta}} \quad \text{and}
$$

$$
\Pi_{nij}^{B} = \frac{\kappa_{nij}^{B}}{\Phi_j} = \frac{T_{ni}(c_n \tau_{nj}^{B})^{-\theta}}{\sum_{n\in N} \sum_{i\in S_n} \sum_{m\in\{W,B\}} T_{ni}\,(c_n\,\tau_{nj}^{m})^{-\theta}}
$$

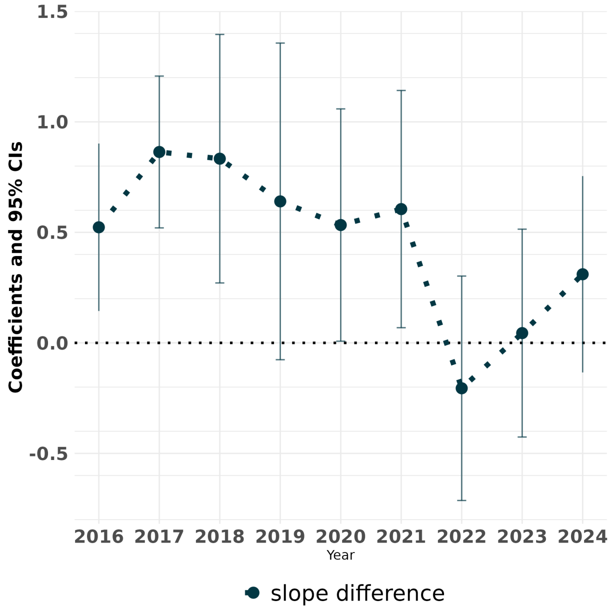
D Appendix - For Section 5

D.1 PPML Results with one-way cluster SE

Figure A7: Heterogeneous distance (semi-)elasticity: related-party vs. arm's-length export - PPML, firm-cluster



(a) Difference Elasticity

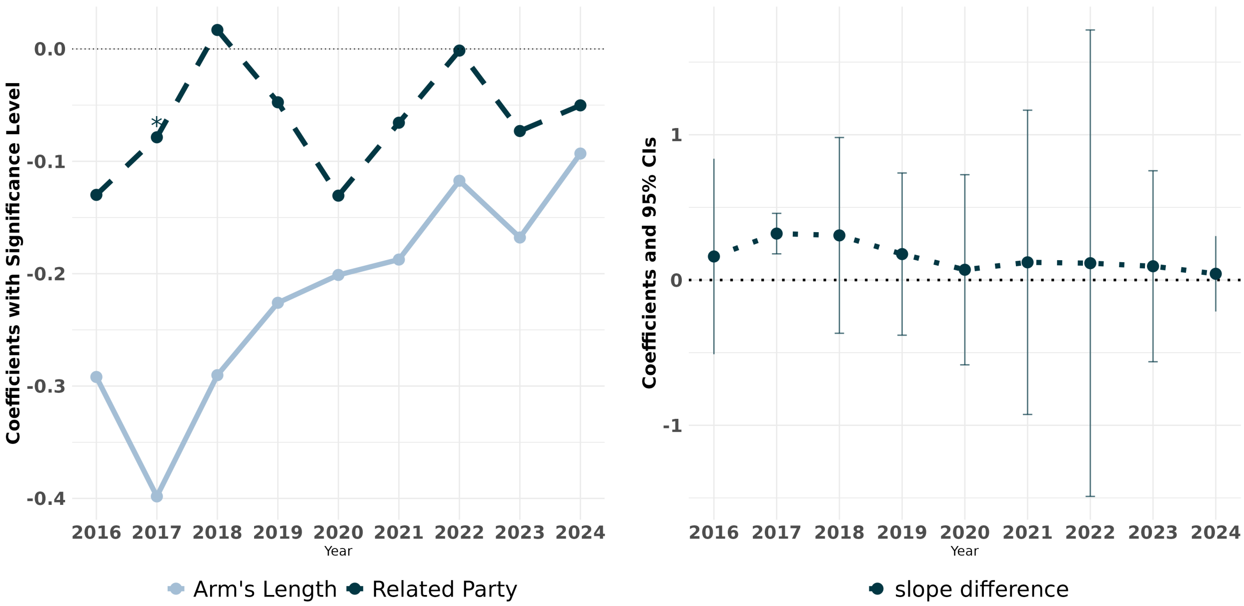


(b) Difference in Distance Elasticity

Notes: Robust standard errors are clustered by MNE affiliate.

D.2 OLS Results

Figure A8: Heterogeneous distance (semi-)elasticity: related-party vs. arm's-length export - OLS, multi-cluster

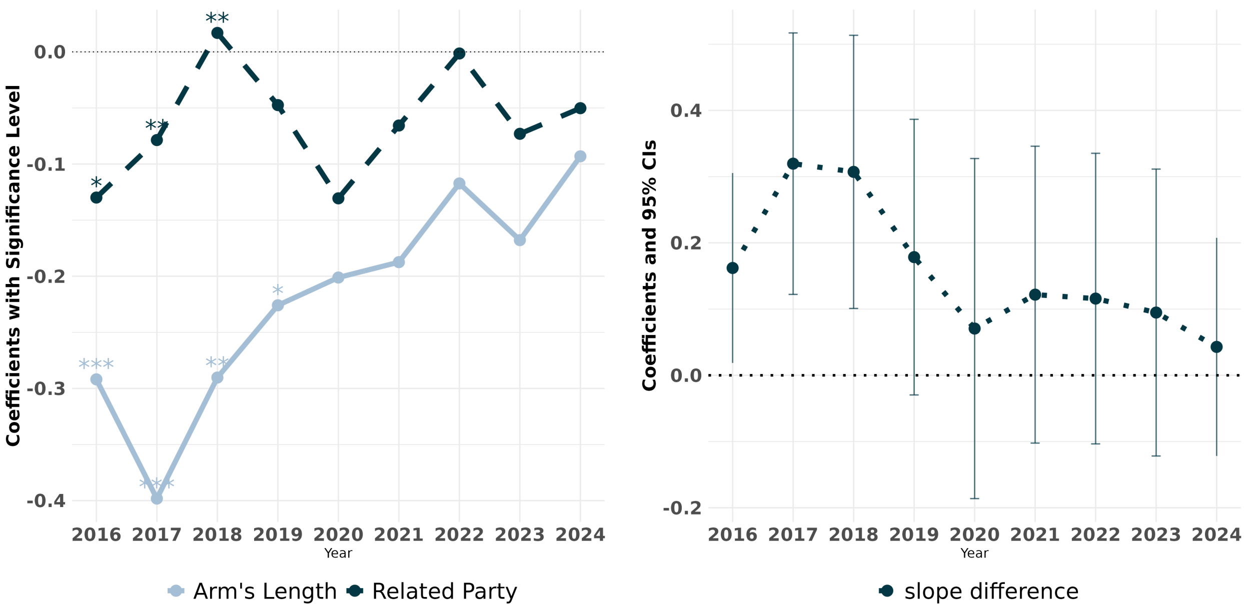


(a) Difference Elasticity

(b) Difference in Distance Elasticity

Notes: Robust standard errors are multi-way clustered by MNE affiliate, source country, product, and destination to account for correlated errors across multiple dimensions.

Figure A9: Heterogeneous distance (semi-)elasticity: related-party vs. arm's-length export - OLS, firm-cluster



(a) Difference Elasticity

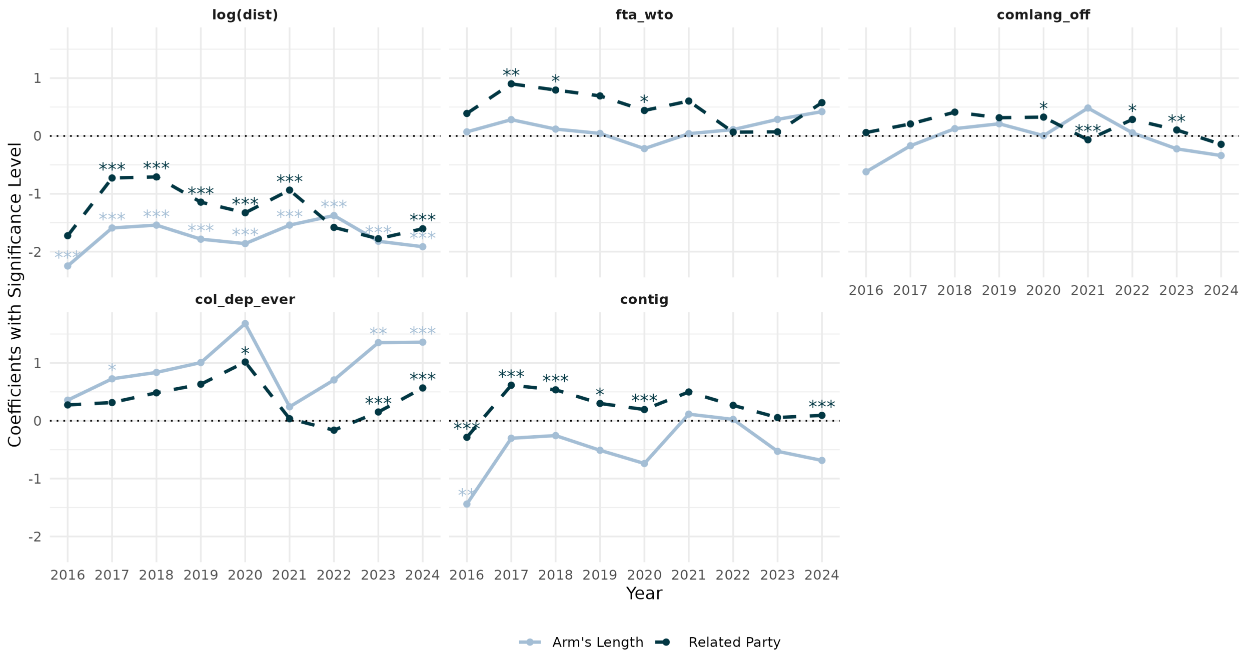
(b) Difference in Distance Elasticity

Notes: Robust standard errors are clustered by MNE affiliate.

D.3 Results on Other Gravity Variables

D.3.1 PPML

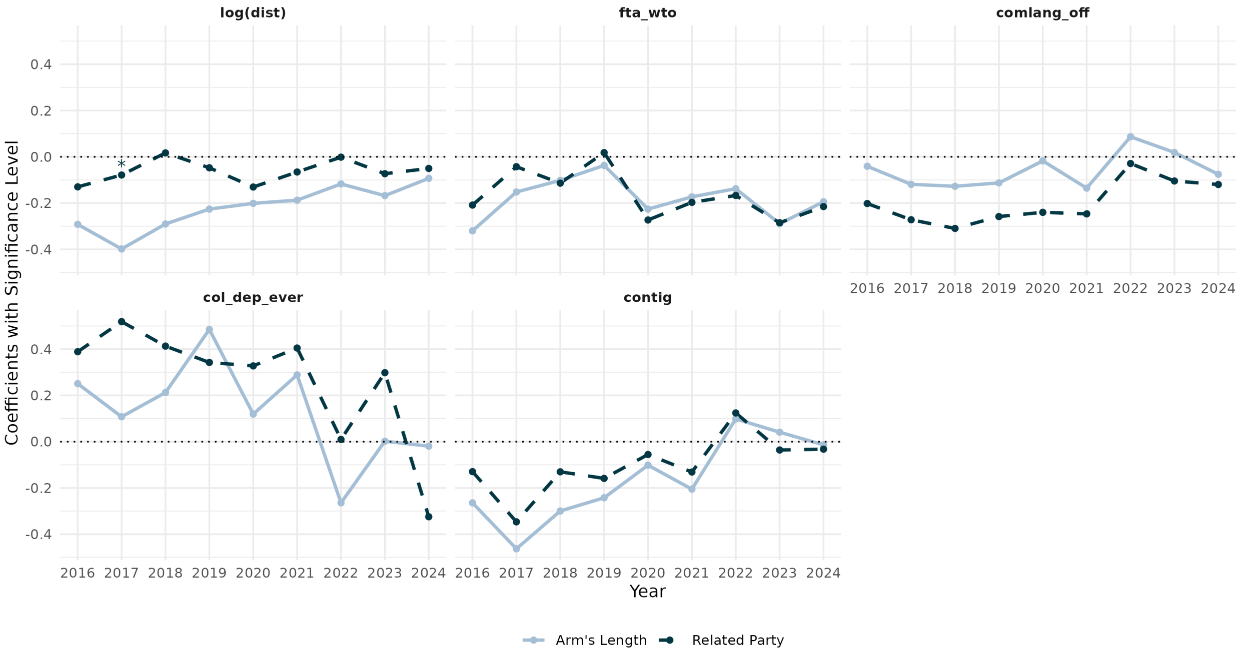
Figure A10: Heterogeneous Effects of Gravity Variables on Related-Party vs. Arm's-Length Exports - PPML, multi-cluster



Notes: The regressors include an interaction term between gravity variables and a within-transfer dummy. The baseline category is arm's-length trade. P-value significance levels are indicated as follows: $^{***}p < 0.001$,** $p < 0.01$,* $p < 0.05$. P-value labels on within-firm trade indicate the significance of the interaction term only, not the linear combination of the baseline and interaction. Robust standard errors are multi-way clustered by MNE affiliate, source country, product, and destination.

D.3.2 OLS

Figure A11: Heterogeneous Effects of Gravity Variables on Related-Party vs. Arm's-Length Exports - OLS, multi-cluster



Notes: The regressors include an interaction term between gravity variables and a within-transfer dummy. The baseline category is arm's-length trade. P-value significance levels are indicated as follows: $^{***}p < 0.001, ^{**}p < 0.01, ^{*}p < 0.05$. P-value labels on within-firm trade indicate the significance of the interaction term only, not the linear combination of the baseline and interaction. Robust standard errors are multi-way clustered by MNE affiliate, source country, product, and destination.